

LONDON
SCHOOL *of*
HYGIENE
& TROPICAL
MEDICINE



Genomic Characterisation of Chromosomally Integrated Human Herpesvirus.

Joshua Thomas Gilmour Tweedy

Thesis submitted in accordance with the requirements for the degree of Doctor of
Philosophy.

February 2016

Supervisor: Dr. Ursula Gompels
Department of Pathogen Molecular Biology
Faculty of Infectious and Tropical Diseases
London School of Hygiene and Tropical Medicine
University of London

Funding Details – LSHTM Graduate Teaching Studentship & HHV-6 Foundation

Declaration

I, Joshua Tweedy, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

This thesis would not have been possible without the support and guidance of my supervisor, Dr. Ursula Gompels. As well as the London School of Hygiene & Tropical Medicine for the graduate teaching assistant position which allowed me to conduct these studies and the HHV-6 foundation which helped fund the materials required to conduct this work. Finally I would like to thank Emma Walker for the support and encouragement provided throughout the final stages of this work.

Abstract

Distinctions in the biological features of human herpesvirus-6A (HHV-6A) and -6B (HHV-6B) have recently led to their reclassification from variants to distinct viral species. Unique for human herpesviruses, these viruses also exist in an endogenous form in ~1% of the human population, resulting from germ-line chromosomal integration. Chromosomally integrated human herpesvirus-6A (CI-HHV-6A) and -6B (CI-HHV-6B) genomes are inherited in a Mendelian manner, leading to the potential for viral gene expression and/or reactivation in every cell of the body. As yet, the effects of these integrated viral genomes on health and their relationship to circulating viral strains remain unclear. To address this, next generation sequencing (NGS) methods were established, to define a readily utilisable pipeline for genomic sequencing of these herpesviruses directly from cell or tissue samples. This was first tested to derive the complete genome sequence of a third HHV-6A strain, AJ. This showed close conservation with a recent North American isolate despite distinct geographic origins, which may reflect highly evolved viral status or recent emergence. Subsequently, using this defined NGS methodology, supplemented by Sanger sequencing, integrated viral genomes were characterised. Integration site analyses revealed a high prevalence of integration on chromosome 17, all with a shared junctional architecture, indicative of a single ancestral germ-line integration event at this chromosome. Genomic analyses revealed CI-HHV-6A retains the full HHV-6A gene complement, with no interrupted open reading frames, and important *cis* acting signals likely required for any viral replication/reactivation mechanism. Deep sequencing with minority variant analysis suggested a potential reactivation mechanism, resulting from superinfection with circulating viral strains. CI-HHV-6A was also found to diverge at a set of genes which have been used as markers of speciation amongst roseoloviruses, while retaining many of the features of HHV-6A. One of the key viral immunomodulatory genes, the chemokine receptor U51, was investigated further. This showed specific coding changes and a panel of viral-human chimeric and viral mutant receptor expression vectors were generated to investigate this. Initial structural modelling and functional characterisations were made showing distinctions. Overall, the results identify unique characteristics of the integrated genomes compared to known circulating viral strains.

Table of Contents

Declaration	2
Acknowledgements	3
Abstract	4
Table of Contents	5
List of Figures & Tables	10
Chapter 1: Introduction	12
1.1 The <i>Herpesviridae</i>	13
1.1.1 Classification & Nomenclature	13
1.1.2 General characteristics of <i>Herpesviridae</i> subfamilies	13
1.1.3 Virion Morphology	15
1.1.4 Genome	16
1.1.5 General Characteristics of Lytic Replication	18
1.1.6 General Characteristics of Latency	19
1.2 Human Herpesvirus-6A & -6B	20
1.2.1 Introduction	20
1.2.2 Genome	21
1.2.3 Cellular & Tissue Tropism	23
1.2.4 Replication Cycle & Latency	24
1.2.5 Transmission and Epidemiology	26
1.2.6 Primary Infection, Reactivation & Reinfection	27
1.2.7 Diagnosis	28
1.2.8 Disease Associations	28
1.2.9 Therapy	32
1.3 Chromosomal Integration of Human Herpesvirus-6A & -6B	32
1.3.1 Inherited Chromosomally Integrated HHV-6A and -6B	32
1.3.2 Somatic Integration of HHV-6A & HHV-6B	34
1.3.3 Reactivation from Chromosomally Integrated HHV-6A & -6B	34
1.4 HHV-6A/-6B and the Chemokine System	35
1.4.1 Immunomodulation by HHV-6A and -6B	35
1.4.2 The Chemokine System	36
1.4.3 HHV-6A and -6B Chemokines	37
1.4.4 HHV-6A and -6B Encoded Chemokine Receptors	38
1.5 Aims	41

Chapter 2: Methods	42
2.1 DNA Samples, Vectors and Reference Sequences	42
2.1.1 Virus Reference Strains and Infected Cell DNA	42
2.1.2 CI-HHV-6A and -6B Patient DNA	42
2.1.3 Virus Genome Reference Sequences	42
2.1.4 Plasmid DNA	43
2.2 Primer Design	43
2.3 Primer Sequences	43
2.3.1 Long Range PCR Amplicons	43
2.3.2 AJ Genomic Sequence Gap Filling	44
2.3.3 CI-HHV-6A/-6B Gene Amplification	45
2.3.4 Viral Chromosomal Integration Site Junction Amplifications	45
2.3.5 U51A Chimeric and Mutant Construct Generation	45
2.3.5.1 <i>U51NCCR3M</i>	45
2.3.5.2 <i>HA-U51NCCR3</i>	46
2.3.5.3 <i>HA-U51ΔE2G</i>	46
2.3.5.4 <i>U51A Alanine-scanning</i>	46
2.3.6 Vector Sequencing	47
2.4 Polymerase Chain Reaction (PCR)	48
2.4.1 General	48
2.4.2 Long Range PCR Amplicons	48
2.4.3 Chromosome 17p Integration Site PCR Amplification	48
2.5 Agarose Gel Electrophoresis and DNA Purification	49
2.6 DNA Quantification	49
2.6.1 General	49
2.6.2 Long Range PCR Amplicon DNA Quantification	50
2.7 Sanger Nucleotide Sequencing and Resolution	50
2.8 Illumina MiSeq NGS Sequencing, Assembly and Annotation	50
2.8.1 Target Enrichment using Agilent SureSelect Target Enrichment	50
2.8.2 Target Enrichment using LPCR Amplicons	51
2.8.3 Illumina MiSeq Sequencing	51
2.8.4 NGS Bioinformatics: Sequence Assembly	51
2.8.5 NGS Bioinformatics: Annotations	52
2.8.6 NGS Bioinformatics: Minor Variant SNP Analyses	52
2.9 Sequence Alignments and Phylogenetic Relationships	52
2.10 Molecular Modelling of U51A-CCL2 Interaction	53
2.11 Plasmid Cloning	53
2.11.1 PCR-driven Overlap Extension	53

2.11.2 HA-U51NCCR3	54
2.11.3 Alanine-scanning Mutagenesis	54
2.11.4 Restriction Digestion	55
2.11.5 Ligation	55
2.11.6 Transformation	55
2.11.7 Plasmid Preparation	56
2.12 Cells and Cell Culture	56
2.12.1 Cell Types	56
2.12.2 Resuscitation of Cells and Storage	57
2.12.3 Cell Counting	58
2.12.4 Expression Vector Transfection	58
2.13 Chemokines and Chemotaxis Assay	58
2.13.1 Chemokines	58
2.13.2 Calcein-AM Staining	58
2.13.3 Chemotaxis Assay	58
2.14 Flow Cytometry	59
Chapter 3: Next Generation Sequencing of Roseoloviruses	60
3.1 Introduction	60
3.2 Establishment of Next Generation Sequencing Protocols for HHV-6A Genomes	62
3.2.1 SureSelect Target Enrichment System	62
3.2.2 Long-range PCR Amplicons	65
3.3 Resequencing of HHV-6A strain U1102 sequence	69
3.4 Construction of the Complete Genome Sequence of HHV-6A Strain AJ	71
3.5 Analyses of the Complete Genome of HHV-6A Strain AJ	72
3.6 Discussion	75
Chapter 4: Genomic Analyses of CI-HHV-6A and Reactivation	80
4.1 Introduction	80
4.2 Characterisation of a Common Integration Site at Chromosome 17p	80
4.3 Genomic Analysis of CI-HHV-6A	84
4.3.1 CI-HHV-6A Genome Determination	84

4.4 CI-HHV-6A Variation	88
4.5 Cis-acting Sites Required for DNA Packaging, Genome replication and Gene Regulation are Conserved in CI-HHV-6A	92
4.6 Deep Sequencing Shows Minor Variants in CI-HHV-6A Genomes with Gene Expression	96
4.7 Discussion	99
Chapter 5: Variation and Functional Analysis of the Viral Chemokine Receptor U51	103
5.1 Introduction	103
5.2 U51 Variation in CI-HHV-6A and -6B	104
5.3 Molecular Modelling of U51A Amino-terminal Domain-CCL2 Interaction	109
5.4 Generation of Chimeric and Mutagenic Receptor Expression Vectors	114
5.5 Surface Expression of U51NCCR3M Receptor Construct	116
5.6 Functional Analysis of U51NCCR3M Construct	117
5.7 Discussion	118
Chapter 6: Discussion	121
6.1 Introduction	121
6.2 Roseolovirus Next Generation Sequencing	121
6.3 Resequencing of HHV-6A Strain U1102	123
6.4 Sequencing of HHV-6A Strain AJ	125
6.5 Germ-line Chromosomal Integration of HHV-6A and -6B	126
6.6 Somatic Chromosomal Integration of HHV-6A and -6B	132
6.7 Reactivation of Integrated HHV-6A/-6B	136
6.8 Functional Consequences of CI-HHV-6A/-6B	138
6.9 CI-HHV-6A U51	139

6.10 Future Work	142
6.11 Concluding Remarks	143
Chapter 7: Appendices	145
7.1 The Human Chemokine System	145
7.1.1 Introduction	145
7.1.2 Chemokines	145
7.1.3 G Protein-Coupled Chemokine Receptors	148
7.1.4 Chemokine-Chemokine Receptor Interactions	150
7.1.5 Chemokine Receptor Structure and Activation	152
7.1.6 G proteins	153
7.1.7 Regulation of Chemokine Receptor Signalling	154
7.1.8 Chemokine Receptor Dimerisation	155
7.2 HHV-6A strain AJ Gene List	157
7.3 Chromosome 17p13.3 CI-HHV-6A Integration Site Junction Nucleotide Sequences	197
7.4 Alignments for Determination of Practical Level of Read Coverage for Accurate Sequence Resolution	202
7.4.1 U38 partial CDS fragment	203
7.4.2 U46	204
7.4.3 U47	205
7.4.4 U51	211
7.4.5 U65	214
7.4.6 U83	217
7.4.7 U47 SNP Resolution from CI-HHV-6A Isolate 5055	218
7.5 CI-HHV-6A Conserved Gene Set Nucleotide Sequences	221
7.6 CI-HHV-6A Variable Gene Set Nucleotide Sequences	277
7.7 Alignment of the Origin of Lytic Replication of CI-HHV-6A	305
7.8 Optimisation of CCR5 Surface Expression	309
7.9 Generation of Stably Expressing Cell Lines	311
7.10 Actin Polymerisation Assay	312
References	313

List of Figures

Figure 1.1 Schematic representation of the herpesvirus virion	16
Figure 1.2 Herpesvirus genome classes	17
Figure 1.3 Arrangement of conserved gene blocks in the <i>Herpesviridae</i> subfamilies	18
Figure 1.4 HHV-6A genome organisation	22
Figure 1.5 CI-HHV-6A/-6B chromosomal integration	34
Figure 1.6 Chemokine ligands and their receptors	37
Figure 2.1 Generation of the U51NCCR3 receptor construct	46
Figure 2.2 PCR-mediated overlap extension to generate an N-terminally swapped chimeric receptor	54
Figure 3.1 Coverage and mapping quality of SureSelect enriched HHV-6A strain U1102 sequence reads across the X83413 reference genome	64
Figure 3.2 Coverage and mapping quality of SureSelect enriched HHV-6A strain AJ sequence reads across the X83413 reference genome	66
Figure 3.3 Coverage and mapping quality of LPCR enriched HHV-6A strain AJ sequence reads across the X83413 reference genome	67
Figure 3.4 Confirmation of previously identified corrections to the HHV-6A strain U1102 genome sequence	70
Figure 3.5 Further sites of variation identified by resequencing of HHV-6A strain U1102	71
Figure 3.6 Phylogenetic analyses of HHV-6A strain AJ to compared to sequenced HHV-6A/-6B strains	73
Figure 3.7 Variation in the DNA polymerase genes of HHV-6A strains and potential effects on the SSPBRT-PCR species typing assay	75
Figure 4.1 Phylogenetic analyses of viral DR _R sequence from viral chromosomal integration site junction amplifications	82
Figure 4.2 Alignments of CI-HHV-6A integration site junctions at chromosome 17p13.3 show common perfect and imperfect telomeric repeat structures	83
Figure 4.3 Comparison of sub-telomere adjacent repeat structure between CI-HHV-6A integrated at chromosome 17p and other defined chromosomal integration sites	84
Figure 4.4 Sequence read coverage of CI-HHV-6A genomes across the HHV-6A strain U1102 reference genome following mapped assemblies	86
Figure 4.5 Phylogenetic analyses of CI-HHV-6A compared to prototypical human herpesvirus species from all three <i>Herpesviridae</i> sub-families	89
Figure 4.6 Phylogenetic analyses of CI-HHV-6A and HHV-6A/-6B conserved and variable gene sets	92
Figure 4.7 Phylogenetic analyses of the origin of lytic replication (Mori) of CI-HHV-6A isolates and HHV-6A/-6B strains	93
Figure 4.8 Alignment of CI-HHV-6A and HHV-6A/-6B pac sites with those of the prototypical strain HHV-6A U1102	94
Figure 4.9 CpG frequency plots across the CI-HHV-6A genomes	95
Figure 4.10 Conservation in CI-HHV-6A of the pre-miRNA sequences identified in HHV-6B	96

Figure 5.1 Phylogenetic analyses of U51 gene from CI-HHV-6A/-6B isolates with HHV-6A/-6B strains	106
Figure 5.2 U51A amino acid sequence variation	108
Figure 5.3 Secondary structure predictions of amino-terminal domain of U51A	110
Figure 5.4 Search space parameters for ligand docking modelling	111
Figure 5.5 Modelling and predicted affinity of the interaction between the amino-terminal domain of HHV-6A strain U1102 U51 and the crystal structure of the chemokine ligand CCL2	113
Figure 5.6 Modelling and predicted affinity of the interaction between the amino-terminal domain of the U51 from the CI-HHV-6A isolates and the crystal structure of the chemokine ligand CCL2	114
Figure 5.7 Surface expression of U51NCCR3M chemokine receptor construct in L1.2 cells	117
Figure 5.8 Chemotaxis of U51NCCR3M cells in response to the chemokine ligand CCL2	118
Figure 7.1 Schematic representation of the structural motifs which retain chemokine tertiary structure and define the four chemokine subfamilies	147
Figure 7.2 Schematic representations of chemokine receptor topology	149
Figure 7.3 Hypothetical model of the chemokine-receptor interaction	152
Figure 7.4 L1.2 cell growth time curve	309
Figure 7.5 Surface expression of CCR5 at time points following transfection	309
Figure 7.6 Effect of transfected DNA on surface expression of CCR5 at 48 hours post transfection	310
Table 1.1 The nine members of the Herpesviridae with humans as a primary host	14
Table 1.2 Chemokine receptors and their expression	36
Table 2.1 LPCR amplicon primer sequences	44
Table 3.1 Percentage identity between HHV-6A CDS features from mapped assemblies methods with the complete sequence of HHV-6A strain AJ	69
Table 3.2 Percentage identity between defined CDS of HHV-6A strain AJ and sequenced HHV-6A strains	74
Table 4.1 Sequence read and coverage statistics for the CI-HHV-6A genomes subjected to NGS	85
Table 4.2 Read coverage across CI-HHV-6A CDS from the UL genomic region	88
Table 4.3 Pairwise comparisons of the CDS features of HHV-6A/-6B strains and CI-HHV-6A isolates to HHV-6A strain U1102	90
Table 4.4 CI-HHV-6A conserved and variable gene sets	91
Table 4.5 Minor variants in CI-HHV-6A isolates identified by deep sequencing	98
Table 5.1 Chemokine binding and chemotaxis properties of HHV-6A U51 and CCR3	115
Table 5.2 Generated chimeric and mutant receptor constructs	115
Table 7.1 Determination of practical level of read coverage for accurate genome sequence resolution	202

Chapter 1: Introduction

Herpesviruses are a group of large DNA viruses capable of causing infections in a wide range of vertebrate hosts, as well as some invertebrate hosts. Nine members of this group of viruses infect humans as their primary host; and it is the two most recently classified of these human herpesviruses, the betaherpesvirus roseoloviruses: human herpesvirus-6A (HHV-6A) and human herpesvirus-6B (HHV-6B) [King et al., 2012; Ablashi et al., 2014], that are the subject of this thesis. A major defining trait of the herpesviruses is their ability to establish life-long latent infection in the host, with periodic reactivation in the face of strong anti-viral immune responses. Therefore this failure of the host to completely eliminate these viruses makes it clear that herpesviruses must be particularly adept at establishing molecular genetic forms and mechanisms for evading or countering the host's immune responses. Indeed, a number of herpesviruses have even evolved means to manipulate host immune responses to aid viral replication.

One mechanism employed by herpesviruses to permit this persistence is the establishment of a latent state, whereby the viral genome is maintained in the host cell in the absence of infectious virion production. The primary mode of latency for many herpesviruses is maintenance of the viral genome as a nuclear episome in the infected cell; with the expression of a limited subset of viral latency genes to permit the replication and persistence of the episomes [Grinde, 2013]. However, recent findings with HHV-6A and HHV-6B suggest a new paradigm. Unique amongst the human herpesviruses, both HHV-6A and -6B can integrate into telomeric regions of the host cell chromosome and it has been proposed that this integration provides an alternative mechanism of somatic latency [Arbuckle et al., 2010; 2013]. Irrespective of whether this represents a natural latency mechanism, *in vivo* integration of these viruses can, and has occurred. This is exemplified by the integration of both HHV-6A and -6B into the germ-line which has resulted in an endogenous form of these viruses in ~1% of populations studied to date [Pellett et al., 2012]. As yet, the effects of these integrated viral genomes on health and their relationship to circulating viral strains remain unclear. A major aspect of this thesis concerns the characterisation of the integrated forms of these viruses with comparisons to circulating viral strains.

Herpesviruses also extensively employ another strategy to achieve this persistence, that of immunomodulation. They dedicate significant amounts of their genomes to encoding products which can directly interfere with a wide array of innate and adaptive immune responses. As they cause infection and establish latency in key effector immune cells, the betaherpesvirus subfamily, to which HHV-6A and -6B belong, are particularly accomplished in this regard. One central immunomodulatory strategy employed most extensively by the betaherpesviruses is subversion of the host chemokine system, a key aspect in immune response signalling and leukocyte movement during an inflammatory response [McSharry et al., 2013; Dagna et al., 2013]. A number of the betaherpesvirus specific genes encode homologues of both chemokines and chemokine receptors. When expressed from infected cells these viral homologues modulate cellular signalling both to escape immune surveillance and assist viral replication, thereby contributing to persistence. This thesis investigates both the nature of the germ-line chromosomally integrated forms of HHV-6A/-6B. As well as an encoded immunomodulatory component, the virally encoded chemokine receptor U51, affecting the chemokine system.

1.1 The *Herpesviridae*

1.1.1 Classification & Nomenclature

Herpesviruses have been classified into the order *Herpesvirales* by the International Committee on Taxonomy of Viruses (ICTV) [Davison et al., 2009; King et al., 2012]. Historically, the primary characteristic for inclusion in the *Herpesvirales* is the possession of a distinct virion morphology. However, the establishment of latency following the resolution of acute primary infection, high level of adaption to their natural hosts, profound immunomodulatory effects and to a lesser extent the genome organisation can also be considered as defining features of the *Herpesvirales*.

The order *Herpesvirales* is divided into three families:

- i. *Alloherpesviridae* – encompassing herpesviruses primarily of bony fish and frogs;
- ii. *Malacoherpesviridae* – containing the viruses of invertebrates;
- iii. *Herpesviridae* – containing the mammal, bird and reptilian herpesviruses.

The latter of these families is the largest, containing ~90 viral species. The viruses within this *Herpesviridae* family are further divided into three subfamilies based originally on shared biological characteristics, but now increasingly including genome sequence phylogeny:

- i. *Alphaherpesvirinae*
- ii. *Betaherpesvirinae*
- iii. *Gammaherpesvirinae*

In addition to their classification, the ICTV have also outlined a formal naming convention for herpesvirus species. Under this scheme all designated virus species have an official nomenclature, named by the Family (or in some cases Subfamily) taxonomic name of its primary natural host followed by 'herpesvirus' and an arabic number [Roizmann et al., 1992]. However, a number of herpesviruses, especially those infecting humans, possessed names which were widely adopted before the introduction of this formal naming convention. Therefore these virus names are regarded as synonymous with the formally assigned name, although they are generally referred to with dual nomenclature e.g. Human cytomegalovirus/Human herpesvirus 5.

1.1.2 General characteristics of *Herpesviridae* subfamilies

Herpesviruses of humans are present in all three of the *Herpesviridae* subfamilies, table 1.1.

Formal Name	Common Name	Subfamily
HHV-1	Herpes simplex virus 1 (HSV-1)	α
HHV-2	Herpes simplex virus 2 (HSV-2)	α
HHV-3	Varicella zoster virus (VZV)	α
HHV-4	Epstein Barr virus (EBV)	γ
HHV-5	Human cytomegalovirus (HCMV)	β
HHV-6A		β
HHV-6B		β
HHV-7		β
HHV-8	Kaposi's sarcoma associated herpesvirus (KSHV)	γ

Table 1.1 The nine members of the *Herpesviridae* with humans as a primary host.

Alphaherpesvirinae

The Alphaherpesvirinae are primarily defined by their replication and latency in neuronal cells [Efstathiou & Preston, 2005]. Three members of the subfamily are regarded as having a human host. Firstly, herpes simplex virus 1 (HSV-1)/human herpesvirus 1 (HHV-1) which is primarily known for the clinical presentation of recurrent small blisters/sores around mouth known as herpes labialis/cold sores, as a result of episodes of reactivation [Roizman et al., 2013]. HSV-1 can be regarded as the prototypic virus of the *Herpesviridae* as a result of its use to elucidate much the basic molecular biology of the family. The closely related Herpes simplex virus 2 (HSV-2)/human herpesvirus 2 (HHV-2) causes a highly similar disease presentation to HSV-1 but with a differing tissue tropism, giving rise to recurrent genital lesions [Roizman et al., 2013]. Finally, varicella-zoster virus (VZV)/human herpesvirus 3 (HHV-3) causes a primary infection, often acquired during childhood, resulting in a acute vesicular rash on the torso known as varicella or chickenpox. In around 15% of cases, and especially seen in immunocompromised individuals, the latent virus can reactivate causing a painful skin rash known as zoster or shingles [Arvin & Gildea, 2013].

Gammaherpesvirinae

The Gammaherpesvirinae are unified by their establishment of latency in lymphocytic cells, as well as their association with lymphoproliferative diseases and other non-lymphoid cancers [Coleman et al., 2010; Amon & Farrell, 2005]. Two members of this subfamily cause widespread infection in humans. Epstein-Barr virus (EBV)/human herpesvirus 4 infecting epithelial and B-cells and is primarily known as an aetiological agent for infectious mononucleosis/glandular fever. However, it is also associated with certain types of cancers including Hodgkin's lymphoma, Burkitt's lymphoma, nasopharyngeal carcinoma [Longnecker et al., 2013]. The second of these is Kaposi's sarcoma-associated herpesvirus (KSHV)/human herpesvirus 8 which is associated with Kaposi's sarcoma and primary effusion lymphoma [Damania & Cesarman, 2013].

Betaherpesvirinae

This thesis concerns the third subfamily of the *Herpesviridae*, the *Betaherpesvirinae* which share the general characteristics of slow growth in comparison to the *Alphaherpesvirinae*

and the distinctive cytopathic effects of multinucleated, 'ballooning' cells which can be observed in infected cells. Latency is also established in subsets of haematopoietic cells, including bone marrow progenitor and monocytic lineage cells [Sinclair, 2008; Luppi et al., 1999; Miyake et al., 2006]. Four human herpesviruses are found within this subfamily: Human cytomegalovirus (HCMV)/human herpesvirus 5 (HHV-5) a member of the *Cytomegalovirus* genus; human herpesvirus 6A (HHV-6A); human herpesvirus 6B (HHV-6B) and human herpesvirus 7 (HHV-7), all members of the *Roseolovirus* genus.

HCMV is regarded as the prototypic virus for this subfamily. It possesses a genome which at ~235 kbp is substantially larger than those found amongst the roseoloviruses. HCMV is a ubiquitous infection worldwide, but has received most notoriety as one of the leading causes of congenital viral infection following intrauterine transmission of the virus to the foetus in approximately 1% of pregnancies [Kenneson & Cannon, 2007]. This can result in severe sequelae for the infant such as hearing loss, cognitive impairment and mental retardation [Whitley, 2004]. Like HCMV, HHV-6A and -6B also cause congenital infections at a similar prevalence, however, recent evidence has suggested these HHV-6A/-6B infections occur by a different mechanism to that of HCMV involving chromosomally integrated forms of these viruses [Hall et al., 2008; 2010], a feature discussed further in section 1.3. Primary infection with HCMV has long been regarded as asymptomatic, rarely giving rise to a mononucleosis-like disease in older children and adults [Grundy, 1990]. Subsequently the virus predominantly establishes latency in cells of myeloid lineage, periodically reactivating, but may also cause chronic low level infection predominantly at endothelial and epithelial sites [Goodrum et al., 2012]. However, increasingly links are being established between HCMV infection and developmental impairment in infants and immunosenescence [Pawelec et al., 2009; Whitley, 2004; Gompels et al., 2012].

The three members of the *Roseolovirus* genus, HHV-6A, HHV-6B and HHV-7, are T-lymphotropic viruses [Lusso et al., 1988; Takahashi et al., 1989; Berneman et al., 1992]. Primary infection with members of this genus is associated with fever and occasionally a rash, exanthem subitum/roseola infantum [Tanaka et al., 1994; Ward, 2005]. For these viruses, both lytic replication and the establishment of latency is achieved in effector cells of the immune system [Kondo et al., 1991; Miyake et al., 2006]; therefore these viruses are under close surveillance by the host's immune system. As such they have evolved a variety of means to directly evade via latency or indirectly interfere with innate and adaptive immune responses.

Chromosomal integration/latency, speciation and immunomodulatory capabilities of HHV-6A and -6B are addressed in this thesis and discussed in sections 1.2, 1.3 and 1.4. However, as mentioned above the *Herpesviridae* are grouped together due to a number of shared characteristics, such as virion morphology and the establishment of latency. Indeed, homology between members of this family has been key to the elucidation of much of the basic molecular biology of these viruses. Therefore the remainder of this section is devoted to the discussion of a number of general aspects of the *Herpesviridae* which relate to the HHV-6A and -6B topics in this thesis.

1.1.3 Virion Morphology

The advent of nucleic acid sequencing and vast advances in high-throughput sequencing technology within the last decade have led to sequence based phylogeny emerging as a major influence on classification. However, historically herpesviruses were identified

through the possession of a distinct morphology of their mature infectious virion. It consists of linear double-stranded DNA (dsDNA) genome packaged into a $T = 16$ icosahedral capsid. This highly organised capsid structure consists of a total of 162 capsomeres, 150 hexameric capsomeres and 11 pentameric capsomeres, with the final pentameric position being occupied by a portal complex required for entry and exit of the genomic DNA [Brown & Newcomb, 2011]. The ~125 nm capsid structure is then coated in a much less ordered proteinaceous matrix known as the tegument. The tegument is composed from a large number of different viral proteins important for virion egress and the very early stages of host cell infection [Pellett & Roizman, 2013; Guo et al., 2010]. Finally the tegument is in turn wrapped in a host-derived lipid envelope producing a ~200 nm in diameter virion, figure 1.1 [Mettenleiter et al., 2009]. A number of viral glycoproteins protrude through the lipid envelope which are essential for viral replication. These contribute to the processes of virion attachment, cell entry as well as the envelopment and maturation of viral particles and thus are a major contributing factor to biological differences that define herpesvirus speciation [Eisenberg et al., 2012].

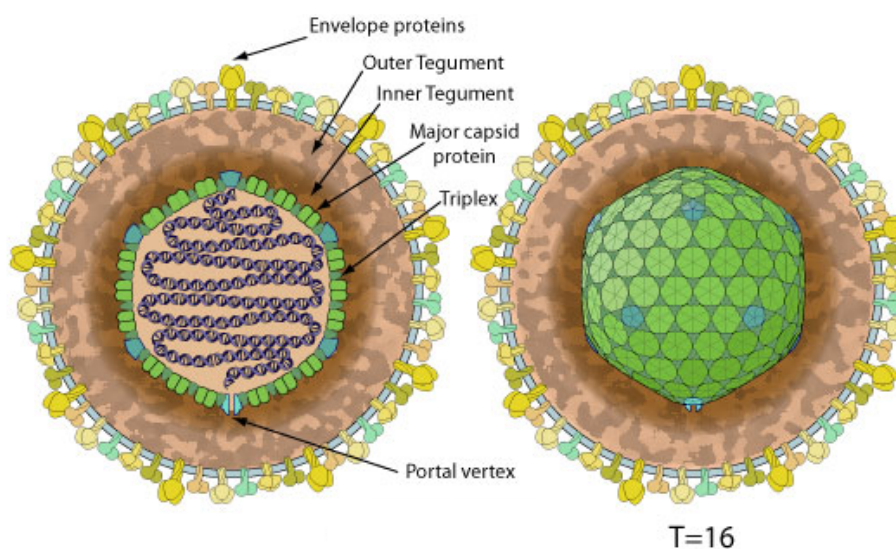


Figure 1.1 Schematic representation of the herpesvirus (HSV-1) virion [Figure from Hulo et al., 2011].

1.1.4 Genome

Densely packaged into the capsid core of the mature infectious herpesvirus virion is the viral genomic DNA. Which despite large differences in size, composition and sequences retain a number of distinct characteristics. Herpesviruses possess linear, dsDNA (Baltimore classification group I) genomes, which contain a number of direct or inverted repeats and possess unpaired nucleotides at their termini, as seen in figure 1.2 [Pellett & Roizman, 2013].

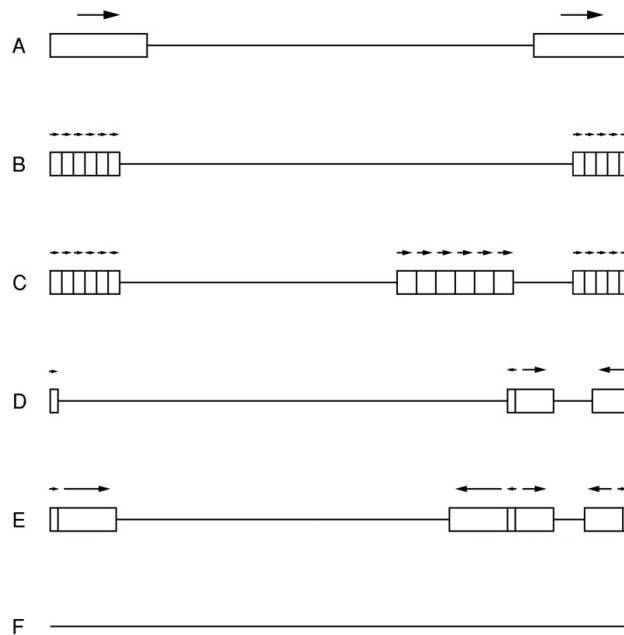
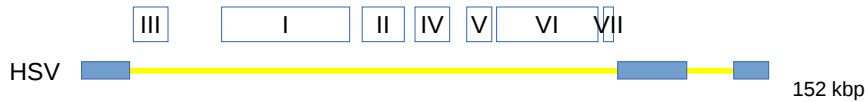


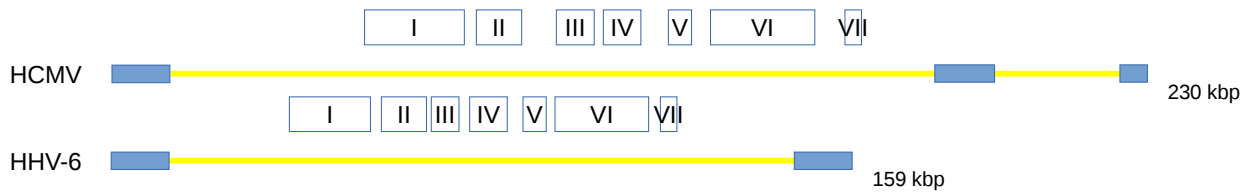
Figure 1.2 Herpesvirus genome classes. Unique regions shown as horizontal lines. Repeat regions shown as boxes with arrows indicating orientation. (A) Class A consists of a unique sequence flanked by direct repeats. (B) Class B again consists of a unique region flanked by repeats, but these repeats consist of tandemly repeated sequence. (C) Class C resembles the class B genome structure with an additional internal set of direct repeats. (D) Class D contain two unique regions (long and short) each flanked by inverted repeats. (E) Class E resembles class D but with larger inverted repeats. (F) Class F is represented by a single member of the Betaherpesvirinae, which lacks the repeat structures found in other herpesviruses [Figure adapted from Davison, 2007].

The relationship between the genomes of members of three families that make up the *Herpesvirales* can be difficult to ascertain, due to the large amount of divergence time since the last common ancestor. In fact, only three genes display clear nucleotide sequence homology between the three families: the DNA polymerase (HSV-1 UL30), dUTPase (HSV-1 UL50) and the terminase (HSV-1 UL15). Even then, for two of these genes, the DNA polymerase and dUTPase, independent gene capture events or convergent evolution in the different families cannot be ruled out [McGeoch et al, 2006]. However, within the relatively closer related *Herpesviridae* family, an increasing amount of distinctive genomic features can be observed. These three subfamilies possess a number of 'core' genes, representing those which appear to be from a common ancestor [Davison, 2007]. Generally speaking these core genes encode products required to make a virus particle and are essential for growth in cell culture. They tend to be positioned towards the centre of the genomes, while genes obtained since the respective divergence events occupy positions nearer the termini of the genomes. In addition to their central location, they are found within seven so called 'gene blocks' of conserved genes (figure 1.3), which in most cases are ordered in the same manner in members of the same subfamily [Gompels et al., 1995; Davison, 2007]. There are 43 *Herpesviridae* core genes, however, three of these genes, the thymidine kinase, helicase and the small subunit from the ribonucleotide reductase, have been lost from certain lineages [McGeoch et al, 2006].

Alpha



Beta



Gamma



Figure 1.3 Arrangement of conserved gene blocks in the *Herpesviridae* subfamilies. Genomic unique regions are designated a solid line, repeat regions as shaded boxes and gene block locations labelled I-VII [Adapted from Gompels et al., 1995].

1.1.5 General Characteristics of Lytic Replication

The replication cycle of all herpesviruses share notable similarities, with HSV-1 used as the general model for herpesvirus replication. Primary attachment of the virion to the host cell surface occurs through relatively non-cell type specific and reversible interactions between viral envelope glycoproteins and glycosaminoglycans (GAGs) [Shukla & Spear, 2001]. This is followed by higher affinity cell type specific interaction(s) with the viral entry receptor(s). Within the *Herpesviridae* three envelope glycoproteins are conserved which include, gH and gL, which form the gH/gL complex and gB. These have been shown to be core components in the viral entry machinery for members of all three subfamilies [Reske et al., 2007; Ryckman et al., 2008; Isaacson & Compton, 2009; Liu et al., 1993; Chou & Marousek, 1992; Hutt-Fletcher, 2007]. However, many herpesviral species possess additional glycoproteins and accessory proteins which are essential to the attachment and entry process, the diversity in which plays a large part in determination of the distinct cell tropisms of each species [Krummenacher et al., 2013; Mori et al., 2003a, 2004].

While endocytosis and viral fusion with the endosomal membrane has been noted for some species *in vitro*, the primary mechanism of entry *in vivo* appears to be fusion of the viral envelope with the plasma membrane of the host cell [Nicola et al., 2003; Spear & Longnecker, 2003]. Following the fusion event, the viral capsid and associated tegument proteins are deposited into the cytoplasm of the host cell, from where they must gain access to the nucleoplasm to continue the replication cycle. As diffusion of molecules larger than ~500 kDa is subject to restrictions in the cytoplasm, the capsid requires an active transport mechanism to reach the nuclear pore complexes for nuclear entry [Sodeik, 2000]. For this, the herpesvirus capsids utilise components of the hosts cytoskeletal transport machinery to traverse the cytoplasm on microtubule networks. The capsids associate with ATP catalysing dynein/dynactin motor complexes, to move to microtubule organising centres anchored in close proximity to the nucleus [Dohner et al., 2002; Dohner & Sodeik, 2005]. Once at this site they bind filaments associated with the nuclear pore complexes; interactions which lead to destabilisation of the capsid, release of the viral

DNA and its transport through the nuclear pore to the nucleoplasm [Ojala et al., 2000; Radtke et al., 2006].

Once in the nucleoplasm, covalent linkage of the termini of the linear genome results in a circularised form [Garber et al., 1993]. Subsequently, with the aid of tegument proteins, host derived transcription machinery is utilised for transcription of viral genes. This gene transcription is tightly regulated and follows a 'cascade' mechanism, in which viral proteins are transcribed and translated in three groups sequentially, with expression of members of the next kinetic class of genes being dependent on expression of genes from the previous class. Firstly, the immediate early (IE) group comprising genes predominantly involved in regulation of viral gene expression and establishment of a suitable replication environment. This is followed by the early (E) group, primarily including those genes responsible for DNA replication. Finally the late (L) gene group, possessing genes for the structural components and glycoproteins of the virions [Hones & Roizman, 1974; Tsao et al., 2009; Jenner et al., 2001].

DNA replication is initiated through DNA rearrangements which permit the access of a number of virally encoded initiator proteins to the origin of lytic replication(s) (*oriLyt*). Subsequently a virally encoded polymerase/elongation complex is formed and genomic replication proceeds via 'rolling circle' replication, in which the circularised parental genome results in the production of head-to-tail full length genomic concatemers [Jacobs et al., 1979; Deiss et al., 1986; Thomson et al., 1994a; Turner et al., 2002; Zimmermann & Hammerschmidt, 1995]. Formation of the precursor capsid structures occurs concurrent with DNA replication. Such that following rolling circle replication, the concatameric units are cleaved into unit length genomes and packaged into precursor capsids through their portal complexes; due to the action of virally encoded terminase proteins on highly conserved cleavage and packaging sequences (*pac1* and *pac2*) at the termini of the unit length genomes [Homa & Brown, 1997; Thomson et al., 1994a; Turner et al., 2002; Zimmermann & Hammerschmidt, 1995].

The current prevailing model suggests the nucleocapsids undergo an envelopment-development process in order to traverse the nuclear membrane. This is mediated by primary tegument proteins and non-structural proteins, resulting in deposition of the nucleocapsid in the cytoplasmic compartment [Mettenleiter, 2002]. This nucleocapsid then associates with remaining tegument proteins before a secondary envelopment process then occurs on endosomal/Golgi complex membranes. Utilisation of the host's vesicular transport and exocytic machinery then permits egress of the mature virion by budding at the plasma membrane [Mettenleiter, 2002; Torrisi et al., 1999; Mori et al., 2008].

1.1.6 General Characteristics of Latency

Productive lytic infection cycle, as described above, is not the only possible outcome of host cell infection. Importantly, the establishment of latency following the resolution of acute primary infection can also be considered a major defining feature of the *Herpesvirales*. This latent state refers to maintenance of the viral genome in the absence of infectious virion production, thereby allowing for life-long persistence of the virus in the host. The ability to reactivate from this latent state to the lytic replication cycle, being the prerequisite that defines latency from abortive infection. Throughout the life of the host this pool of latently infected cells permits sporadic instances of viral reactivation; allowing for the initiation of lytic replication from the latent genomes and thus production of

infectious virions, primarily based upon the immune status of the host. [Pellett & Roizman, 2013].

The establishment of latency is highly cell type specific, a feature which plays a key role in the classification of members of the *Herpesviridae*. However, for all subfamilies the primary mode of latency has long been believed to be the maintenance of the genome as a nuclear episome associated with cellular histones. This episomal structure displays a highly restricted pattern of gene expression maintained by both host and viral factors. Those genes which are expressed mainly involve factors contributing to maintenance of latency, immune evasion, and reactivation [Grinde, 2013].

However, recent findings with some members of the *Herpesviridae*, including HHV-6A and -6B, suggest that this central dogma may not be true for all herpesviruses [Delecluse & Hammerschmidt, 1993; Delecluse et al., 1993; Kaufer et al., 2011; Arbuckle et al., 2010; 2013]. Both HHV-6A and -6B can integrate into telomeric regions of the host cell chromosome and it has been proposed that this integration may also provide an alternative mechanism of latency [Arbuckle et al., 2010; 2013]. Uniquely amongst herpesviruses, in ~1% of populations studied to date, integration of HHV-6A and -6B has occurred in the germ-line [Pellett et al., 2012]. Not only is viral genomic DNA present in every nucleated cell of the body, the integrated form of the virus can be transmitted vertically between generations in a Mendelian manner, as originally shown by Daibata and colleagues [Daibata et al., 1998b; Daibata et al., 1999c]. These chromosomally integrated forms of HHV-6A and -6B are a major topic in the work presented here. Therefore extensive discussion of this viral integration will be discussed in its own section below (section 1.3).

1.2 Human Herpesvirus-6A & -6B

1.2.1 Introduction

In 1986, during the cultivation of peripheral blood mononuclear cells (PBMCs) from AIDS patients and other lymphoproliferative disorders, the existence of large, refractile mono- or binucleated cells often with nuclear and/or cytoplasmic inclusion bodies, was noted. Examination of these cells led to the isolation of a novel virus possessing herpesvirus-like morphology, initially named human B-lymphotropic virus (HBLV) [Salahuddin et al., 1986]. Subsequent studies showed it was predominantly a T lymphotropic herpesvirus, therefore the name of this virus was revised to Human Herpesvirus-6 (HHV-6) to reflect this and to adhere to the ICTV classification scheme used for other herpesviruses [Ablashi et al., 1987]. As such this represents the first recorded isolation of one of the HHV-6 species (now known as HHV-6A strain GS). This was closely followed by isolation of further strains, HHV-6A U1102 and AJ [Downing et al., 1987; Tedder et al., 1987] and HHV-6B strains Z29 and HST [Lopez et al., 1988; Yamanishi et al., 1988]. The first complete clones of the genome were then derived for HHV-6A strain U1102 [Martin et al., 1991a]. Followed by derivation of the complete genome sequence [Gompels et al., 1995]. This combined with better replicative properties in cell culture, led to the establishment of HHV-6A strain U1102 as the prototype for the Roseoloviruses.

Soon after the initial isolation of HHV-6, it became clear that notable differences in the restriction endonuclease cleavage patterns, the ability to replicate in certain T cell lines, reactivity to a number of monoclonal antibodies and nucleotide sequence, could be

observed among the HHV-6 strains isolated [Wyatt et al., 1990; Schirmer et al., 1991; Ablashi et al., 1991; Aubin et al., 1991]. These distinctions led to the classification of HHV-6 strains into two variant groups: A and B [Ablashi et al., 1993]. However, as data was accrued on the strains of HHV-6 falling into these two variant categories, the distinction between them with regards to epidemiology, disease associations and biological features became clearer. This in combination with a lack of data suggesting recombination between the variants implied the two groups occupy different ecological niches [Ablashi et al., 2014]. As such, despite strains in the two variants sharing in the region of 90% overall nucleotide sequence homology, the clear distinctions between these variant groups has meant that they have recently been reclassified as distinct viral species, HHV-6A and HHV-6B, by the ICTV [King et al., 2012; Ablashi et al., 2014]. The prototypic strains for each species are now strain U1102 for HHV-6A and strain Z29 for HHV-6B.

1.2.2 Genome

Currently, complete genome sequences for two HHV-6A strains, the 159 kbp U1102 strain and the 157 kbp GS strain, and also two HHV-6B strains, the 162 kbp Z29 strain and the 162 kbp HST strain, have been deposited in the GenBank sequence database (accession numbers X83413.1, KC465951.1, AF157706.1 and AB021506.1, respectively) [Gompels et al., 1995; Gravel et al., 2013a; Dominguez et al., 1999; Isegawa et al., 1999]. HHV-6A strain U1102 was isolated from reactivated infection in an adult HIV/AIDS patient from Uganda, while GS is from a mixture of American/Jamaican patients with lymphoproliferative disorders, including a HIV/AIDS patient [Downing et al., 1987; Salahuddin et al., 1986]. For HHV-6B, Z29 was isolated from a HIV/AIDS patient from Zaire (now Democratic Republic of the Congo) [Lopez et al., 1988], while strain HST was isolated from a Japanese infant during primary infection [Yamanishi et al., 1988], representing the only childhood primary infection strain sequenced to date.

A conserved genome architecture can be observed between HHV-6A and HHV-6B. This consists of a large ~144 kbp central unique region, which is interrupted by three repeat regions (R1-R3), as well as being flanked by two ~8 kbp direct repeats (DRs) located at the left (DR_L) and right (DR_R) termini of the genome. Predicted ORFs in the DRs are designated DR followed by an arabic number. Both DRs are flanked by herpesviral conserved genome cleavage and packaging signals (pac) as well as vertebrate telomeric repeat arrays which appear to play a role in integration of HHV-6A and -6B genomes into chromosomal DNA [Thomson et al., 1994a; Gompels & Macaulay 1995; Achour et al., 2009; Arbuckle et al., 2010]. Variation in the size of these DR and R1-R3 regions, is predominantly responsible for the size difference observed between the two species [Dominguez et al., 1999; Isegawa et al., 1999]. Differences in these repeat regions are also a major contributor to the sequence variation observed between the two species. The overall nucleotide sequence identity of HHV-6A and HHV-6B is in the region of 90% [Dominguez et al., 1999; Isegawa et al., 1999]. But further analyses shows the divergence includes these mismatched repetitive sequences as well as a number of hypervariable genes, such as those coding for the envelope glycoproteins, involved in immunomodulation and those important in the early stages of infection (U86-U95, excluding U94) [Gompels & Kasolo, 2006]. Therefore, much of the rest of the unique region of the genome displays a higher level of sequence conservation, averaging ~95%.

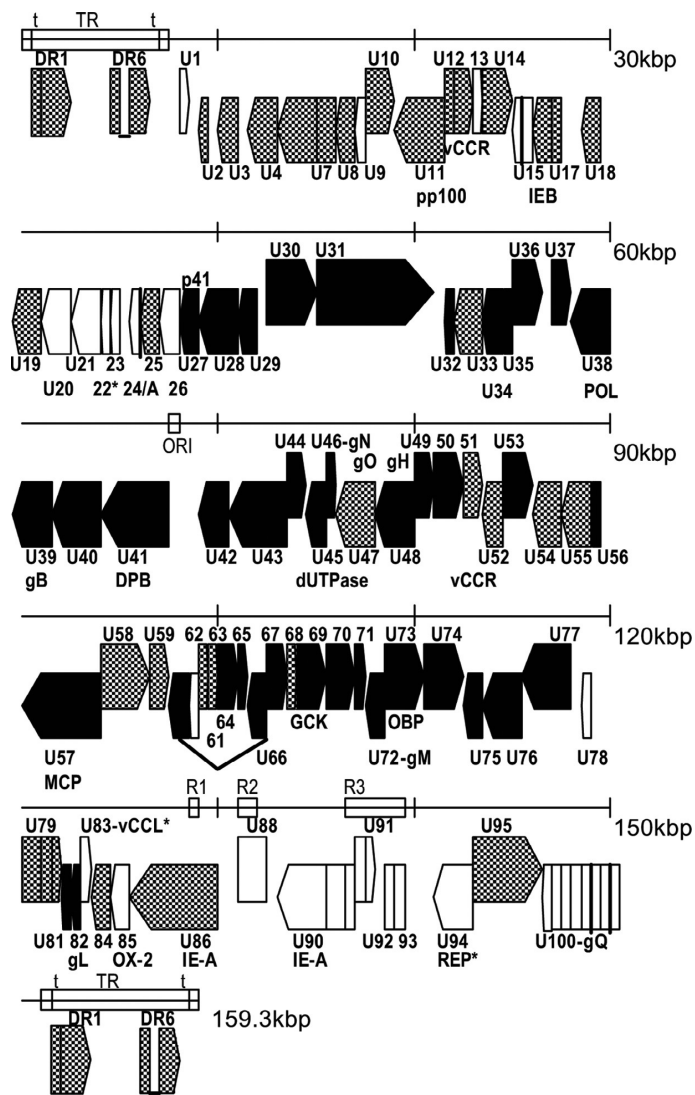


Figure 1.4 HHV-6A genome organisation. Human herpesvirus conserved genes (black), betaherpesvirus conserved genes (shaded), HHV-6A unique (white). Direct repeat regions are labelled TR and internal repeat regions are labelled R1-R3. [Figure from Gompels & Kasolo, 2006].

The central unique region of HHV-6A and HHV-6B contains at least 100 genes, ORFs designated U1 – U100, with a minority encoding multiple exons. HHV-6B has been suggested to contain nine additional ORFs (119 ORFs) compared to HHV-6A (110 ORFs), which have been termed B1 – B9. However, the protein coding status of a number of these ORFs has yet to be determined [Megaw et al., 1998; Dominguez et al., 1999]. Comparisons of related genomes suggests a verified list of 85 genes [Megaw et al., 1998]. The unique region displays some further general subdivisions. The left end of the genome contains predominantly betaherpesviral conserved genes, totalling around 27 genes. The central region contains 41 of the 43 core genes, that is those conserved among the *Herpesviridae* family, which are found in the characteristic seven gene blocks, in a betaherpesviral conserved arrangement [Gompels & Kasolo, 2006]. While the right end of the genome contains predominantly Roseolovirus and HHV-6A and -6B specific genes, figure 1.4.

Interestingly, HHV-6A and -6B appear to encode a number of genes which share some homology to cellular and other viral family genes. As such it seems gene acquisition, or possibly convergent evolution, have played an important role in HHV-6A and HHV-6B

evolution. A number of these genes are of note with regard to the topic of this work. U83 encodes a functional chemokine homologue, while U12 and U51 encode functional chemokine receptor homologues. All of which may contribute an immunomodulatory role to HHV-6A and -6B. U94 encodes a protein with homology to the Adeno-associated parvovirus type 2 (AAV-2) Rep 68/78 proteins and as such has been strongly implicated as playing a role in HHV-6A and -6B chromosomal integration events, discussed further in section 1.3.

1.2.3 Cellular & Tissue Tropism

Both HHV-6 species display a broad cellular tropism *in vitro*, presumed to be due to the ubiquitous expression of their primary cellular receptor, CD46; however, productive lytic infection of HHV-6A and -6B, both *in vitro* and *in vivo*, predominantly occurs in non-naive CD4+ T lymphocytes [Lusso et al., 1988; Takahashi et al., 1989; Grivel et al., 2003; De Bolle et al., 2005a]. *In vitro* differences in the ability of HHV-6A and -6B to infect certain T-cell lines and other cell types was a major basis for the classification of both viruses as separate species [Ablashi et al., 2014]. Notably, HHV-6A has a wider cellular tropism with productive infection in a range of cytotoxic effector cells: CD8+ T cells [Lusso et al., 1991; Grivel et al., 2003], natural killer cells [Lusso et al., 1993], and gamma/delta T cells [Lusso et al., 1995]. In addition HHV-6A also appears to replicate more efficiently in a variety of neuronal cells, potentially providing a basis for increased neuropathology of HHV-6A: neural stem cells [De Filippis et al., 2006], astrocytes [Donati et al., 2005] and oligodendrocytes [Ahlqvist et al., 2005].

A number of factors have been proposed to contribute to the differential cell tropism observed between HHV-6A and -6B. Variation in the affinity of HHV-6A and -6B for CD46 has been noted, with HHV-6B displaying a lower affinity than HHV-6A [Santoro et al., 1999] and the potential utilisation of additional or alternative cellular receptors, exemplified by the recent identification of CD134 as a specific receptor for HHV-6B [Tang et al., 2013], discussed further in section 1.2.4. Closely linked to this is the large sequence variation observed between a number of the HHV-6A and HHV-6B envelope glycoprotein homologues, gO and gQ [Mori, 2009a]. While, the sequence variation in the viral chemokine, U83, between the species and the subsequent ability to chemoattract different cell populations has also been implicated in determining cell tropism [Luttichau et al., 2003; Dewin et al., 2006; Catusse et al., 2007; 2009; Clark et al., 2013].

Latent infection has been demonstrated to occur within bone marrow progenitor cells [Luppi et al., 1999]; monocytes/macrophages [Kondo et al., 1991]; as well as reports in myeloid, astrocytoma and oligodendrocyte cell lines [Yasukawa et al., 1999; Yoshikawa et al., 2002; Ahlqvist et al., 2005]. However, again the greater appreciation of the differential target cells of these viral species, means the definition of the true latency reservoirs of these viral species requires further clarification. In addition, the recognition of germ-line chromosomally inherited HHV-6A and -6B (CI-HHV-6A/-6B), discussed further in section 1.3, could be considered to cloud this distinction in a subset of the population, depending on the reactivation capabilities in certain cell types.

HHV-6A and -6B DNA can also frequently be detected in an array of both healthy and pathological human tissues [Di Luca et al., 1996; De Bolle et al., 2005b]. Of note, the salivary glands and adenoid tissue have been identified as a site of viral replication and persistence for HHV-6B [Jarrett et al., 1990; Fox et al., 1990; Di Luca et al., 1995; Roush et

al., 2001; Sato et al., 2009], which is consistent with the major route of transmission of HHV-6B being salivary secretions from close contacts. In addition, while both HHV-6A and HHV-6B are considered neurotropic viruses, evidence suggests HHV-6A replicates more readily in the CNS and is associated with increased neuropathology [Hall et al., 1998; Chan et al., 2001; Ablashi et al., 2014].

1.2.4 Replication Cycle & Latency

In many respects the replication cycles of HHV-6A and -6B follow the generalised Herpesviral cycle outlined in section 1.1.5 and 1.1.6. However, an exception to this may arise in the possible mode for the establishment and maintenance of latency. Recent evidence has suggested integration of the viral genome into the host cell chromosome may provide an alternative mechanism for latency in HHV-6A and -6B, instead of nuclear episomal maintenance as predominantly observed in the rest of the *Herpesvirales* (discussed in detail in section 1.3). The following will therefore focus on supplementing the discussion of the replication cycle previously outlined with features specific to HHV-6A and -6B.

Initial reports identified the cellular receptor for both HHV-6A and HHV-6B as the human complement regulator protein CD46, also known as membrane cofactor protein (MCP) [Santoro et al., 1999]. CD46 is a type I transmembrane glycoprotein primarily noted as an inhibitor in the complement system, but which has also been found to possess important roles in T cell regulation and fusion during fertilisation. It has also been identified as a receptor for a diverse array of human viral and bacterial pathogens, including measles virus, a number of adenoviruses, *Streptococcus pyogenes*, *Neisseria gonorrhoeae*, and *Neisseria meningitidis* [Cardone et al., 2011]. Structurally it consists of four short consensus repeats (SCR1 – 4) at the N-terminus, followed by a serine, threonine and proline rich domain, a transmembrane domain and a short C-terminal cytoplasmic domain [Liszewski et al., 1991]. It is expressed on all nucleated cells, which provides some insight into the relatively broad *in vitro* human cell tropism of HHV-6A and -6B. In the initial reports of the HHV-6 cellular receptor, an anti-CD46 mAb as well as a soluble form of CD46 were shown to inhibit HHV-6A and -6B-mediated cell fusion (HHV-6A strain GS and HHV-6B strains Z29 and PL1). In addition, gain of function was demonstrated through the expression of human CD46 rendering non-human cells susceptible to HHV-6A and -6B entry [Santoro et al., 1999]. Subsequently, the viral ligand important for HHV-6A and -6B fusion and entry was found to be a glycoprotein complex consisting of the gH/gL/gQ1/gQ2 glycoproteins (U48, U82 and U100), gH being primarily responsible for CD46 binding [Santoro et al., 2003; Mori et al., 2003b; Tang et al., 2011; Kawabata et al., 2011; Tanaka et al., 2013; Jasirwan et al., 2014]. This complex was found to be interacting with at least the SCR2 and SCR3 domains, and possibly also the SCR4 domain, of CD46 in the case of HHV-6A [Mori et al., 2002; Greenstone et al., 2002; Santoro et al., 2003].

However, HHV-6A and -6B may utilise additional receptors, or especially with regard to HHV-6B, utilise an alternative receptor(s). Firstly, for the HHV-6B strain HST the gH/gL/gQ1/gQ2 was found not to bind CD46 [Mori et al., 2004] and did not mediate fusion from without in cells expressing human CD46 [Mori et al., 2002]. Secondly, variability in the ability of HHV-6A strain GS to infect a panel of CD4+ T-cell lines was observed, even with the transfection of CD46 into the cell lines to account for different isoforms of CD46 [Santoro et al., 1999].

Thirdly, both HHV-6A and -6B have been found to possess another viral envelope glycoprotein complex, gH/gL/gO (U48, U82 and U47), also containing the *Herpesviridae* conserved glycoproteins gH and gL [Mori et al., 2004]. Since these glycoproteins have been implicated as participants in herpesvirus cell fusion and entry, a number of attempts are currently being made to characterise the role of this gH/gL/gO glycoprotein complex [Paterson et al., 2002]. Also of note is that the products of the gO gene from HHV-6A and -6B only share ~77% amino acid sequence identity, much lower than that found amongst the other envelope glycoproteins. Therefore interactions of the gH/gL/gO complex have been postulated to be involved in the different cell tropisms observed for the two viral species [Mori et al., 2004]. Indeed, recently it has been suggested that at least HHV-6B strains HST and KYO utilise CD134, also known as OX40, as the cellular receptor [Tang et al., 2013]. CD134 is a member of the tumour necrosis factor receptor superfamily which is preferentially expressed on the surface of activated CD4+ T cells, but also present on activated CD8+ T cells and other T cell subsets at lower levels [Croft, 2010]. Interestingly, CD134 is rarely present on glial cells [Croft, 2010], which may play a role in the increased neurotropic capacity of HHV-6A over that of HHV-6B, which has been observed.

Another envelope glycoprotein, gB, also appears to have a role in the entry process [Takeda et al., 1996; Mori et al., 2002]. While the mechanism of action for gB in HHV-6A and -6B currently remains undetermined, identification of its role in other herpesviruses combined with high conservation of the gene among the *Herpesviridae* suggests it likely plays a role in the membrane fusion event [Pereira, 1994; Isaacson & Compton, 2009].

Entry results in release of the capsid and associated tegument proteins in the cytoplasm of the host cell. Cytoplasmic transport of the capsid and nuclear entry has received little attention for HHV-6A/-6B. But it is presumed to undertake a mechanism similar to that of HCMV which utilises the hosts microtubule network to reach the nucleus and entry of the viral genome via nuclear pore complexes [Ogawa-Goto et al., 2003; Dohner & Sodeik, 2005]. Subsequently lytic replication proceeds almost immediately following viral entry via 'cascade' gene expression, DNA replication and capsid packaging and assembly, as outlined in section 1.1.5. There are however, notable differences in the egress pathway. Unlike the other herpesviruses which have received extensive study, HHV-6A/-6B glycoproteins were found to be undetectable at the plasma membrane in lytically infected cells [Cirone et al., 1994]. Instead it was found they accumulate predominantly in lipid rafts in ER/Golgi apparatus membranes [Cardinali et al., 1998; Kawabata et al., 2009], as such it is believed that association with the viral envelope glycoproteins occurs during a secondary envelopment on these structures prior to egress via the host's vesicular transport system [Torrise et al., 1999].

Lytic replication of HHV-6A and -6B has profound effects on host cells. The most readily identifiable of these are the cytopathic effect of multinucleated, ballooning cells and apoptosis of cells [Inoue et al., 1997; Yasukawa et al., 1998; Ichimi et al., 1999]. But lytic infection also produces notable effects on the production of cytokines, upregulation of IFN- α [Kikuta et al., 1990], IL-1 β [Flamand et al., 1991], TNF- α [Flamand et al., 1991] and IL-15 [Flamand et al., 1996], downregulation of IL-2 [Flamand et al., 1995] and IL-12 [Smith et al., 2003] and modulation of CCL5 production [Milne et al., 2000; Caruso et al., 2003], which have been linked to roles in both immunomodulation and viral infection control [Takahashi et al., 1992; Smith et al., 2003; Milne et al., 2000; Caruso et al., 2003]. As well as affecting the expression of cell surface markers, most notably the induction of CD4 expression permitting infection of cells by HIV-1 [Lusso et al., 1991; Furukawa et al., 1994].

Following initial infection, lytic replication and seroconversion, the persistence of HHV-6A and -6B has long been thought to occur either through entering a true latent state due to the formation of circular genomic episomes as observed among other herpesviruses or possibly through low level replication at distinct sites. However, recent findings have led to the suggestion that HHV-6A and -6B may in fact utilise an alternative or additional mechanism for latency, that of chromosomal integration [Arbuckle et al., 2012, 2013]. Since the integration of HHV-6A and -6B is a major theme of the work presented in this thesis, extensive discussion of this feature has been moved to it's own section below (section 1.3).

Latency associated transcripts and latency associated gene expression have been detected for HHV-6A and HHV-6B. The U94 gene product was found to be expressed at higher levels during latency. Stable expression of U94 in a lymphoid cell line permitted HHV-6 infection but restricted viral replication, and as such a role in the establishment and/or maintenance of latency was proposed [Rotola et al., 1998; Caselli et al., 2006]. While, due to their expression at higher levels prior to reactivation four latency associated transcripts located in the IE1/IE2 gene region, have been implicated in the promotion of reactivation [Kondo et al., 2002; Kondo et al., 2003]. Expansion of these IE1/IE2 latency associated transcripts, and subsequent reactivation, was associated with cellular stimuli [Kondo et al., 2003]. But superinfection with HHV-7 has also been shown to reactivate latent HHV-6B, presumed to be by a transactivating function [Katsafanas et al., 1996; Tanaka-Taya et al., 2000].

1.2.5 Transmission and Epidemiology

In those populations studied to date primary infection with a HHV-6 species, most commonly HHV-6B, occurs within the first two years of life [Okuno et al., 1989; Enders et al., 1990; Zerr et al., 2005a] coinciding with a decline in the levels of maternal antibodies, which are believed to confer at least some level of protection against infection [Farr et al., 1990; Hall et al., 1994]. Despite some variance in the estimates, possibly in part due to geographic variation but also likely due to the different methodologies employed, the overall prevalence of HHV-6A and -6B infection is now believed to be >90%, but there appear to be geographic differences [Enders et al., 1990; Di Luca et al., 1994; Zerr et al., 2005a].

In most populations, HHV-6B has been identified as the initial HHV-6 species acquired and predominantly responsible for fever and exanthem subitum disease associated with primary infection [Yamanishi et al., 1988; Dewhurst et al., 1993; De Bolle et al., 2005a]. Primary infection with HHV-6A is rare, but symptomatic cases have been noted [Ward et al., 2005; Hall et al., 2006]. Instead HHV-6A is thought to occur asymptotically later in life [De Bolle et al., 2005a]. Co-infection with multiple HHV-6B strains, as well as HHV-6A and HHV-6B together has also been observed in adult populations [Cone et al., 1996]. It has been shown that there may be geographic differences in primary infection and species prevalence. Studies in Zambia show HHV-6A is equally prevalent with HHV-6B in symptomatic febrile infant infection and predominant in a population study of asymptomatic children [Kasolo et al., 1997; Bates et al., 2009]. Global seroprevalance studies will require specific reagents to further explore relative distributions, but these are not yet available since immunodominant proteins in the two HHV-6 species are cross-reactive.

As highlighted, primary infection with either HHV-6A or -6B primarily occurs due to horizontal transmission in the first years of life. Indicative of transmission from close contacts during this period. The frequent detection of HHV-6A and/or -6B DNA in saliva [Di Luca et al., 1995; Tanaka-Taya et al., 1996; Zhao et al., 1997] has led to suggestions that asymptomatic shedding of HHV-6A and -6B in the saliva of close family members likely represents the main route of transmission. However, while HHV-6B is frequently detected in saliva/salivary gland tissue, detection of HHV-6A has been limited [Collot et al., 2002]. Therefore, alternative routes of transmission may exist. Recent demonstration of the presence of HHV-6A and -6B DNA in the nasal cavity, has led to suggestions that this site represents an *in vivo* reservoir which could permit host to host transmission [Harberts et al., 2011]. Detection of HHV-6A and -6B in the cervix [Leach et al., 1994; Okuno et al., 1995] and stool samples [Suga et al., 1998] have been used to implicate the possibility of perinatal and faecal-oral transmission routes, however, evidence for such routes remains limited.

Congenital HHV-6A and -6B infections can also occur, albeit with a much rarer occurrence of approximately 1% of the population, similar to that of HCMV. This can be due to transplacental passage of maternal HHV-6A or -6B, however, recent findings suggest this detection is most frequently from the inheritance of chromosomally integrated form of these viral species, discussed in section 1.3 [Hall et al., 2008; Hall et al., 2010].

1.2.6 Primary Infection, Reactivation & Reinfection

Primary infection with one of the HHV-6 species in infants is associated with an acute febrile illness, lasting 3-7 days [Asano et al., 1994; Hall et al., 1994]. In a subset of patients (~25%), the initial high temperature fever, is followed by the development of a maculopapular rash, clinically referred to as exanthem subitum/roseola infantum/sixth disease, initially on the torso but later spreading to the extremities [Asano et al., 1994]. In the majority of populations studied, HHV-6B has been found to be the causative agent of symptomatic primary infections [Yamanishi et al., 1988; Dewhurst et al., 1993; De Bolle et al., 2005a], however, the opposite has been found to be true in a Sub-Saharan African population [Kasolo et al., 1997; Bates et al., 2009]. As highlighted by epidemiological studies, primary infection in immunocompetent adults is a rare occurrence, but where studied there has been suggestions of an increased severity compared to the infant infection, with presentations of extended lymphadenopathy and mononucleosis-like disease noted [Niederman et al., 1988; Akashi et al., 1993]. In immunocompetent individuals these primary infections are generally regarded as benign and self-limiting. However, in rare cases, and also more frequently amongst immunosuppressed individuals, complicated primary infections may occur associated with much more severe disease manifestations, such as liver dysfunction/hepatitis, myocarditis, thrombocytopenia, gastrointestinal symptoms and central nervous system complications [De Bolle et al., 2005a].

As with all members of the *Herpesviridae*, primary infection with HHV-6A/-6B results in life-long persistence of the virus. The virus remains in a latent state, with the ability to reactivate periodically. This reactivation is primarily associated with immunosuppression and again in immunocompetent individuals is most frequently asymptomatic and self-limiting [Caserta et al., 2004]. However, in cases of prolonged immunosuppression such as seen in transplant recipients and HIV positive individuals, symptomatic reactivation occurs

at high frequency, often associated with severe disease manifestations [Clark & Griffiths, 2003]. In addition, following primary infection, reinfection with further HHV-6A/-6B strains and associated disease manifestations remains a possibility, however, to date the effects of these secondary acquisitions remains poorly defined [Agut, 2011].

1.2.7 Diagnosis

As in much of virology, the gold standard for diagnosis for active infection by HHV-6A/-6B, be that primary, reactivated or re-infection, remains virus isolation and culture. However, this method is both time and resource intensive, as well as being less useful in the detection of latent and chromosomally integrated infections. As such methods based on serology, antigen detection and DNA/RNA detection tend to be more readily employed, yet due to the similarity between HHV-6A and -6B these assays possess a number of limitations.

Serologic methods that have been employed include indirect immunofluorescence assays (IFA), enzyme-linked immunosorbent assays (ELISA) and antibody avidity assays. However, these methods provide difficulties in the diagnosis of reactivations and have the potential for cross-reactivity with other betaherpesviruses [Nakagawa et al., 1997]. But most importantly, due to the high level of similarity observed between the antigens in HHV-6A and -6B, all of these serological assays currently lack of ability to discriminate between HHV-6A and HHV-6B [Flamand et al., 2010].

Detection of viral DNA/RNA by means of quantitative (qPCR) and/or reverse transcriptase (RT-PCR) has provided highly sensitive methods which also allow for the typing of the species and genotype strains. However, in the case of qPCR issues still remain in the discrimination of active infection from CI-HHV-6A/-6B, as there can be considerable overlap between the viral loads obtained from samples in each of these cases [Flamand et al., 2010; Caserta et al., 2010]. Such issues have been shown to lead to unnecessary administration of antiviral therapy [Lee et al., 2012] and also have the potential for causing delays in the diagnosis/intervention of active HHV-6A or -6B infection. Accurate diagnosis of CI-HHV-6A/-6B currently relies on the use of fluorescence in situ hybridisation (FISH) analysis or qPCR on hair follicle samples [Pellet et al., 2012]. However, FISH analysis where available, is time intensive and hair follicle samples are not routinely obtained. The recent development of a CI-HHV-6A/-6B specific droplet digital PCR assay may provide a means of rapid diagnosis from readily obtained and stored cellular samples [Sedlak & Jerome, 2013; Sedlak et al., 2014]. Indeed recent evidence suggests this third generation PCR technology allows for accurate discrimination between the HHV-6 species, as well as more accurate detection of co-infection [Leibovitch et al., 2014].

1.2.8 Disease Associations

Due to the latent nature of the HHV-6A/-6B and propensity to reactivate under times of immunosuppression, the ability to demonstrate causality of a disease represents a difficult task. In addition to the now well established clinical symptoms of primary infections described above, numerous disease associations have been attributed to infections with HHV-6A and/or -6B. The evidence for some of these provide compelling arguments for the involvement of the HHV-6 species, while for others the associations still remain controversial.

Both HHV-6A and HHV-6B have been demonstrated to infect a variety of neuronal cells *in vitro*, see section 1.2.3, where they may establish persistent infections. In addition, DNA and mRNA transcripts from both viruses can also be readily detected in the brain tissue and cerebrospinal fluid (CSF) of healthy immunocompetent adults [Chan et al., 2001; Luppi et al., 1998a; Cuomo et al., 2001; Opsahl & Kennedy, 2005]. Therefore, it seems clear that the central nervous system (CNS) represents a site of replication and latency for both HHV-6 species. This complicates attempts to definitively attribute neurological pathologies to these viruses and has led to much debate about involvement of the HHV-6 species in such diseases. However, it also clearly demonstrates the neuroinvasive properties of these viruses, which in addition to its noted proinflammatory properties provide a potential mechanism for neuropathology. This in combination with a large number of association studies and the recent demonstration of HHV-6A as a trigger for neurological disease in a marmoset (*Callithrix jacchus*) model; and HHV-6A also responsible for viral persistence and production of proinflammatory chemokines via TLR9 signalling in the brain tissue of CD46 transgenic mice [Leibovitch et al., 2013; Reynaud et al., 2014], strongly suggest a role of HHV-6A and HHV-6B in at least some of these pathologies. However, there do seem to be some differences between the two species and their neurotropism and pathological outcomes. HHV-6A possesses a greater propensity to replicate and cause cytopathic effects in a number of the neuronal cell types [Ahlqvist et al., 2005; Donati et al., 2005; De Filippis et al., 2006;]. Also, HHV-6B has been detected more frequently in the brain tissue, likely as a consequence of its higher prevalence in the populations sampled [Chan et al., 2001], however, association studies and the recent animal models have implicated a greater role for HHV-6A in neuropathology [Hall et al., 1998; Akhyani et al., 2000; Soldan et al., 2000; Leibovitch et al., 2013; Reynaud et al., 2014].

The most well defined neurological disease associations are those of febrile seizures, encephalopathy. Similarly, strong links have been established between HHV-6A/-6B reactivation in immunocompromised populations and these neurological complications [Zerr et al., 2001]. Links between HHV-6A/-6B and encephalopathy in immunocompetent adults remain difficult to prove. However, HHV-6A/-6B has been detected in cerebrospinal fluid samples from immunocompetent patients suffering encephalitis and there have been case reports of the successful treatment of such patients with antiherpesviral drugs [Isaacson et al., 2005; Birnbaum et al., 2005].

Infections with HHV-6 viral species also have some less well defined links to a number of other neurological pathologies. The most notable being that due to the proinflammatory properties of HHV-6A and -6B, these viruses have frequently been cited as a causative agent or progression factor in the inflammatory neurological disease multiple sclerosis (MS). These links have been determined by a variety of methods. The more frequent amplification of HHV-6A/-6B DNA and immunohistochemical detection of HHV-6A/-6B antigens in the brain tissue and especially the MS lesion samples of MS patients [Wilborn et al., 1994; Challoner et al., 1995; Chapenko et al., 2003; Goodman et al., 2003]. Increased detection of both HHV-6A/-6B DNA and specific antibody titres in the sera and CSF of MS patients [Soldan et al., 1997; Chapenko et al., 2003; Ablashi et al., 1998a]. Increased lymphoproliferative responses to HHV-6A antigens in MS patients [Soldan et al., 2000]. The increased detection of HHV-6A/-6B DNA and HHV-6A/-6B specific antibody titres in relapsing remitting MS patients, undergoing relapse [Alvarez-Lafuente et al., 2004; Simpson et al., 2012]. While potential links to MS can also be made from the observation

that HHV-6A infection in the marmoset animal model caused myelin abnormalities in some of the marmosets that displayed both clinical symptoms and anti-HHV-6A antibody responses [Leibovitch et al., 2013]. So while there appear to be strong links between MS and HHV-6A/-6B infection, the true role of the involvement of the HHV-6 species, be it causative agent, progression factor or consequence of MS, remain to be determined.

Transplantation Complications

HHV-6A/-6B infection is detected with high frequency, incidence rate estimated to be 30-50%, in the setting of transplantation, with HHV-6B accounting for the majority of these infections [Ljungman & Singh, 2006; De Bolle et al., 2005a; Zerr, 2012; Drobyski et al., 1993; Wang et al., 1999; Ogata et al., 2013; Inazawa et al., 2015]. Generally occurring within a month post transplantation, this infection is often asymptomatic or produces the fever and rash manifestations observed during primary infection [Griffiths et al., 1999; Ablashi et al., 2010]. However, complications are reported with higher frequency than seen in immunocompetent hosts, and more often the complications are associated with fatality [Ablashi et al., 2010; Ohashi et al., 2008; Ogata et al., 2013; Inazawa et al., 2015]. These infections are thought to predominantly arise from the reactivation of latent endogenous virus as a result of the prolonged immunosuppression induced through the utilisation of immunosuppressive drugs to maximise the success of transplantation. However, primary infection or reinfection with a secondary strain due to transfer of virus from host to recipient during transplantation remain a possibility [De Bolle et al., 2005a].

In haematopoietic stem cell transplant recipients, encephalopathy is the most frequently reported complication and the available reports include both cases of fatalities and successful treatment with anti-herpesviral drugs [Drobyski et al., 1994; Rieux et al., 1998; Yoshida et al., 2002; MacLean & Douen, 2002; Ogata et al., 2013; Inazawa et al., 2015]. Delayed engraftment or graft rejection are also often reported [Rosenfeld et al., 1995; Ljungman et al., 2000; Zerr et al., 2005b; Dulery et al., 2012]. While in solid organ transplant recipients, disease similar to acute primary infection is most frequently observed [Griffiths et al., 1999; Humar et al., 2002]. However, encephalopathy [Montejo et al., 2002; Paterson et al., 1999; Ljungman & Singh, 2006], hepatitis [Griffiths et al., 1999] and graft dysfunction/rejection [Acott et al., 1996; Tong et al., 2002] are also reported complications.

HIV/AIDS Interactions

Since its original isolation from an AIDS patient [Salahuddin et al., 1986], HHV-6A/-6B has long been associated with HIV-1 infection. Indeed HHV-6A/-6B infection is frequently detected in AIDS patients [Knox & Carrigan, 1994; Secchiero et al., 1995; Ablashi et al., 1998b]. In addition, HHV-6A, HHV-6B and HIV-1 all have a tropism for lytic replication in CD4+ T-cells and productive co-infections with accelerated cytopathic effects have been observed *in vitro* [Lusso et al., 1989]. As such it has been proposed that while HHV-6A/-6B active infection may be a consequence of the immunosuppression induced by HIV-1 infection, it could also play a role in the progression of HIV-1 to AIDS.

A number of lines of evidence point towards interplay between HHV-6A/-6B and HIV-1. Some of these have suggested a negative impact of HHV-6A/-6B infection on HIV-1. Firstly, the HHV-6A viral chemokine, U83, has been demonstrated to be a high affinity ligand for the important HIV-1 co-receptor CCR5. Its action at CCR5 inhibits the infection of CCR5

expressing cells by CCR5 tropic HIV-1 at nanomolar concentrations [Catusse et al., 2007]. Also, early in the lytic cycle, HHV-6A infection has also been found to upregulate the production of CCL5 [Grivel et al., 2001]. CCL5 binds CCR5 with high affinity, albeit a lower affinity than U83A, and has been demonstrated to be a potent inhibitor of CCR5 tropic HIV-1 [Cocchi et al., 1995], thereby providing another mechanism for the inhibition of HIV-1 replication. However, as to be outlined shortly, these features may in fact promote the evolution of HIV-1 strains towards CCL5 resistance or alternative co-receptor usage, therefore actually aiding in the progression to AIDS.

A greater body of evidence supports a role of HHV-6A/-6B infection in the the progression to AIDS. Firstly, infection of cells with HHV-6A/-6B leads to transactivation of the CD4 promoter by a mechanism involving the immediate early genes U86 and U89 [Flamand et al., 1998]. Not only does this infection upregulate CD4 expression on CD4+ cells potentially making them more susceptible to HIV-1 infection, it also induces the surface expression of CD4 on normally CD4- cells (CD8+ T-cells, $\gamma\delta$ T-cells and NK cells), thereby rendering them susceptible to HIV-1 infection [Lusso et al., 1991; 1993; 1995]. HHV-6A also has the ability to transactivate the HIV-1 LTR promoter resulting in stimulation of HIV-1 gene expression [Ensoli et al., 1989; Horvat et al., 1989; McCarthy et al., 1998]. This transactivation is via a region which is distinct from that required for HIV-1's own transactivating protein Tat [Ensoli et al., 1989]. There is evidence of a synergistic interaction of HHV-6A/-6B proteins with Tat, as well as some suggestion that Tat may in fact reciprocally enhance HHV-6A/-6B gene expression [Di Luca et al., 1991; Garzino-Demo et al., 1996; Sieczkowski et al., 1995]. HHV-6A infection also increased the production of a number of proinflammatory cytokines, notably TNF- α and IL-1 β [Flamand et al., 1991], which have been implicated in the activation of HIV-1 gene expression [Duh et al., 1989; Swingler et al., 1992; Copeland, 2005]. While in a pig-tailed macaque (*Macaca nemestrina*) model co-infection with HHV-6A and simian immunodeficiency virus (SIV) accelerated the progression of disease through the depletion of CD4+ T-cells [Lusso et al., 2007]. Thereby providing the first *in vivo* links of HHV-6A/-6B to AIDS progression. Further work with this macaque model has demonstrated that SIV isolates obtained from HHV-6A and SIV co-infected animals one year post infection had developed resistance towards CCL5 inhibition, which was not seen in singly SIV infected animals [Biancotto et al., 2009]. Thereby implicating the evolution of the SIV strains to CCL5 resistance, as a result of the HHV-6A mediated generation of high levels of CCL5 and potentially providing an explanation for the rapid progression of SIV disease.

Myocarditis & Cardiomyopathy

A viral agent is often cited in myocarditis and cardiomyopathy, and parvovirus B19 and HHV-6A/-6B are the three most frequently found viruses in the heart muscle [Bultmann et al., 2005; Kuhl et al., 2005]. This link is extended by a number of case reports of both HHV-6A and HHV-6B associated myocarditis and even the recent successful treatment of myocarditis in an infant, with the promising anti-herpesviral drug candidate artesunate [Bigalke et al., 2007; Chang et al., 2009; Leveque et al., 2011; Ashrafpoor et al., 2013; Hakacova et al., 2013].

1.2.9 Therapy

In immunocompetent hosts, HHV-6A/-6B primary infection, reactivation or reinfection represent self-limiting infections and therefore does not require treatment. However, where individuals present with some of the more severe disease manifestations outlined above, intervention strategies are attempted. Currently there are no formally approved vaccines or drugs for the treatment of HHV-6A or -6B, therefore intervention involves the use of drugs originally developed as anti-HCMV compounds (or anti-HSV in the case of acyclovir/valaciclovir) which have subsequently been demonstrated to have some efficacy against HHV-6A and -6B [Agut et al., 1989; Manichanh et al., 2000; De Clercq et al., 2001; De Clercq & Naesens, 2006]. All of these drugs specifically inhibit the catalytic site of the viral DNA polymerase that (with the exception of cidofovir) result in premature chain termination, with acyclovir/valaciclovir, ganciclovir/valganciclovir and cidofovir acting as nucleotide/nucleoside analogues which are activated by viral kinases and foscarnet directly binding the DNA polymerase to bring about these effects [Wagstaff & Bryson, 1994; Reardon & Spector, 1989; De Clercq, 2003]. In an *in vitro* analysis cidofovir was the most potent in cord blood lymphocytes, while in the T-cell lines HSB-2 and MOLT-3, foscarnet was the most potent [De Bolle et al., 2004]. This study also demonstrated that generally, all of the drugs are more potent against HHV-6A than HHV-6B. Ganciclovir/valganciclovir appear to be the most readily utilised drugs, resulting in successful resolutions, in case reports [Flamand et al., 2010].

However, the efficacy of these drugs has yet to be assessed in controlled clinical trials, they frequently have strong adverse effects associated with their use, and the potential prolonged use required to control HHV-6A/-6B infection in immunocompromised settings, opens up the possibility of the development of resistance [De Clercq & Naesens, 2006]. As such there remains a search for new anti-HHV-6 drugs, especially those which do not target the viral DNA polymerase. From this search two non-nucleoside analogues have emerged as promising candidates. Firstly, CMV422, presumed to target a cellular protein tyrosine kinase, has been shown to exhibit a high level of activity and selectivity for HCMV and HHV-6A/-6B [Snoeck et al., 2002; De Bolle et al., 2004]. While artesunate, a derivative of artemisinin often used in the treatment of *Plasmodium falciparum* malaria, has also proven to have broad activity against herpesviruses, including HHV-6A/-6B [Efferth et al., 2008; Milbradt et al., 2009] and has recently been successfully utilised in a clinical setting, for suspected HHV-6B myocarditis [Hakacova et al., 2013].

1.3 Chromosomal Integration of Human Herpesvirus-6A & -6B

1.3.1 Inherited Chromosomally Integrated HHV-6A and -6B

In 1993, analysis of uncultured peripheral blood mononuclear cells (PBMCs) from three patients, two suffering lymphoproliferative disorders and the third multiple sclerosis (MS), led to the discovery of a high copy number of HHV-6 specific sequences present in the samples. Following up this finding, the discovery of higher than expected molecular weights for the viral DNA during restriction analysis of the patients PMBCs indicated that the viral genomic sequences were linked to cellular DNA [Luppi et al., 1993]. These findings provided the first indication that HHV-6A/-6B DNA may integrate into the human genome. Subsequently, Gardella gel, fluorescent in situ hybridisation (FISH), and PCR analysis was used in a number of different studies to confirm chromosomal integration of

viral genomic DNA (or the high DNA copy number indicative of integration) in a number of different tissue types including hair follicles, brain, lymph nodes and cardiac tissues [Daibata et al., 1998a; Morris et al., 1999; Ward et al., 2006; Nacheva et al., 2008; Hall et al., 2008; Strenger et al., 2010; Arbuckle et al., 2010]. It has now been demonstrated that in these individuals HHV-6A and/or HHV-6B genomic DNA has been inherited through an integration event(s) into the chromosome of germ-line cells, hence the apparent presence of CI-HHV-6A/-6B in every nucleated cell of the body. Therefore, CI-HHV-6A/-6B can be transmitted vertically between generations in a Mendelian manner, with integration at the same chromosomal sites in parent and child [Daibata et al., 1998b; Hall et al., 2008; Morissette & Flamand, 2010; Arbuckle et al., 2010]. As such, this inherited CI-HHV-6A/-6B represents a unique mechanism among human herpesviruses for the vertical transmission of HHV-6 genomic DNA, in addition to the horizontal transmission through viral shedding in saliva which is believed to be the major route of acquiring HHV-6A/-6B. The prevalence of this CI-HHV-6A/-6B in the world's population has been estimated to be in the region of 0.8 – 1% [Tanaka-Taya et al., 2004; Leong et al., 2007; Hubacek et al., 2009; Potenza et al., 2009].

Even prior to the acquisition of the complete genome sequences for strains of HHV-6A and -6B, it was noted that human telomeric repeat sequences were present in the genomes of both species, close to the termini [Thomson et al., 1994a; Gompels & Macaulay 1995]. Within each of the direct repeats (DR) located at the left (DR_L) and right (DR_R) termini of the genome is a left terminal (pac1 adjacent) imperfect human telomeric repeat array and a right terminal (pac2 adjacent) perfect repeat array, each with variable copy numbers ranging from 15-180 repeats between different isolates which have been examined [Thomson et al., 1994a; Gompels & Macaulay 1995; Achour et al., 2009; Arbuckle et al., 2010]. Soon after the initial indications that HHV-6A/-6B could integrate into human chromosomal DNA, it was hypothesised that these telomeric repeat sequences may be involved in the integration event [Luppi et al., 1998b]. Support for this gathered as mapping of the integration sites demonstrated they were all localised to telomeric regions (1q44 [Watanabe et al., 2008], 9q34.3 [Nacheva et al., 2008; Ward et al., 2006], 10q26.3 [Nacheva et al., 2008; Huang et al., 2013], 11p15.5 [Nacheva et al., 2005; Clark et al., 2006b; Ward et al., 2006], 17p13.3 [Nacheva et al., 2008; Ward et al., 2006; Clark et al., 2006b; Arbuckle et al., 2010; Luppi et al., 1993; Morris et al., 1999; Torelli et al., 1995], 18p11.3 [Hubacek et al., 2009], 18q23 [Arbuckle et al., 2010], 19q13.4 [Nacheva et al., 2008; Ward et al., 2006] and 22q13.3 [Arbuckle et al., 2010; Daibata et al., 1999b]). Anecdotally, it seems there may be some preference for integration into the telomeric region of chromosome 17p13.3, however, further study is needed to confirm this, this is also one of only two sites for which integration has been observed for both HHV-6A and HHV-6B [Morris et al., 1999; Torelli et al., 1995; Morissette & Flamand, 2010]. This telomeric integration was confirmed through the *in vitro* demonstration that single copy full length genomic DNA of both HHV-6A and -6B can integrate into the telomeres of human chromosomes at a location close to the subtelomere [Arbuckle et al., 2010; Arbuckle et al., 2013]. The integration event results in insertion of the HHV-6A or -6B genome (figure 1.5) in an orientation, where U100 is located towards the centromere and U1 towards the telomeric cap, as well as the loss of the pac2 site from the DR_R [Arbuckle et al., 2010].



Figure 1.5 CI-HHV-6A/-6B chromosomal integration. Direct repeats (DR_L and DR_R), retained pac sites (pac1 and pac2) and gene locations (U1 and U100) labelled in relation to key telomeric features [Figure adapted from Prusty et al., 2013b].

1.3.2 Somatic Integration of HHV-6A & HHV-6B

The recent work by Arbuckle and colleagues has also led to the hypothesis that HHV-6A and -6B may be using chromosomal integration after primary infection of somatic cells as a means of achieving latency in a different manner to that observed in other human herpesviruses [Arbuckle et al., 2010; Arbuckle et al., 2013]. Arbuckle et al. were able to demonstrate the maintenance of viral genomic DNA in the apparent absence of episomal DNA, following infection of cell lines with HHV-6A and -6B (JJHan, Molt3 and HEK 293 cells). This was found to be due to the ability of a single copy of the viral genome to rapidly integrate into the chromosomes of some cells within a population of lytically competent cells, with the resultant cells remaining viable [Arbuckle et al., 2010; Arbuckle et al., 2013]. As such they proposed that integration represents an alternative mechanism of achieving latency upon primary infection of somatic cells, with telomeric integration providing a good site for latency due to comparatively limited active transcription [Arbuckle et al., 2010; Morissette & Flamand 2010; Arbuckle & Medveczky, 2011; Arbuckle et al., 2013]. With inherited CI-HHV-6A/-6B being a result of the rare occurrence of integration into the germ-line resulting from the broad cellular tropism of HHV-6A/-6B. In the case of somatic integration the number of cells harbouring CIHHV-6 would be relatively small, therefore not detected by conventional assays. Further compelling evidence for this hypothesis is added due to the suggestion that CI-HHV-6A/-6B can reactivate, discussed below.

1.3.3 Reactivation from Chromosomally Integrated HHV-6A & -6B

Increasing evidence is emerging with regard to the reactivation of CI-HHV-6A and -6B, both that of the inherited and somatic integration forms. Treatment of PBMCs from CI-HHV-6 patients and a CI-HHV-6 cell line with trichostatin A (TSA) and to a lesser extent 12-*O*-tetradecanoyl-13 acetate (TPA), compounds known to reactivate latent herpesviruses, could reactivate both CI-HHV-6A and -B to produce infectious virions [Arbuckle et al., 2010; Arbuckle & Medveczky, 2011; Arbuckle et al., 2013]. The sequence of the gB gene in transplacentally acquired HHV-6A cases was shown to be identical to that of mothers CI-HHV-6A yet divergent from other known HHV-6 isolates, implicating transmission of reactivated CI-HHV-6A [Gravel et al., 2013b].

At the start of this thesis, the relationship of these chromosomally integrated forms of HHV-6A/-6B to those of their circulating counterparts was unknown. This addition of more than 150 kb of viral DNA into the host chromosome, if intact, leads to the possibility of viral gene expression and/or reactivation in every nucleated cell of the body; as well as potentially significant effects on host chromosome function and integrity. Furthermore, along with the presence of the genome in every cell of the body, germ-line integrated virus can be transmitted vertically following the laws of Mendelian inheritance, as a congenital 'infection' [Hall et al., 2010]. Which may have effects on the generation and

regulation of the immune system, as well as infant development. Therefore part of this thesis concerns characterisation of the CI-HHV-6A/-6B genomes with comparison to circulating viral strains. Focusing on genetic variation which may define phenotypic and pathogenic variation between circulating viral strains and the inherited chromosomally integrated forms.

1.4 HHV-6A/-6B and the Chemokine System

1.4.1 Immunomodulation by HHV-6A and -6B

Mechanisms of immunomodulation by HHV-6A/-6B are important for virus persistence since they cause both primary and latent infection in key effector immune cells [Lusso et al., 1988; Takahashi et al., 1989; Kondo et al., 1991]. As such they have evolved a variety of means to interfere with immune responses. These include modulation of both innate and adaptive immunity.

Antigen presentation

Both HHV-6A and -6B have evolved means of specifically targeting antigen presentation pathways. Briefly, for antigen presentation in the MHC class I pathway, proteasomal degradation of cytosolic proteins leads to the generation of antigenic peptides. These antigenic peptides are translocated to the endoplasmic reticulum (ER) through the action of an ATP-dependent transporter protein known as transporter associated with antigen presentation (TAP), where they are loaded onto MHC class I complexes. Proper loading of the peptide on the complex permits vesicular transport and subsequent presentation at the cell surface. HHV-6A and -6B U21 specifically associates with newly synthesised MHC class I diverting them to lysosomal compartments, thereby reducing surface-expression. While both HHV-6A and -6B utilised this mechanism, and despite 90% similarity between the pU21 of HHV-6A and -6B, HHV-6B was shown to be much less efficient at downregulation [Hirata et al., 2001; Glosson & Hudson, 2007].

T cell responses

For HHV-6A it has also been shown that the product of ORF U24 poses a role in modulating T cell interactions. pU24 downregulated surface expression of both the T-cell receptor complex and CD3, targeting them to endosomal compartments for degradation. Thereby interfering with the activation of the T-cells by antigen presenting cells and potentially having profound effects on the development of adaptive immune responses [Sullivan & Coscoy, 2008].

Cytokine responses

Pattern recognition receptor (PRR) mediated recognition of pathogen-associated molecular patterns (PAMPs) with the subsequent cellular signalling and gene transcription involved in a multitude of immune processes, represents one of the earliest responses by the host to pathogen challenge. In the case of viral infection, the upregulation of cytokines known as interferons (IFNs), their signalling and induction of interferon stimulated genes (ISGs) is the key player in early host immune responses. As such this represents another vital target of herpesvirus immunomodulation. Interferon regulatory

factors (IRFs) are important in both the regulation of IFN and ISG transcription, while signal transducers and activators of transcription (STATs) are important in upregulating ISGs.

While not specifically an immunomodulation mechanism the general host protein synthesis shutdown that occurs upon infection by a number of herpesviruses, including HHV-6A/-6B [Di Luca et al., 1990], likely inhibits many immune responses. Additionally HHV-6B and to a lesser extent -6A, encode a means of directly interfering with IFN responses. HHV-6B infected cells are resistant to IFN- α and - β due to effects on the expression of IFN-stimulated genes. While the effects were much less pronounced in HHV-6A infected cells, representing a clear biological difference between the strains [Jaworska et al., 2007; 2010]. This effect was found to be mediated by IE1 expression, which interacts with STAT2 sequestering it to the nucleus. Thereby preventing the binding of the ISGF3 transcription factor to IFN-responsive gene promoters [Jaworska et al., 2010]. The differences between HHV-6A and HHV-6B with regard to IE1 mediated IFN resistance were found to map to a 41 amino acid region present in IE1B yet absent in IE1A [Jaworska et al., 2010].

1.4.2 The Chemokine System

Another immunomodulatory strategy employed most extensively by the betaherpesviruses is subversion of the host chemokine system, a key aspect in immune response signalling and leukocyte movement during an inflammatory response [McSharry et al., 2013; Dagna et al., 2013]. The human chemokine system, reviewed in appendix 7.1, is complex with multiple ligands and receptors expressed on specific leukocyte populations, figure 1.6 and table 1.2. The betaherpesviruses encode homologues of both chemokines and chemokine receptors. When expressed from infected cells these viral homologues modulate cellular signalling both to escape immune surveillance and assist viral replication.

Receptor	Expressing Cell
CCR1	N, Mo/M Φ , Th1, Tmem, NHC
CCR2	Mo/M Φ , Th1, iDC, Ba, NK
CCR3	Eo, Ba, Th2, MC, MG, DC, PC, NHC
CCR4	Th2, Th17, Treg, iDC, Mo, B, CD4+ & CD8+ T
CCR5	DC, Mo/M Φ , NK, Th1, Th17, Treg, CTL, NHC
CCR6	Th17, NK, NKT, Treg
CCR7	mDC, B, naïve T
CCR8	DC, Mo/M Φ , Th2, CD8+ T, Tregs, NHC
CCR9	Gut-homing T, B, DC
CCR10	Skin-homing T, skin Fb, EC
CXCR1	N, Mo, NK, MC, Ba, DC, CD8+ T, Treg, EC
CXCR2	N, Mo, NK, MC, Ba, DC, T, EC
CXCR3	B, Th1, CD8+ T, pDC, NK, NKT, Treg
CXCR4	Most leukocytes, NHC
CXCR5	B, CD8+ T
CXCR6	Th1, Th17, NKT, NK, PC
XCR1	DC
CX3CR1	Mo/M Φ , Th1, CTL, DC, NK, MG, neuron

Table 1.2 Chemokine receptors and their expression. Abbreviations: N, neutrophil; Mo, monocyte; MC, mast cell; Ba, basophil; DC, dendritic cell; EC, endothelial cell; pDC, plasmacytoid DC; NHC, nonhaematopoietic cell; PC, plasma cell; iDC, immature DC; Eo, eosinophil; MG, microglial cell; Fb, fibroblast.

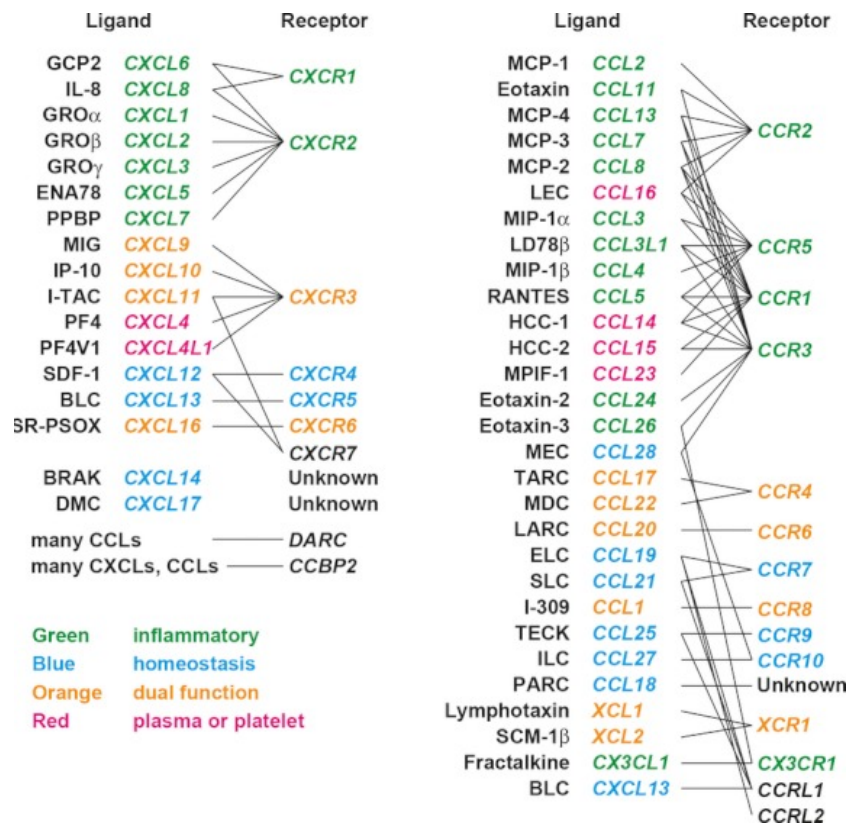


Figure 1.6 Chemokine ligands and their receptors. Both common and systematic nomenclature are included. Colours denote functional classification of the chemokines. [Figure from Nomiyama et al. 2013].

A number of viruses have developed methods to subvert or exploit the chemokine system as a means of establishing a niche or evading immune detection. Arguably, the most famous being the chemokine co-receptor usage by the HIV virus [Haggani & Tilton, 2013]. Additionally, some members of the *Poxviridae* and most notably the *Herpesviridae*, possess genes which can directly interfere with the chemokine system. The poxviruses seem to have predominantly adopted a strategy of encoding soluble chemokine binding proteins [Epperson et al., 2012]. While the herpesviruses, tend to encode homologues of cellular chemokines and chemokine receptors, likely pirated from the host during their co-evolution. Of the herpesviruses, those found within the *Betaherpesvirinae* subfamily make extensive use of these virally encoded chemokine component homologues. While most of these vGPCRs are dispensable for growth in tissue culture, their significance *in vivo* is increasingly being recognised. HHV-6A and -6B both encode a chemokine homologue, U83 and two chemokine receptor homologues, U12 and U51 [Gompels et al., 1995; Dominguez et al., 1999; Isegawa et al., 1999].

1.4.3 HHV-6A and -6B Chemokines

U83

HHV-6A and HHV-6B both encode a single chemokine homologue, U83 [Zou et al., 1999; Dewin et al., 2006]. This is one of the few genes which is unique to the HHV-6 species but absent from other betaherpesviruses. Original annotation of the genomes of HHV-6A and -6B noted that these ORF encoded a protein with homology to mammalian CC chemokines [Gompels et al., 1995; Dominguez et al., 1999; Isegawa et al., 1999]. Subsequent analysis of

the ORF in a number of strains from both species identified the hypervariable nature of this gene. Not only are large differences observed between the two species, U83 also displays a high level of variation between different strains from the same species [French et al., 1999; Sjahril et al., 2009]. Due to the variation of U83 observed between the two species and its association with an immunomodulatory role, U83 was proposed to be a prime candidate for many of the biologic and pathogenic differences observed between the two species.

Subsequently, the U83 genes from HHV-6A [Dewin et al., 2006; Catusse et al., 2007] and HHV-6B were functionally characterised. This did indeed highlight a number of differences between the gene in the two species. U83B was found to be an agonist of CCR2 capable of inducing chemotaxis in CCR2 expressing leukocytes [Zou et al., 1999; Luttichau et al., 2003; Clark et al., 2013]. In contrast U83A was a high potency agonist for CCR1, CCR4, CCR5, CCR6 and CCR8, capable of inducing chemotaxis of leukocytes [Dewin et al., 2006; Catusse et al., 2007; Clark et al., 2013]. Additionally, in contrast to human chemokines, U83A does not interact with the atypical chemokine receptors ACKR1/DARC or ACKR2/D6 [Catusse et al., 2009], receptors which play a vital role in regulation of the chemokine system [Nibbs & Graham, 2013]. Initially U83 was described to be expressed as a late kinetic class gene [Zou et al., 1999]. But soon after it was found that while the full length U83 protein was only expressed with late kinetics, U83 transcripts can be detected at early times. These transcripts, however, undergo splicing leading to introduction of an early stop codon resulting in the generation of truncated peptide, termed U83-N [French et al., 1999]. This N-terminal peptide form of U83 retained the receptor binding properties of the full length chemokine, however, acted as an antagonist at the receptor instead of an agonist as seen with the full length chemokine [Dewin et al., 2006; Catusse et al., 2007]. These properties led to suggestions of a role of inhibition of proinflammatory responses during the early stages of infection, followed by chemoattraction of susceptible cell types for latency and/or dissemination during the later stages of infection. It has also been shown that both the full length and truncated forms of U83A are highly potent inhibitors of HIV-1 infection [Catusse et al., 2007].

1.4.4 HHV-6A and -6B Encoded Chemokine Receptors

U12

Both HHV-6A and -6B also encode two GPCRs, U12 and U51, with homology to mammalian chemokine receptors. To date, information regarding the U12 ORFs of HHV-6A and HHV-6B remains sparse. Functional analysis of the U12 ORF and its products has only occurred for the HHV-6B species (strain HST) [Isegawa et al., 1998; Isegawa et al., 1999]. The U12 ORF was found to be expressed as a spliced transcript with late kinetics in HHV-6B infected cells [Isegawa et al., 1998]. While generation of stably expressing human erythroleukemic cell line, allowed for initial attempts at functional characterisation of pU12. Using these cell lines pU12 was found to bind the β -chemokine ligands CCL2, CCL3, CCL4 and CCL5, and be activated by these same ligands as measured by mobilisation of intracellular Ca^{2+} , with CCL2 and CCL5 acting as more effective Ca^{2+} mobilising agonists than CCL3 and CCL4 [Isegawa et al., 1998]. The U12 ORF of HHV-6B strain Z29 contains a deletion, respective to that of strain HST and other clinical isolates examined, which is predicted to result in the production of a truncated protein, implying U12 is dispensable for *in vitro* culture [Dominguez et al., 1999].

U51

The HHV-6A (strains U1102, GS) and HHV-6B (strains Z29 and HST) pU51 ORFs predict protein products of 301 amino acids in length, which show ~94% sequence identity with one another [Gompels et al., 1995; Gravel et al., 2013a; Dominguez et al., 1999; Isegawa et al., 1999]. A couple of particular features of note from examination of the U51 ORF were that the predicted N-terminal domain of the 7TM structure of these vGPCRs is unusually short in comparison to the human chemokine receptors with which it shares homology (~15 amino acids in comparison to ~30-40 amino acids), while the highly conserved DRY motif is actually ERI in the viral receptor.

Initial characterisation of HHV-6A (strain U1102) pU51 identified it as an early lytic gene [Menotti et al., 1999]. This initial study also led to suggestions that HHV-6 may modulate the surface expression of pU51 in a cell type dependent manner, since they found the expressed protein was retained in intracellular vesicles unless expressed in cells of T-lymphocytic lineage [Menotti et al., 1999]. However, later work has found that pU51 of HHV-6A, is also surface expressed in non-lymphoid cells, with the implication that the virus may be exploiting codon usage for variable surface expression of pU51 [Milne et al., 2000; Bradel-Tretheway et al., 2003]. The data from these studies also suggests that HHV-6A pU51 can form homodimers, however, while the importance of homo- and heterodimerisation of GPCRs is increasingly being recognised, the functional significance of this dimerisation (if any) with regard to pU51 function still remains unclear [Menotti et al., 1999; Milne et al., 2000; Bradel-Tretheway et al., 2003].

Initial attempts at characterising the ligand binding properties of HHV-6A pU51 led to the identification of its ability to bind a broad range of β -chemokines (CCL2, CCL5, CCL7, CCL11 and CCL13) and the KSHV viral chemokine homologue vMIP-II [Milne et al., 2000; Fitzsimons et al., 2006]. While later a further β -chemokine ligand (CCL19) as well as the γ -chemokine, XCL1, were found to bind at physiologically relevant concentrations [Catusse et al., 2008]. Thus, HHV-6A pU51 combined binding profile overlaps that of the predominantly inflammatory human chemokine receptors CCR1 (CCL5, CCL7), CCR2 (CCL2, CCL7, CCL13), CCR3 (CCL5, CCL7, CCL11 and CCL13), CCR5 (CCL2, CCL5, CCL7, CCL11, CCL13), CCR7 (CCL19) and XCR1 (XCL1).

With regards to the signalling capabilities, HHV-6A pU51 has been found to signal in a constitutive fashion through the $G\alpha$ subunits, but not $G\beta\gamma$ subunits, of G proteins of the the $G\alpha_{q/11}$ family activating PLC and inhibiting CRE-mediated gene transcription [Fitzsimons et al., 2006; Catusse et al., 2008]. In addition the constitutive signalling activity of the viral receptor can be differentially modulated through the binding of the human chemokine ligands [Fitzsimons et al., 2006; Catusse et al., 2008]. CCL5 was found to promote trafficking of the receptor signal to G proteins of both the $G\alpha_{q/11}$ and $G\alpha_{i/o}$ (specifically $G\alpha_{i3}$ and $G\alpha_{o1}$) families; while CCL2 and CCL11 trafficked the signal to only $G\alpha_{i/o}$ family G proteins ($G\alpha_{i1/12/13}$ and $G\alpha_{i1/12}$, respectively) [Fitzsimons et al., 2006]. This modulation of the receptor-G protein coupling resulted in decreased PLC activation and a abolishment of the inhibition of CRE-mediated gene transcription, compared to levels seen for the constitutively active receptor state [Fitzsimons et al., 2006].

The constitutive signalling activity exhibited by HHV-6A pU51 has been shown to specifically lead to the transcriptional down-regulation of two cellular targets. The first of these is CCL5, which has been demonstrated in both epithelial and haematopoietic cell

lines, where its roles have been hypothesised to include immune evasion or modulation of recruitment of circulating inflammatory cells for systemic spread of the virus [Milne et al., 2000; Caruso et al. 2003; Catusse et al., 2008]. The second cellular target specifically down-regulated by HHV-6A pU51 is FOG-2 in a haematopoietic cell line, FOG-2 is a transcriptional repressor which may have a role in influencing the Th1/Th2 balance by repressing Th2 development, but it is also hypothesised that this dysregulation may contribute to cardiac and inflammatory complications [Catusse et al., 2008]. Currently, the broader ranging functional consequences of the differential ligand dependent signalling remain unclear.

Finally, RNAi knockdown of HHV-6A pU51 was found to result in a reduction of viral replication and virally induced cytopathic effects, with the same study also suggesting pU51 may also be involved in direct cell-cell fusion mediated by viral proteins thus potentially providing a mechanism for this inhibition of viral replication [Zhen et al., 2005].

U51 has complex regulatory signalling, resembling properties of the vGPCR US28 from the closely related HCMV, which also combines constitutive and inducible signalling. Furthermore, US28 is expressed during latency and is a key immunomodulator [Beisser et al., 2001; Miller et al., 2012]. Expression of such potent immunomodulators such as the chemokine receptor U51 and chemokine U83 could affect many pathological processes. In the context of germ-line integrated CI-HHV-6A/-6B there could be profound effects, given possible gene expression in every cell. In this thesis the genetic relationship between the integrated genomes and that of infectious virus are investigated and then also evaluated for possible effects of any strain variation on potential immunomodulatory properties.

1.5 Aims

Therefore in order to investigate the nature of the CI-HHV-6A/B genomes and their potential effects, three major aims were established for the research presented here:

1. In order to avoid issues related to genomic changes resulting from laboratory culture and to be able to characterise clinical material directly, the thesis aimed to develop methods for the enrichment of Roseolovirus genomic DNA directly from clinical samples to permit whole genome characterisation of these viral species by NGS, with validation through sequencing of HHV-6A/-6B genomes. Thereby establishing methods to permit characterisation of the chromosomally integrated forms of these viral species to allow in depth investigation of the genetic variation between circulating and integrated genomes.
2. To compare the integrated viral genomes to exogenous infectious HHV-6A/-6B genomes. Through the application of the established NGS methods, to the recently defined germ-line chromosomally integrated forms of HHV-6A/-6B, in order to analyse any differences or similarities. With a focus on genetic variation which may define phenotypic and pathogenic variation between circulating viral strains and the inherited chromosomally integrated forms.
3. To characterise a key immunomodulatory gene from the CI-HHV-6A/-6B genomes, that of the virally encoded chemokine receptors, U51. Focussing on the molecular basis of the chemokine ligand-receptor interactions and possible implications of gene expression from CI-HHV-6A/-6B genomes.

Chapter 2: Methods

2.1 DNA Samples, Vectors and Reference Sequences

2.1.1 Virus Reference Strains and Infected Cell DNA

HHV-6A strain U1102 was isolated from a reactivated infection in a Ugandan HIV/AIDS patient [Downing et al., 1987] and strain AJ was originally isolated in the UK from a reactivated infection in an adult HIV/AIDS patient from the Gambia [Tedder et al., 1987]. Strain U1102 and AJ infected cell DNA was extracted previously in this laboratory from infected JJhan T leukaemic cell lines, as described by Dr. D. Clark [Clark, 2011].

2.1.2 CI-HHV-6A and -6B Patient DNA

Two cohorts of CI-HHV-6A/-6B samples were collected in collaboration with Dr. D. Lassner at the Institute of Cardiac Diagnostics and Therapy, Berlin, Germany and Dr. P. Hubacek at Motol University Hospital, Prague, Czech Republic. CI-HHV-6A/-6B was identified as described in Kuhl et al., 2005, Boutolleau et al., 2006 and Hubacek et al., 2013, with additional samples described in Tweedy et al., 2015a. Samples were DNA extracted from endomyocardial biopsies (EMB) or peripheral blood leukocytes (PBL). These were from European patients with haematological disorders, malignancy or inflammatory disease in the Czech Republic and cardiac disease in Germany where germline CI-HHV-6A/-6B had been indicated by diagnostic viral loads, with positive identification from hair or nail samples in some instances.

In the German cardiac cohort, viral load was quantified through real-time PCR screening of the conserved U94 locus, with comparison to human genome copy number calculated from the molecular mass of extracted blood DNA sample. In a subset of the PBL or EMB samples, RNA was extracted and cDNA prepared using Trizol reagent and reverse transcriptase as described previously [Kuhl et al., 2005; 2015].

In the Czech inflammatory disease cohort, relative copy number was determined following real-time PCR screening of the viral U65/66 locus and human albumin genes using a HHV-6A/B-specific Taqman assay, as described previously [Boutolleau et al., 2006]. Germline integration was also confirmed by Taqman detection of virus genomes in DNA extracted from hair or nail samples, as described by Hubacek et al. 2013.

2.1.3 Virus Genome Reference Sequences

Reference sequences used for HHV-6A and -6B were those where complete genomic sequences are available. The HHV-6A prototype strain U1102 as sequenced by Gompels et al., 1995 (Accession X83413/RefSeq NC_001664). Two reports, from different passaged cultures, of the HHV-6A strain GS isolated from a reactivated infection in an American patient with lymphoproliferative disease [Salahuddin et al., 1986] and as sequenced by Gravel et al., 2013a (Accession KC465951) and Bhattacharjee et al., 2014 (Accession KJ123690).

HHV-6B strain Z29 isolated from a reactivated infection HIV/AIDS patient from Zaire [Lopez et al., 1988] and sequenced by Dominguez et al., 1999 (Accession AF157706/RefSeq NC_000898). HHV-6B strain HST isolated from a Japanese infant during

primary infection [Yamanishi et al., 1988] and sequenced by Isegawa et al., 1999 (Accession AB021506).

2.1.4 Plasmid DNA

An amino-terminally haemagglutinin (HA)-epitope tagged human chemokine receptor CCR3 construct inserted into the pcDNA3 vector [Comabadiere et al., 1995; Auger et al., 2002], designated HA-CCR3, was generously donated by Dr. J. Pease (Imperial College London, London, UK). A HHV-6A strain U1102 chemokine receptor U51 coding sequence construct (designated pcDNA3-U51) and an amino-terminally HA-epitope tagged U51 construct cloned in pcDNA3 (designated U51D) as described in Milne et al., 2000 were obtained from Dr. U.A. Gompels (London School of Hygiene and Tropical Medicine, London, UK) of this laboratory. A human chemokine receptor CCR5 clone in pcDNA3 was obtained from the cDNA Resource Centre (Rolla, Missouri, USA). All vector construct identities were confirmed by Sanger capillary sequencing, as described in section 2.7, before use.

2.2 Primer Design

The primer design program Primer3Plus [Untergasser et al., 2007] was primarily used to aid design of primers. Due to location and mutational requirements, some primers were designed manually adhering as closely as possible to the generalised concepts of PCR primer design [Dieffenbach et al., 1993]. Multiple alignment programs ClustalW2 [Larkin et al., 2007] and Clustal Omega [Sievers et al., 2011] were used to confirm specificity and sequence identity between viral strains. Unless otherwise stated, all oligonucleotide primers were synthesised by Sigma-Aldrich (Gillingham, UK).

2.3 Primer Sequences

2.3.1 Long Range PCR Amplicons

Primer sets were designed based on the HHV-6A strain U1102 genome, with comparison to related HHV-6A strains. These primers sets were designed to generate 36 overlapping amplicons, ranging in size from 1-7 kbp, spanning the entire HHV-6A genome (table 2.1). Primer design was by Maria-Alexandra Spyrou, an MSc student in the laboratory, and described in our subsequent publications Tweedy et al., 2015a; 2015b.

Amplicon	Forward Primer (5'-3')	Reverse Primer (5'-3')	Amplicon Size (bp)
DR1	ACAACCGCATCTTCTTCAG	ATGAACGGAGATCTGGGAG	1750
DR1DR6	CCTCATCTGTTATTCCCTTCC	ACTCACCGCAGTCTTGAAC	2872
DR6	TCCAAACTGTACGGCTATG	TGGCGTCTAATCCAGGTAC	1858
DR_UL	CATGAGAGGACACTGGACGTACTC	GACATCTGTGGACCATGCA	2339
0203	ACCATGTGCGGAGAAAAGTTG	ACGAGAATGAATATGCCGATG	2174
0306	ACTCCTATCCGTCTAAACTGCT	ATTTCTCACGCCGGTATTC	4417
0611	ATGCCAAATCTAGCTGCTG	GAATCGGAGGAAGCTAACGT	4757
1114	TCTTCAGGTGTCTCAACCTTC	TGATGATCTGCTGTTGGAAC	5737
1419	TTGAGATGCAGAGCCTCTC	ACGTGGTGGATAATTCTTCAG	5370
1923	CTACATCATCGGTTGCCATA	TACAGAAGTTTCGGCAACTG	5216
2328	TAGTCGCTGGATGTCGATAC	GAGGAATTGTGGACAACGCT	4721
2830	CATCCAACTCCTTGTCCGT	TCGTTGACGCTGTCTGAC	5753
3031	GACACTGTCCAACGCATCT	CTCTTGGGCATAAAGTCCAGTA	6552
3134	ACAGAGCGACACCCAATATTC	AGCTGTGTCGATGTCTTGCTA	2787
3439	TGCATTCTCTCACAGCATG	TGTTTCCTTATGCCACCAATC	6309
3941	TGAGGCATGAGACTCAGAAG	AGCGGCATCCATAATGCTTA	4796
4142	AGCATCGTTTTGAAGACACC	GAGTTGTTGTTTCGCTGGTCT	2810
4142INT	TTGTTTCAGGCCAGCATC	ATACCGAAAACCTCAGGCAAC	2983
4245	AGACCGCATTTCGTATCGCT	CACGATCAAGTTCGAATCGTA	5092
4549	GAGTGCTCGGATAAGTTCATT	ACGATACAATCCGGAATGC	5922
4953	TGGAACCTCGGAATTGTTTCT	GTTTGAGACTGTATCTGCGCA	4282
5357	TATGGTGAAGAACCGTCCG	TCCGCAGATAACAGAATGG	4707
5758	CAGTGCCATCTGTTATGAATAG	AGTCGCAGTAAGGTCCACGT	6792
5864	TGCTGAAATTACCTCAGTGAG	CCATTGAGTATAATGCATCC	4229
6469	CTGTATCAATATCAAGGCGG	ACCACTGAATACGAACGCTC	5111
6974	GCGATTGTTCTGCGATAGAG	CTTATATCCGAGCCTTGCACT	6472
7476	GAGAACATGCTAGACAATTGG	CAAGAACTGCGACTCAATCC	2668
7679	GTTGCTAGTTGTATGACTTGG	CAATATCCACCGTTAGAGAAC	7681
7984	AGATGTATGCTGAAGAACGTG	GGATCGTCAACCGTTAGTG	5109
8486	CACAGTGTGTTTCGCCGGAAG	ACAGACAATGCACATCCTCTG	4773
8690	AGGTTGATACGGCAACGAG	GCAATCATTAGCATAACAGATG	3326
8690INT	TGAGGAATCACGTGTTTTG	ATCGAATCTATCCATGAAGATG	1746
90R3	CATCATTGTTATCGCTTTCAC	GCAACCGCAGTTCCTGTT	4912
R395	GCGGTACCCACTGATCTT	AGTCTACCAGGCATTCCGT	6250
95100	GAGGAGGGTCTGTCTAGATGT	TCGGAGATATCATAATCTGCGT	4031
100DR	TTATAGTTGCTCCCGAAAGC	AAGAAGATGCGGTTGTCTTG	2358

Table 2.1 LPCR amplicon primer sequences and expected sizes. Primers were designed based on the HHV-6A strain U1102 genome. Size indicates the expected base pair size of amplicons from LPCR reactions with HHV-6A strain U1102 template DNA.

2.3.2 AJ genomic Sequence Gap Filling

The following primer sets were used for the amplification and subsequent Sanger capillary sequencing of contig gaps, ambiguities and repetitive regions. For the DR_L-UL junction: DRL_ULF 5' CGTACACACGCAGACACACA 3' and DRL_ULR 5' GTATCCTCGTCTGCCCTCTG 3'. For the R2 region: R2F 5' GTGCTTTTTGTGTATGTCTCTATG 3' and R2R 5' CTCTATCTCTTATTTGTTTCCCTC 3'. For the R3 region R3F 5' GTACCCACTGATCTTTATCTTTATG 3' and R3R 5' CTAAAATCTGCTAACGGGG 3'. For the UL-DR_R junction: UL_DRRF1 5' TACTCTGACGCTCATCTTTTC 3', UL_DRRR1 5' GTTGTCTTGTATACAGGGGTAAC 3', UL_DRRF2 5' GGGCCGAGACTCCTTTTTT 3' and UL_DRRR2 5' TGGAGGAAGAAGGTGAAGTTG 3'.

2.3.3 CI-HHV-6A/-6B Gene Amplification

Primers used to amplify a fragment of the U38 gene, and the complete coding sequences of U46 and U83 were provided by Dr U.A. Gompels and described in Tweedy et al., 2015a; by Bates et al., 2009 or by Clark et al., 2013. These were: MaSU38F 5' AAGACGGGTTATTATGCTGTG 3' and MaSU38R 5' ACAGACATAAAGATGCTATCC 3'; U46F 5' TGTCTGCTGTTAATCACGTC 3' and U46R 5' GCGATCTAATAACCCTTCAC 3'; BHU83F1 5' GAAGGATCCTAATATGAGTGACATTAC 3', ERU83R1 5' TTCGAATTCTTTCATGATTCTTTGTC 3', U83FP1 5' AAGTTAACACGACGGGAACAA 3' and U83RP1 5' TGCCATATCACACATCGAG3'. To amplify the U83 locus, a semi-nested PCR procedure was used for some samples using primers BHU83F1 and U83RP1, then a second step using primers U83FP1 and U83RP1.

The complete coding sequences of U47, U51 and U65, were amplified using the following primers: U47aF 5' ATGTGTGATGCGTTGCATTT 3', U47aR 5' TCAGCGTCTCTCACATCACC 3', U47bF 5' TTCCGTTGTTTGC GTTGTTA 3' and U47bR 5' GTAATCGCGGTCCAACACA 3'; IU51SF1 5' GTCAATACGGATGGGGTTTTG 3' and IU51SR2 5' CAGCGCCGAAGATCTATTCT 3'; U65F 5' CAATCACCGGTAAATTCG 3' and U65R 5' CAGCTGCTACAATACACAG 3'.

2.3.4 Viral Chromosomal Integration Site Junction Amplifications

Primers for the amplification of the integration site junctions of HHV-6A at specific human chromosomal subtelomeric regions were as previously described by Britt-Compton et al., 2006 and Arbuckle et al., 2010. The 17p subtelomeric primer was 5' AACATCGAATCCACGGATTGCTTTGTGTAC 3' and the HHV-6 DR_R primer was 5' CATAGATCGGGACTGCTTGAAAGCGC 3'. Further chromosomal subtelomeric region primers, described in Huang et al., 2014 were also trialled in conjunction with the above HHV-6 DR_R primer. These included chromosome 10q 5' ATCCTTCCTCTTTGCAGCCG 3', chromosome 11q 5' CAGACCTTGGAGGCACGGCCTTCG 3', chromosome 18q 5' CTCATGTCCTCGGTCTCTTGCCCTC 3' and an alternative chromosome 17p primer: SubT17 5' CCAATTTACTGGTAATGGACT 3'.

Nucleotide sequencing of these amplicons was performed with the primer sequences above and a custom nested primer set designed as follows. For this, a ClustalW2 [Larkin et al., 2007] alignment of the previously published CI-HHV-6A and -6B, known and suspected 17p integration sites [GenBank accession numbers: GU784872.1, KF366419.1 and KF366420.1, respectively] was generated, then used to design primers located at a conserved region within the subtelomeric region of the amplicon, designated 17Tel-DR_IntF 5' TGTCGTTTCATTCCATTTATTCC 3', and a semi conserved region (1bp difference across 22bp between CI-HHV-6A and -6B) of the DR_R, designated 17Tel-DR_IntR 5' GGAAGACGACAACAGGTTT 3'.

2.3.5 U51A Chimeric and Mutant Construct Generation

2.3.5.1 U51NCCR3M

For the U51NCCR3M construct, the PCR-driven overlap extension protocol devised by Heckman & Pease, 2007 was utilised (summarised in section 2.11.1 and figure 2.2). For this, amplification of the predicted amino-terminal domain of U51A used, U51NF1 5'

HA-U51ΔM1A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCT**GCG**GAGAAAGAAACGAAGTCTTTGGC
 TTGG 3'
 HA-U51ΔE2A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATG**GCG**AAAGAAACGAAGTCTTTGGCT
 TGG 3'
 HA-U51ΔK3A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAG**GCA**GAAACGAAGTCTTTGGCT
 TGG 3'
 HA-U51ΔE4A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAA**GCA**ACGAAGTCTTTGGCT
 TGG 3'
 HA-U51ΔT5A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAA**GCG**AAGTCTTTGGCT
 TGG 3'
 HA-U51ΔK6A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAAC**GCGT**CTTTGGCT
 TGG 3'
 HA-U51ΔS7A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAG**GCT**TTGGC
 TTGG 3'
 HA-U51ΔL8A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAGTCT**GCGGC**
 TTGG 3'
 HA-U51ΔW10A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAGTCTTTGGC
TGCGCCA 3'
 HA-U51ΔP11A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAGTCTTTGGC
 TTGG**GCA**GCC 3'
 HA-U51ΔT13A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAGTCTTTGGC
 TTGGCCAGCC**GCT**GCG 3'
 HA-U51ΔE15A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAGTCTTTGGC
 TTGGCCAGCCACTGCG**GCG**TTT 3'
 HA-U51ΔF16A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAGTCTTTGGC
 TTGGCCAGCCACTGCGGAG**GCT**TAT 3'
 HA-U51ΔY17A 5'
 TCGAAGCTTATGTTACCCATACGATGTTCCAGATTACGCTATGGAGAAAGAAACGAAGTCTTTGGC
 TTGGCCAGCCACTGCGGAGTTT**GCT**GGT 3'

2.3.6 Vector Sequencing

Vector inserts were sequenced prior to use to confirm identities, as well as following cloning of newly generated inserts. The pGEM-T (Promega, Southampton, UK) and pcDNA3 (Life Technologies, Paisley, UK) vectors were used for cloning. Primers designed against the T7 and SP6 promoter sequences contained in these vectors, T7: 5' TAATACGACTCACTCTAGGG 3' and SP6: 5' ATTTAGGTGACACTATAG 3', were used to

sequence multiple cloning site inserts, as described in section 2.7.

2.4 Polymerase Chain Reaction (PCR)

2.4.1 General

Except for the specific cases outlined in sections 2.4.2 and 2.4.3, PCR amplifications were performed using 25 μ l reactions with GoTaq green mastermix (Promega, Southampton, UK), following the accompanying protocol. For these, 12 μ l GoTaq DNA polymerase green mastermix (Promega, Southampton, UK) was combined with 7 μ l nuclease-free water (Sigma-Aldrich, Gillingham, UK), 2.5 μ l forward primer (10 μ M for a final concentration of 1 μ M), 2.5 μ l reverse primer (10 μ M for a final concentration of 1 μ M) and 1 μ l of template DNA (at concentration of 50 ng/ μ l). A negative control (1 μ l nuclease-free water, described above, as template) and positive control (1 μ l of relevant DNA at concentration of 50 ng/ μ l) was included per 10 reactions, for all reaction preparations. Reaction preparation, DNA template addition, amplification and amplicon analyses were undertaken in separate facilities to prevent contamination.

Amplification reactions were performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). Thermocycling conditions were: a hot-start of 95°C for 2 minutes, followed by 35 cycles of 95°C for 20 seconds denaturation step, 58°C for 30 seconds annealing step, 72°C for 2 minutes extension step, and a final elongation at 72°C for 5 minutes.

2.4.2 Long Range PCR Amplicons

PCR amplifications were performed using 50 μ l reactions with GoTaq Long PCR mastermix (Promega, Southampton, UK), following the accompanying protocol. For these, 25 μ l GoTaq Long PCR DNA polymerase mastermix (Promega, Southampton, UK) was combined with 13 μ l nuclease-free water (Sigma-Aldrich, Gillingham, UK), 5 μ l forward primer (10 μ M for a final concentration of 1 μ M), 5 μ l reverse primer (10 μ M for a final concentration of 1 μ M) and 2 μ l of template DNA (242-460 ng of DNA). A negative control (2 μ l water as template) and positive control (2 μ l of relevant DNA corresponding to 200 ng of DNA) was included per 10 reactions, for all reaction preparations. Reaction preparation, DNA template addition, amplification and amplicon analyses were undertaken in separate facilities to prevent contamination.

Amplification reactions were performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). Thermocycling conditions were: a hot-start of 95°C for 2 minutes, followed by 35 cycles of 95°C for 20 seconds denaturation step, 59°C for 30 seconds annealing step, 70°C for 6 minutes extension step, and a final elongation at 72°C for 10 minutes.

2.4.3 Chromosome 17p Integration Site PCR Amplification

Reaction set up was performed as in section 2.4.1, with amplification reactions again performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). However, template DNA was increased from 50 ng to 242-460 ng of DNA depending on the CI-HHV-6A/B sample. Additionally, thermocycling conditions were based on those described in Britt-Compton et al., 2006 and Arbuckle et al, 2010, with

adjustments to account for different polymerases and amplicon sizes. These included of hot-start 94°C for 5 minutes, followed by 25 cycles of 94°C for 15 seconds denaturation step, 59°C for 30 seconds annealing step, 72°C for 5 minutes extension step, and a final elongation at 72°C for 10 minutes.

2.5 Agarose Gel Electrophoresis and DNA Purification

Separation of DNA was performed by agarose gel electrophoresis using 1.5% or 0.7% (the latter for the LPCR amplicons) agarose gels. The gels were prepared from 1.5 g or 0.7 g agarose (Fisher Scientific, Loughborough, UK) dissolved in 100 ml TBE buffer (made as a 10X stock solution: 108 g Tris base, 55 g Boric acid and 7.5 g EDTA disodium salt all from Sigma-Aldrich, Gillingham, UK dissolved in 1 L deionised water. Further diluted to 1X solution with deionised water prior to use) supplemented with 5µg/ml ethidium bromide (Sigma-Aldrich, Gillingham, UK) for DNA visualisation. Restriction digested and PCR reaction DNA were mixed with blue loading dye (New England Biolabs, Hitchin, UK), where necessary, prior to gel loading. Electrophoresis was carried out using a Sunrise 96 Horizontal Electrophoresis Apparatus (Gibco BRL, now Life Technologies, Paisley, UK) with 100 volts supplied by a BioRad Model 200/2.0 Electrophoresis Power Supply (Bio-Rad, Hemel Hempstead, UK). All samples were run with a relevant DNA ladder for the expected size of the amplicon, either a 100 bp or 1 kb DNA ladder (both from New England Biolabs, Hitchin, UK). DNA bands were visualised under UV light (365 nm) using a Gene Genius Bioimaging machine in conjunction with GeneSnap Image Acquisition software (Syngene, Cambridge, UK).

After confirmation of correct size, DNA bands were excised from the gel using a clean scalpel and DNA was extracted using silica membrane spin columns either the PureLink Quick Gel Extraction Kit (Life Technologies, Paisley, UK), Wizard SV Gel and PCR Clean-Up System (Promega, Southampton, UK) or the Zymoclean Gel DNA Recovery Kit (Zymo Research, Freiburg, Germany) and their accompanying protocols, respectively. While the exact specifics vary, all follow a generalised protocol whereby the extracted gel slice was weighed and in turn incubated with 3 volumes of a gel solubilisation buffer for 10 minutes at ~50°C, until completely dissolved. The dissolved solution was then applied to a silica spin column and centrifuged at >12,000 g for 1 minute using a Micro Centaur Plus microcentrifuge (MSE, Lower Sydenham, UK). Run-through was discarded, and two rounds of a wash step were performed, whereby a wash buffer was applied to the column and centrifuged at >12,000 g for 1-5 minutes, with run-through being discarded after each washing round. Sample DNA was then eluted into a fresh microcentrifuge tube, through the addition of 30 µl nuclease-free water (Sigma-Aldrich, Gillingham, UK) to the column, which was left to stand for 1 minute prior to a final centrifugation step at >12,000 g for 1 minute. Eluted DNA was then either used immediately or stored at -20°C.

2.6 DNA Quantification

2.6.1 General

Except in the case highlighted in section 2.6.2, DNA concentrations and purity were calculated using 1 µl of sample DNA on a Nanodrop 1000 spectrophotometer (ThermoFisher Scientific, Loughborough, UK), according to the manufacturers guidelines.

2.6.2 Long Range PCR Amplicon DNA Quantification

Prior to equimolar pooling, long range PCR (LPCR) purified amplicon DNA concentrations were quantified as in section 2.6.2. Both LPCR amplicon and SureSelect enriched DNA library concentrations were calculated post-shearing and post-library preparation using the Agilent High Sensitivity DNA Kit and 2100 Bioanalyzer Instrument (both Agilent, Stockport, UK), in accordance with the manufacturers guidelines.

2.7 Sanger Nucleotide Sequencing and Resolution

Sanger capillary DNA sequencing was performed either using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, now Life Technologies, Paisley, UK) and resolved on an ABI Prism 3730 DNA Analyzer (Applied Biosystems, now Life Technologies, Paisley, UK) at LSHTM or purified and sent for Sanger sequencing by Source Bioscience (Nottingham, UK).

The BigDye Terminator v3.1 Cycle Sequencing Kit was utilised according to the accompanying protocol. Briefly, 10 µl reactions were prepared consisting of 1 µl BigDye termination mix, 3 µl dilution buffer, 1 µl sequencing primer, 1 µl template DNA and 4 µl nuclease free water (Sigma-Aldrich, Gillingham, UK). Unless otherwise stated, the PCR amplification primers were used in the sequencing reactions. Amplification reactions were performed on a DYAD PTC-220 peltier thermal cycler (MJ Research, now Bio-Rad, Hemel Hempstead, UK). Thermocycling conditions were: a hot-start of 96°C for 4 minutes, followed by 25 cycles of 96°C for 20 seconds denaturation step, 50°C for 10 seconds annealing step, 60°C for 2 minutes extension step, and a final elongation at 60°C for 5 minutes. DNA from the sequencing reactions were precipitated through the addition of a 0.1M sodium acetate/70% ethanol solution and incubation on ice for 20 minutes, followed by centrifugation at 4°C and 3000 g for 30 minutes and removal of the supernatant. The DNA was then subjected to two rounds of a wash step, involving resuspension in 70% ethanol and centrifugation at 4°C and 3000 g for 10 minutes, prior to removal of the supernatant. DNA was then resuspended in Hi-Di Formamide (Life Technologies, Paisley, UK) and resolved on an ABI Prism 3730 DNA Analyzer (Applied Biosystems, now Life Technologies, Paisley, UK).

Resulting sequence traces from both methods were visualised and manually corrected using ChromasPro version 1.7.6 (Technelysium, Brisbane, Australia). Sequences were resolved from both forward and reverse primers and combined to generate consensus sequences. Consensus sequences were compared to National Centre for Biotechnology Information (NCBI) and European Bioinformatics Institute (EBI) archived sequences using their respective Basic Local Alignment Search Tools (BLAST) [Camacho et al., 2009; Lopez et al., 2003].

2.8 Illumina MiSeq NGS Sequencing, Assembly and Annotation

2.8.1 Target Enrichment using Agilent SureSelect Target Enrichment

The solution hybrid selection of HHV-6A genomic DNA was performed as described previously [Gnirke et al., 2009; Depledge et al., 2011], using the SureSelect Target

Enrichment System (Agilent Technologies, Stockport, UK). Briefly, HHV-6A, strain U1102 or AJ, infected cell total DNA was sheared to an average size of 200 bp (6x60 seconds: duty cycle 10%, intensity 5 and 200 cycles per burst) using a Covaris E210 focused-ultrasonicator (Covaris, Brighton, UK), followed by end repair, dA-tailing, adapter ligation and PCR enrichment according to the manufacturers guidelines with all required purification steps performed using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK). Overlapping, 120-mer biotinylated RNA baits, custom designed against the HHV-6A strain U1102 genome sequence (Accession X83413, RefSeq NC_001664) were then hybridised to the viral DNA component of this total cellular fragmented DNA library for 36 hours. Before streptavidin magnetic beads were used to capture HHV-6A genomic DNA. This was followed by a limited number of rounds of PCR to amplify the captured library.

2.8.2 Target Enrichment using LPCR Amplicons

Long-range PCR (LPCR) amplicon based sequencing was performed as described previously [Depledge et al., 2011]. For this, primer sets (synthesised by Sigma-Aldrich, Gillingham, UK), described in section 2.3.1, were designed against the HHV-6A strain U1102 reference genome [Gompels et al., 1995] (Accession X83413, RefSeq NC_001664) in overlapping 1-7kb amplicons. 36 overlapping PCR amplicons were generated using GoTaq Long PCR mastermix (Promega, Southampton, UK) and nuclease-free H₂O (Sigma-Aldrich, Gillingham, UK) with thermocycling using a hot start 95°C for 2 minutes, then 35 cycles of 95°C for 20 seconds, 59°C for 30 seconds, 70°C for 6 minutes, and a final elongation step of 72°C for 10 minutes. Amplicons were size-selected on 0.7% agarose gels, then purified using the Wizard SV gel and PCR clean-up kit (Promega, Southampton, UK). Pooled equimolar amplicons were sheared using an E210 focused-ultrasonicator (Covaris, Brighton, UK) to an average size 200 bp. End repair, dA-tailing, adapter ligation, and PCR enrichment used NEBNext DNA library prep master mix set for Illumina with multiplex oligos (New England Biolabs, Hitchin, UK). Again all required purification steps were performed using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK).

2.8.3 Illumina MiSeq Sequencing

DNA quality and quantification of prepared libraries was assessed as described in section 2.6.2. Final, indexed DNA libraries were denatured, diluted, then loaded onto a MiSeq v2 reagent cartridge following the MiSeq guide (Illumina, Little Chesterford, UK) for 2 x 250 bp paired-end sequencing. Post-sequencing, raw sequence data quality was assessed with FastQC (Babraham Bioinformatics, Cambridge, UK). The Fastq file reads had adapters removed and quality trimming using a phred score of 33 and minimum length of 100 base pairs with trimmomatic version 0.32 [Bolger et al., 2014].

2.8.4 NGS Bioinformatics: Sequence Assembly

For mapped assemblies, trimmed read-pairs were mapped to the HHV-6A strain U1102 reference genome [Gompels et al., 1995] (Accession X83413, RefSeq NC_001664) using the BWA-MEM alignment algorithm of BWA version 0.7.10 and SAMtools version 1.0 [Li and Durbin, 2009; Li et al., 2009]. Average read coverage was calculated using the DepthOfCoverage tool from the Genome Analysis Toolkit (GATK) software version 3.3 [McKenna et al., 2010] and alignment quality was assessed using Qualimap version 2.0

[Garcia-Alcade et al., 2012]. Variant calling used both a SAMtools version 1.0 mpileup, BCFtools, vcfutils varFilter pipeline [Li et al., 2009; Danecek et al., 2011] and the GATK version 3.3 UnifiedGenotyper tool [DePristo et al., 2011]. For *de novo* assemblies a VelvetOptimiser version 2.2.5, Velvet version 1.2.10 [Zerbino & Birney, 2008], ABACAS version 1.3.1 [Assefa et al., 2009] pipeline was used to optimise assembly and contig ordering using HHV-6A strains U1102 [Gompels et al., 1995] (Accession X83413; RefSeq NC_001664) and GS [Gravel et al., 2013a] (Accession KC465951), with manual adjustments using Artemis version 16 [Rutherford et al., 2000; Carver et al., 2005; 2012]. Gaps, ambiguities and repetitive regions were confirmed by PCR and Sanger capillary sequencing as described in section 2.4 and 2.7.

2.8.5 NGS Bioinformatics: Annotations

Initial annotations were transferred onto generated consensus sequences using the Rapid Annotation Transfer Tool (RATT) [Otto et al., 2011] and the annotation of the prototypical HHV-6A strain U1102 [Gompels et al., 1995] (Accession X83413, RefSeq NC_001664). These were in turn updated and manually corrected in Artemis version 16 [Rutherford et al., 2000] guided by GeneMark ORF predictions [Besemer et al., 2001] and the experimentally determined annotation corrections found in more recently sequenced HHV-6A and HHV-7 strains [Megaw et al., 1998; Donaldson et al., 2013; Gravel et al., 2013a].

2.8.6 NGS Bioinformatics: Minor Variant SNP Analyses

Variant calling was performed as described in section 2.8.4. As described in our subsequent publication [Tweedy et al., 2015a] “A cut-off for SNP sensitivity was applied using an database control, a CG>GC inversion in the U83 CDS, previously identified and corrected [French et al., 1999; Dewin et al., 2006] but still present on the archived NCBI database sequence. At this site, the correct sequence was called to a depth of 6148 reads with a level of detection of 0.02 %; a cut-off 1 log above this was applied to verify SNPs, with a sensitivity of 0.2 %, and SNPs confirmed by Sanger capillary sequencing.”

2.9 Sequence Alignments and Phylogenetic Relationships

Multiple alignments of nucleotide and encoded amino acid sequences were calculated using either Clustal W [Larkin et al., 2007] or MUSCLE [Edgar, 2004] using the Molecular Evolutionary Genetics Analysis (MEGA) software version 6.0.5 [Tamura et al., 2013]. For nucleotide alignments, translated amino acid sequences were aligned before back-translation. For gene subset alignments, amino acid sequences or nucleotide coding sequences were concatenated and aligned as described above, after which any gaps were removed prior to downstream analysis.

Phylogenetic relationships were calculated by Maximum Likelihood statistical method [Felsenstein, 1981] using MEGA version 6.0.5 [Tamura et al., 2013]. Nucleotide sequence trees were constructed using the Tamura–Nei model [Tamura & Nei, 1993], with uniform rates, close neighbour interchange, invariant sites allowance, and checked with 1050 bootstrap replicates. Amino acid sequence trees were constructed using the Jones–Taylor–Thornton model [Jones et al., 1992], again using close neighbour interchange, invariant site allowance, and tested with 1050 bootstrap replicates.

2.10 Molecular Modelling of U51A-CCL2 Interaction

Secondary structure predictions were made using RaptorX [Kallberg et al., 2012], JPred [Cole et al., 2008] and SOPMA [Geourjon & Deleage, 1995] web servers. All viral receptor amino-terminal domain peptide structures were prepared using HyperChem (Hypercube Inc, Gainesville, Florida, USA). Helices were modelled with ionic ends removed and optimised in Amber99 force field in vacuo for 8000 cycles (utilising the following parameters: Constant dielectric scale factor: 1; Switched cut-offs, outer: 14Å, inner: 10Å; Electrostatic scale factor: 0.8333 Van der Waals scale factor: 0.5) [Cornell et al., 1995]. The X-ray crystal structure of the CCL2 I-form refined to 2.40Å [Lubkowski et al. 1997] was obtained from the RCSB protein data bank (PDB ID: 1DOL). Modelled viral amino-terminal peptides were docked to the CCL2 crystal structure using the AutoDock Vina software (search space parameters: centre x 29.451, y 45.273, z 23.508; size x 20Å, y 40Å, z 16Å) [Trott & Olson, 2010].

2.11 Plasmid Cloning

2.11.1 PCR-driven Overlap Extension

For the generation of chimeric constructs, PCR-driven overlap extension was performed following the protocol devised by Heckman and Pease, 2007 (figure 2.1), except that the expand high fidelity PCR system was substituted for the KOD Hot Start DNA Polymerase system (Merck Millipore, Watford, UK) with the accompanying protocol being followed. The primers U51NF1 and U51NR1 (section 2.3.5) and a vector containing a U51 construct were used to amplify the predicted amino-terminal domain of the U51 gene coding sequence. Amplification with these mutagenic primers was also designed to add an EcoRI restriction site before the amino-terminus domain sequence and a section of sequence complementary to the required chimeric joining site of CCR3. In addition, two further primers CCR3MF1 and CCR3MR1 and a vector containing a HA tagged-CCR3 construct were used to amplify the coding sequence of CCR3 minus its amino-terminal domain. This included the addition of a XhoI restriction site at the carboxyl-terminus. Subsequently, these two PCR products were both utilised in a second round of PCR, along with the primers U51NF1 and CCR3MR1 to allow for the generation of a chimeric receptor construct. The generated chimeric construct was then purified and inserted into a HA tag containing pcDNA3 vector using the EcoRI and XhoI restriction sites, as described in sections 2.11.3 – 2.11.6, to generate a vector designated HA-U51NCCR3M.

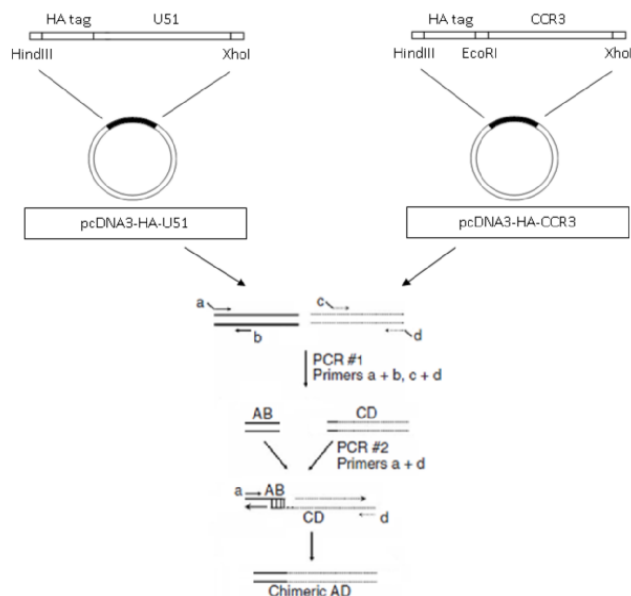


Figure 2.2 PCR-mediated overlap extension to generate an N-terminally swapped chimeric receptor. The first round of PCRs generates two PCR products with overlapping sequences due to the inclusion of nucleotides in the internal primers, b and c, that span the junction of segments AB and CD. Here AB is used to signify the N terminal region of the U51 gene and CD signifies the CCR3 gene minus the N terminus. These two products are then used as the template for a second PCR to generate the hybrid gene product AD. The inclusion of restriction enzyme site in the sequences of the two external primers then allows for insertion of this chimeric receptor construct into an expression vector [Figure adapted from Heckman & Pease, 2007].

2.11.2 HA-U51NCCR3

The HA-U51NCCR3 was generated using the T7 and U51DNR1 primers to amplify the amino-terminal domain of U51A with a HA tag and an EcoRI restriction site at the 3' end. This construct was in turn inserted into the HA-CCR3 vector using the HindIII and EcoRI restriction sites.

2.11.3 Alanine-scanning Mutagenesis

For alanine-scanning mutagenesis of the entire amino-terminal domain of U51A, the Heckman and Pease, 2007 protocol described above was adapted to allow for the generation of mutant constructs with a single PCR step. The primers HA-U51ΔM1A – HA-U51ΔY17A described in section 2.3.5.4, which were designed to contain the entire amino-terminal domain of U51A with relevant mutations for alanine scanning, plus the addition of an amino-terminal HA tag and EcoRI restriction site, were used in conjunction with the U51MR1 primer and the U51A vector to amplify alanine-scanning mutant constructs. These constructs were inserted into pGEM-T as a transfer vector, a selection of these constructs (U51A-ΔE2G, -ΔE2A, -ΔK3A, -ΔE4A, -ΔK6A and -ΔE15A) were then transferred to a HA tagged pcDNA3 expression vector using the EcoRI and XhoI restriction sites, as described in sections 2.11.3 – 2.11.6, to generate vectors designated HA-U51ΔE2G, HA-U51ΔE2A, HA-U51ΔK3A, HA-U51ΔE4A, HA-U51ΔK6A and HA-U51ΔE15A.

2.11.4 Restriction Digestion

All restriction digestions were performed using restriction enzymes from New England Biolabs (Hitchin, UK), following the accompanying protocols. For double digestions 50 µl reactions were prepared as follows: 1 µl restriction enzyme 1 (10 enzyme units), 1 µl restriction enzyme 2 (10 enzyme units), 5 µl relevant digestion buffer, X µl of sample DNA corresponding to 1 µg total DNA, Y µl nuclease-free water (Sigma-Aldrich, Gillingham, UK) to make up to 50 µl total reaction volume. Digestion reactions were incubated at 37°C for 1 hour. Subsequently, restriction enzymes were heat inactivated by incubation at 65°C (in some cases 80°C) for 20 minutes. Reaction DNA was separated, visualised and gel purified, as described in section 2.5.

For single restriction enzyme digestions, the volume of nuclease-free water was increased accordingly. In addition, where necessary, post-digestion DNA was treated with 1 µl Antarctic Phosphatase (1 enzyme unit) (New England Biolabs, Hitchin, UK) and incubated at 37°C for 15 minutes prior to heat inactivation in order to prevent self-ligation.

2.11.5 Ligation

All ligations were performed using T4 DNA ligase (New England Biolabs, Hitchin, UK), in accordance with the manufacturers guidelines. Briefly, the size and DNA concentration of purified digested DNA, quantified as described in section 2.5, was used to calculate the mass of insert required for a 3:1 molar insert:vector ratio. 20 µl ligation reactions were then set up as follows: 1 µl T4 DNA ligase (1 enzyme unit), 2 µl ligation buffer, X µl insert DNA corresponding to 3:1 molar insert:vector ratio, Y µl vector DNA corresponding to 3:1 molar insert:vector ratio, Z µl nuclease-free water (Sigma-Aldrich, Gillingham, UK) to make up to 20 µl total reaction volume. Ligation reactions were then incubated at room temperature for 1 hour or at 16°C overnight, followed by incubation at 65°C for 10 minutes for enzyme heat inactivation.

2.11.6 Transformation

Vectors were transformed into competent *Escherichia coli* (*E. coli*) strain JM109 cells. Either commercially prepared cells: JM109 Competent Cells $>10^7$ cfu/µg (Promega, Southampton, UK) or MAX Efficiency DH5α Competent Cells (Life Technologies, Paisley, UK) following their accompanying protocols, respectively. Or prepared according to the protocol outlined by Chung et al., 1989. For this, *E. coli* strain JM109 cells were grown in LB broth (10g tryptone, 5g yeast extract and 10g NaCl dissolved in 1L de-ionised water, all from Sigma-Aldrich, Gillingham, UK) to the early exponential phase as estimated by absorbance at 600 nm (OD_{600}). Cells were then diluted 1:1 with transformation and storage solution (TSS) (LB broth containing 10% (wt/vol) PEG, 5% (vol/vol) DMSO and 50 mM $MgCl_2$ all from Sigma-Aldrich, Gillingham, UK) and stored at -80°C for long-term storage.

Transformation competent cells were thawed on ice, up to 50 ng of sample DNA is mixed with the cells and incubated on ice for 10 minutes. The cells were then heat-shocked for 45 seconds by immersion in a water bath at 42°C, followed by a further 2 minute incubation on ice. Cells were then mixed with 900 µl of SOC medium (Life Technologies, Paisley, UK) and incubated for 60 minutes at 37°C with shaking. Followed by plating onto LB/ampicillin agar plates (LB medium prepared as above, supplemented with 15 g/L agar and 100 µg/mL ampicillin, both from Sigma-Aldrich, Gillingham, UK) and incubated at 37°C

overnight.

2.11.7 Plasmid Preparation

Streaking of picked colonies on LB/ampicillin plates, prepared as described in section 2.11.5, was performed for single colony isolation. Individual colonies were then grown overnight in LB broth, also as described in section 2.11.5, supplemented with 100 µg/mL ampicillin (Sigma-Aldrich, Gillingham, UK), at 37°C with shaking. Plasmid DNA was then purified using either the PureLink Quick Plasmid Miniprep Kit or the PureLink HiPure Plasmid Maxiprep Kit (both Life Technologies, Paisley, UK). In general, the overnight LB/ampicillin cultures were harvested by centrifugation at 5000 g for 10 minutes and the supernatant removed, followed by cell lysis using a modified version of the alkaline lysis method [Birnboim & Doly, 1979], with isotonic resuspension buffer supplemented with RNase, and an alkaline lysis buffer containing a detergent, followed by incubation at room temperature for 5 minutes. A precipitation/neutralisation buffer was added to the lysed cell mixture and the lysate is centrifuged for 10 minutes at >12,000 g. The plasmid DNA containing supernatant was then transferred to either a silica membrane spin column, or in the case of maxiprep purification an anion-exchange resin column with solutions passing through the column by gravitational instead of centrifugal force, for purification as described in section 2.5. Plasmid DNA insert identities were confirmed by Sanger capillary sequencing as described in section 2.7.

2.12 Cells and Cell Culture

2.12.1 Cell Types

The human lymphoblastic K562 cell line (derived from chronic myelogenous leukemia) [Klein et al., 1976] was obtained from Dr. U.A. Gompels (London School of Hygiene and Tropical Medicine, London, UK) and maintained in RPMI 1640 media (Sigma Aldrich, Gillingham, UK) supplemented with 10% v/v Fetal Bovine Serum (Life Technologies, Paisley, UK), 50 IU/ml penicillin, 50 µg/ml streptomycin and 2 mM glutamine (all Life Technologies, Paisley, UK), as described by Milne et al., 2000. The human monocytic THP-1 cell line (derived from monocytic leukemia) [Tsuchiya et al., 1980] was also obtained from Dr. U.A. Gompels (London School of Hygiene and Tropical Medicine, London, UK) and maintained in RPMI 1640 media with GlutaMAX-I and 25 mM HEPES pH 7.4 (from Sigma Aldrich, Gillingham, UK) supplemented with 10% v/v Fetal Bovine Serum (Life Technologies, Paisley, UK), 2 mM glutamine, 50 IU/ml penicillin, and 50 µg/ml streptomycin (all Life Technologies, Paisley, UK), as described by Dewin et al., 2006. The murine pre-B (derived from Abelson murine leukemia virus transformation of adult C57L mouse cells) were generously donated by Dr. J. Pease (Imperial College London, UK) and maintained in RPMI 1640 media with GlutaMAX-I and 25 mM HEPES (Sigma Aldrich, Gillingham, UK) supplemented with 10% v/v Fetal Bovine Serum (Life Technologies, Paisley, UK), 50 IU/ml penicillin, 50 µg/ml streptomycin, 1x non-essential amino acids, 1 mM sodium pyruvate and 50 µM 2-mercaptoethanol (all Life Technologies, Paisley, UK), as described by Vaidehi et al., 2009.

Cells were grown at 37°C, 5% CO₂ in a Sanyo MCO-15 incubator (Sanyo Electric Biomedical, Loughborough, UK). Cells were counted every 36 hours, as described in section 2.12.3, and maintained at log phase, at a concentration between 2 x 10⁵ and 1 x 10⁶ cells/ml. This was performed by centrifugation of sufficient total viable cells numbers at 125 g for 5 minutes

in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK), supernatant removal and resuspension in fresh media, as described above. Cultures exceeding 1.5×10^6 cells/ml were discarded due to reported suboptimal expression, post transfection [Vaidehi et al., 2009].

2.12.2 Resuscitation of Cells and Storage

All procedures were carried out in a sterile environment using a class II safety Heraeus LaminAir HB 2436 laminar flow hood (ThermoFisher Scientific, Loughborough, UK). All surfaces, bottles and equipment were cleaned with 70% ethanol prior to their use or addition to the flow hood.

For resuscitation, cells from liquid nitrogen storage were thawed in a pre-warmed 37°C water bath, added to 9 ml of pre-warmed culture media, and centrifuged at 125 g for 6 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK). The supernatant was carefully removed before resuspension in fresh pre-warmed media and transfer to a T25 tissue culture flask (Corning Life Sciences, Kennebunk, Maine, USA) which was incubated at 37°C , 5% CO_2 in a Sanyo MCO-15 incubator (Sanyo Electric Biomedical, Loughborough, UK).

For the production and storage of cell stocks, cultured cells were counted as described in section 2.12.3. Media containing 5×10^6 cells was centrifuged at 125 g for 5 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and the supernatant removed. Cells were then resuspended in 1 ml 90% Fetal Bovine Serum (Life Technologies, Paisley, UK)/10% DMSO (Sigma Aldrich, Gillingham, UK), transferred to a cryotube, frozen slowly at -80°C before transfer to liquid nitrogen storage.

2.12.3 Cell Counting

Viable cell number per ml and total viable cell number were estimated using a Neubauer haemocytometer. For this cultured cell suspension and trypan blue (Sigma Aldrich, Gillingham, UK) were mixed in a 1:1 ratio then added to the haemocytometer chambers. On the basis of trypan blue exclusion, viable cell counts per ml were calculated, which in turn was multiplied by total cell suspension volume to estimate total viable cell numbers.

2.12.4 Expression Vector Transfection

DNA electroporation of L1.2 cells was performed as described by Vaidehi et al., 2009. 24 hours prior to electroporation cells were split to 0.5×10^6 cells/ml so they would be in logarithmic growth phase (approximately 1×10^6 cells/ml) prior to electroporation. Cells were counted as described in section 2.12.3, then a cell suspension corresponding to 10×10^6 cells was centrifuged at 300 g for 5 minutes and resuspended in 800 μl RPMI 1640 media with GlutaMAX-I and 25 mM HEPES (Sigma Aldrich, Gillingham, UK). These cells were added to an 0.4-cm electrode gap electroporation cuvette (VWR, Lutterworth, UK) along with 0.5 mg/ml tRNA (Sigma Aldrich, Gillingham, UK) and 1 μg expression plasmid DNA before incubation at room temperature for 30 minutes. Cells were then electroporated at 330 volts and 975 μF using a Gene-Pulser electroporator (BioRad, Hemel Hempstead, UK) and incubated at room temperature for a further 30 minutes. Cells were then transferred to a T25 tissue culture flask (Corning Life Sciences, Kennebunk, Maine, USA) and resuspended in complete media, as described in section 2.12.1, at a final concentration of 1×10^6 cells/ml and incubated for 3 hours at $37^\circ\text{C}/5\% \text{CO}_2$ in a Sanyo MCO-15 incubator

(Sanyo Electric Biomedical, Loughborough, UK) before addition of sodium butyrate to a final concentration of 10 mM. Cells were then incubated for a further 24 - 48 hours to allow gene expression.

K562 and THP-1 were electroporated based on an adaptation of the protocol described by Milne et al., 2000. These electroporation were performed as described above except that cells were added to the cuvette along with 0.1 mM Dithiothreitol, 10 mM Glucose (both Sigma Aldrich, Gillingham, UK) along with 20 µg expression plasmid DNA and sodium butyrate induction was omitted.

2.13 Chemokines and Chemotaxis Assay

2.13.1 Chemokines

All chemokines were purchased from Peprotech (Rocky Hill, New Jersey, USA), and prepared according to the manufacturers guidelines. Lyophilised chemokines were reconstituted in nuclease-free water (Sigma Aldrich, Gillingham, UK), working aliquots were prepared by dilution with RPMI 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK) and stored at -20°C.

2.13.2 Calcein-AM Staining

Receptor expressing cells were centrifuged for 5 minutes at 300 g in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and resuspended in RPMI 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK) with 4 µM calcein-AM (Life Technologies, Paisley, UK) for live cell staining. Cells were then incubated for 30 minutes at 37°C, before being washed twice by centrifugation for 5 minutes at 300 g in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and resuspension in RPMI 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK) prior to use in the chemotaxis assay, described in section 2.13.3. Post chemotaxis, remaining cell suspension was removed from the top of the chemotaxis filter and live cell migration was assessed using a Wallac Victor 1420 Multilabel Counter (Perkin Elmer, Seer Green, UK) with an excitation wavelength of 485 nm and an emission wavelength of 535 nm.

2.13.3 Chemotaxis Assay

Chemotaxis was performed as described by Vaidehi et al., 2009. The Neuroprobe ChemoTX chemotaxis plate wells (Receptor Technologies, Warwick, UK) were blocked with RPMI 1640/1% BSA (both Sigma Aldrich, Gillingham, UK) and incubated for 30 minutes, then rinsed with de-ionised water. Dilutions of chemokines, prepared in RPMI 1640/0.1% BSA together with controls containing chemokine buffer only, were placed in the lower wells of the chemotaxis plate in triplicate at a volume of 31 µl per well. The 5 µm pore size filter was then overlaid on the well plate. 50 µl resuspended calcein-AM stained cells at a cell density of 3×10^6 cell/ml, prepared as described in section 2.13.2, were then applied atop the chemotaxis filter above each well. The lid of the plate was fitted and the entire plate was incubated at 37°C/5% CO₂ for 3-5 hours. Live cell migration was calculated as described in section 2.13.2.

2.14 Flow Cytometry

Cells were counted as described in section 2.12.3, 2×10^5 cells were collected, centrifuged at 400 g for 4 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK), with supernatant removed before resuspension in PBS/0.1% BSA (both Sigma Aldrich, Gillingham, UK). Cells were then stained with relevant fluorescently labelled antibodies (all from R&D Systems, Abingdon, UK, except for anti-HA-FITC which was from Roche, Burgess Hill, UK) and incubated for 30 minutes at 4°C. Subsequently, cells were washed with PBS/0.1% (both Sigma Aldrich, Gillingham, UK) with centrifugation at 400 g for 4 minutes and fixed through resuspension in PBS/2% PFA (both Sigma Aldrich, Gillingham, UK) and incubation at room temperature for 15 minutes. Cells were then subjected to a further wash step as described above. Stained cells were run on a FACS Calibur flow cytometer (BD Biosciences, Oxford, UK).

Chapter 3: Next Generation Sequencing of Roseoloviruses

3.1 Introduction

The dideoxy or chain termination method developed by Sanger and colleagues [Sanger et al., 1977] has underpinned DNA sequencing technology for the past three decades. Advances in the chemistry and technology associated with this technique [Prober et al., 1987; Panussis et al., 1998; Marsh et al., 1997], allowed automation which has increased throughput to a level that allows large complex genomes, such as the human genome, to be sequenced [Lander et al., 2001]. However, for whole genome sequencing this method suffers from a number of limitations. Requirements for the amplification of DNA fragments prior to sequencing typically involves bacterial cloning and electrophoretic separation of chain terminated fragment libraries before base detection; placing limitations on the throughput potential of this method. Therefore, whole genome Sanger sequencing remains a laborious and expensive process.

Attempts to overcome these limitations led to the development of next generation sequencing (NGS). NGS describes a group of diverse technologies unified by the rapid, low cost per base generation of large volumes of sequence read data when compared to traditional Sanger sequencing. In the decade since these technologies became commercially available, five major platforms have emerged: 454, Illumina, Ion Torrent, PacBio and SOLiD. While all of these platforms have their relative advantages and disadvantages [Reviewed in Mardis et al., 2013; van Dijk et al., 2014], the lowest cost per base and highest per run throughput displayed by the Illumina platform has contributed greatly to a dominant share in the NGS marketplace [Liu et al., 2012; Loman et al., 2012; van Dijk et al., 2014]. The introduction of Illumina MiSeq in 2011, a benchtop high-throughput sequencing platform, allowed this technology to become a viable option for smaller laboratory and in clinical settings. As such advances in these second-generation sequencing technologies have revolutionised DNA sequencing and are now beginning to allow such data to be available to a much wider user base. Which in turn has permitted much greater consideration of species at the whole genome level.

Due to their small size, sequencing of viral genomes is well within the remit of a single run from many NGS technologies. However, if this sequencing is performed directly from infected tissue samples, the presence of host genomic nucleic acid vastly complicates the process. In such cases, the levels of viral nucleic acid may be in a minority compared with host nucleic acid; proportions which will be reflected in resultant sequence reads. For certain applications this may represent sufficient viral sequence reads. However, in other cases, such as viral genomic sequencing, this will affect the ability to achieve the levels of read coverage required for accurate sequence resolution. This is especially an issue with restricted tissue infections or in characterisation of latent infections, such as that observed in herpesviruses. As well as for characterisation of the recently defined germ-line chromosomally integrated form of HHV-6A/-6B (CI-HHV-6A/-6B), under investigation in this thesis.

In these cases, methods for the isolation of viral DNA from the host nucleic acids prior to sequencing are advantageous [Mamanova et al., 2010; Radford et al., 2012]. This enrichment is typically achieved by *in vitro* viral culture, long-range PCR (LPCR) amplification of viral genomic DNA or hybridisation capture. Each of which have their relative advantages and disadvantages mostly centring around time, cost, failure to detect

large genomic changes or the introduction of mutations [Reviewed in Mertes et al., 2011]. However, of particular note in this regard, *in vitro* culture of the betaherpesvirus HCMV has been shown to rapidly alter the genetic content of the viral population [Dargan et al., 2010]. Which may have implications for attempts to sequence Roseoloviruses utilising this enrichment method. While all of these enrichment methods have been successfully utilised for the next-generation sequencing (NGS) of whole genomes of various other members of the *Herpesviridae* [Lin et al., 2013; Szpara et al., 2014; Cunningham et al., 2010; Donaldson et al., 2013; Depledge et al., 2011; Depledge et al., 2014; Renzette et al., 2011], at the start of this thesis, these technologies had not been applied to Roseoloviruses.

Additionally, different methods for the assembly of NGS sequencing reads can affect their interpretation. If a suitable reference genome is available, the sequencing reads may be aligned/mapped onto this genome and any variants used to generate a consensus sequence. Whilst this method has time advantages and less computational resource requirements. Issues arise with read alignment at regions of high divergence and novel genes/sequence may be missed [Scheibye-Asling et al., 2009]. Alternatively reads can be subjected to *de novo* assembly whereby specific algorithms are used to identify overlapping sequence in the reads permitting their assembly into larger contigs. This allows for better resolution of novel sequences, but also requires greater computational power and read depth coverage [Scheibye-Asling et al., 2009]. Additionally both of these sequence assembly methods suffer with the resolution of repetitive sequences [Treangen & Salzberg, 2011; Lee & Schatz, 2012], which can be found at a number of locations throughout HHV-6A/-6B genomes, most notably the T1 and T2 regions in the direct repeat (DR) regions of the genomic termini, the origin of lytic replication (*oriLyt*) towards the centre of the unique long (UL) region, and the R1, R2 and R3 regions located towards the right end of the UL region, indicated in figure 1.4. As with enrichment strategies, at the start of this thesis the utility of these NGS short read assembly methods for use in Roseoloviruses had not been determined.

In order to assess the effects of different NGS methods, comparisons were made between different enrichment and sequence assembly methods for the Illumina platform based NGS of HHV-6A genomic DNA. The aim was to establish a readily utilisable pipeline for viral genomic sequencing directly from cell or tissue samples. This would also be widely applicable to HHV-6A/-6B strains, and key to determine genomic sequences of the previously uncharacterised integrated forms of these viruses, CI-HHV-6A/-6B. These methods were first validated by resequencing the HHV-6A reference strain U1102. Before further validation through the application of this technology for the sequencing of a second previously unsequenced strain of HHV-6A, AJ. Also one of the first Roseolovirus isolates.

Available reports of complete HHV-6A/-6B genomes are currently limited. Complete genome sequences of two HHV-6B strains, the 162 kbp Z29 strain and the 162 kbp HST strain, have been determined (accession numbers AF157706 and AB021506, respectively) [Dominguez et al., 1999; Isegawa et al., 1999]. Z29 was also isolated from a HIV/AIDS patient from Zaire, now Democratic Republic Congo [Lopez et al., 1988], while strain HST was isolated from a Japanese infant during primary infection [Yamanishi et al., 1988], and represents the only primary childhood infection strain currently sequenced.

To date, genomes have been derived from HHV-6A strains isolated from reactivated infections in immunosuppressed patients. The first to be sequenced was strain U1102,

originally isolated from a Ugandan HIV/AIDS patient [Downing et al., 1987]. This was sequenced by Gompels and colleagues (accession number X83413) [Gompels et al., 1995] and subsequently has become the prototypical reference sequence for HHV-6A. After the start of this thesis, two versions of strain GS were derived (accession numbers KC465951 and KJ123690), only one of which (KC465951) was reported [Gravel et al., 2013a]. With GS originally being isolated from American/Jamaican patients with lymphoproliferative disease [Salahuddin et al., 1986]. HHV-6A strain AJ, was also one of the original HHV-6A isolates and was characterised in the UK from an adult HIV/AIDS patient from the Gambia [Tedder et al., 1987], however, at the start of this thesis the sequence of its genome had not been resolved.

Although the sequencing projects that gave rise to the strain GS genome sequences utilised NGS technologies, methods were not explicitly described. Importantly, no viral target enrichment methods were described, meaning these methods could not be readily applied to the sequencing of CI-HHV-6A/-6B genomes directly from clinical samples. Comparisons were made between all these fully sequenced HHV-6A and -6B genomes and the resequenced U1102 and strain AJ derived here.

3.2 Establishment of Next Generation Sequencing Protocols for HHV-6A Genomes

3.2.1 SureSelect Target Enrichment System

Solution hybrid selection has previously been utilised for the specific capture and subsequent second-generation sequencing of viral genomic DNA [Matranga et al., 2014]. Including from some members of the *Herpesviridae*, namely VZV, EBV and KSHV [Depledge et al., 2011]. However, its use for the enrichment of HHV-6A genomic DNA has not been reported. Therefore, initially attempts were made to investigate the validity of the use of solution hybrid selection to specifically capture HHV-6A genomic DNA from an infected cell, total DNA fragment library and then to compare it to a long-range PCR amplicon method of target enrichment. The enriched fragment libraries were then used for the determination of whole viral genomic sequence by Illumina MiSeq based NGS and read assembly. Initially the SureSelect methodology was applied to resequencing of the prototypical HHV-6A strain, U1102 (Accession X83413/RefSeq NC_001664), originally sequenced by Gompels et al., 1995 from plasmid, cosmid and lambda bacteriophage clones containing fragments covering the entire genome [Martin et al., 1991a]. This allowed an assessment of the accuracy of the solution hybrid selection and NGS methodology for HHV-6A strains. It also provided a means for confirmation of the strain U1102 genome sequence, confirmation of previously reported errors in the U1102 sequence (three substitutions in the DR6, U83 and U86 ORFs) [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002] and identification of any further errors or minor strain variants.

The SureSelect Target Enrichment System, was used due to its proven ability in the whole genome sequencing of other human herpesviruses from clinical samples [Depledge et al., 2011]. Briefly, following sequence library generation, overlapping, 120-mer biotinylated RNA baits, custom designed against known HHV-6A/-6B and HHV-7 genomes [Depledge et al., 2011; Donaldson et al., 2013] were used to capture HHV-6A strain U1102 genomic DNA from the JJhan infected cell total DNA sequence library, followed by a limited number of rounds of PCR to generate sufficient quantities of nucleic acid [Gnirke et al., 2009]. The enriched sequence library was then subjected to paired-end sequencing on an Illumina

MiSeq. Post-sequencing, quality trimmed read-pairs were mapped to the strain U1102 genome using Burrows-Wheeler Aligner-MEM algorithm (BWA-MEM) [Li & Durbin, 2009], with variant calling, read coverage analysis and consensus sequence generation performed using the Genome Analysis Toolkit (GATK) [McKenna et al., 2010; DePristo et al., 2011] as well as a SAMtools/BCFtools/vcfutils pipeline [Li et al., 2009; Danecek et al., 2011]. In addition, a VelvetOptimiser, Velvet [Zerbino & Birney, 2008] and ABACAS [Assefa et al., 2009] pipeline was also used to optimise *de novo* assembly of the trimmed read-pairs and subsequent contig assembly against the U1102 reference genome.

For the consensus sequences generated by both assembly methods, the Rapid Annotation Transfer Tool (RATT) [Otto et al., 2011] was used to transfer the annotation of the prototypical HHV-6A strain U1102 onto the assembled genomes sequences, these initial annotations were in turn corrected and updated using a combination of GeneMark ORF predictions [Besemer et al., 2001] and further annotation corrections found on HHV-6A and HHV-7 strains derived during the work towards this thesis [Donaldson et al., 2013, Gravel et al., 2013a; Megaw et al., 1998]. In addition, the original annotation of the HHV-6A strain U1102 (Accession X83413, RefSeq NC_001664) was also updated to reflect these corrections and thereby allowing for accurate comparisons between the archived and experimental genome sequences.

A total of 1,444,266 reads were generated from the MiSeq run, with a mean read length of 220.89 bp post trimming. For the read mapped assemblies 1,403,841 of these reads (97.20%) could be mapped to the U1102 reference genome with a mean mapping quality score, a Phred-scaled quality score of the mapping, of 52.95, 1,396,309 (96.68%) of which were properly paired. Unmapped read pairs were subjected to BLASTn searches [Camacho et al., 2009], which revealed they were of human genomic origin. Of the 40,425 unmapped reads, 97.4% shared homology with some of the viral bait sequences used for DNA capture, presumed to have been captured along with viral DNA due to the presence of human genomic-like sequences within the HHV-6A genome. The remaining 2.6% of unmapped reads were believed to result from incomplete post-hybridisation wash steps. Read depth coverage across the entire genome is shown in figure 3.1A. The mean read depth coverage across the genome was 1,557.36 (SD 481.21), with 100% of the genome covered at a read depth coverage greater than 10, and 99.95% at greater than 100 depth. Read mapping quality scores were also compared across the genome, figure 3.1B. Regions with lowest read depth and mapping quality scores were predominantly found in the direct repeats (DR) at the genomic termini and in the R2 and R3 regions within the unique long (UL) section of the genome. These included regions with higher G+C content, which are known to be under-represented in Illumina sequencing results [Benjamini & Speed, 2012].

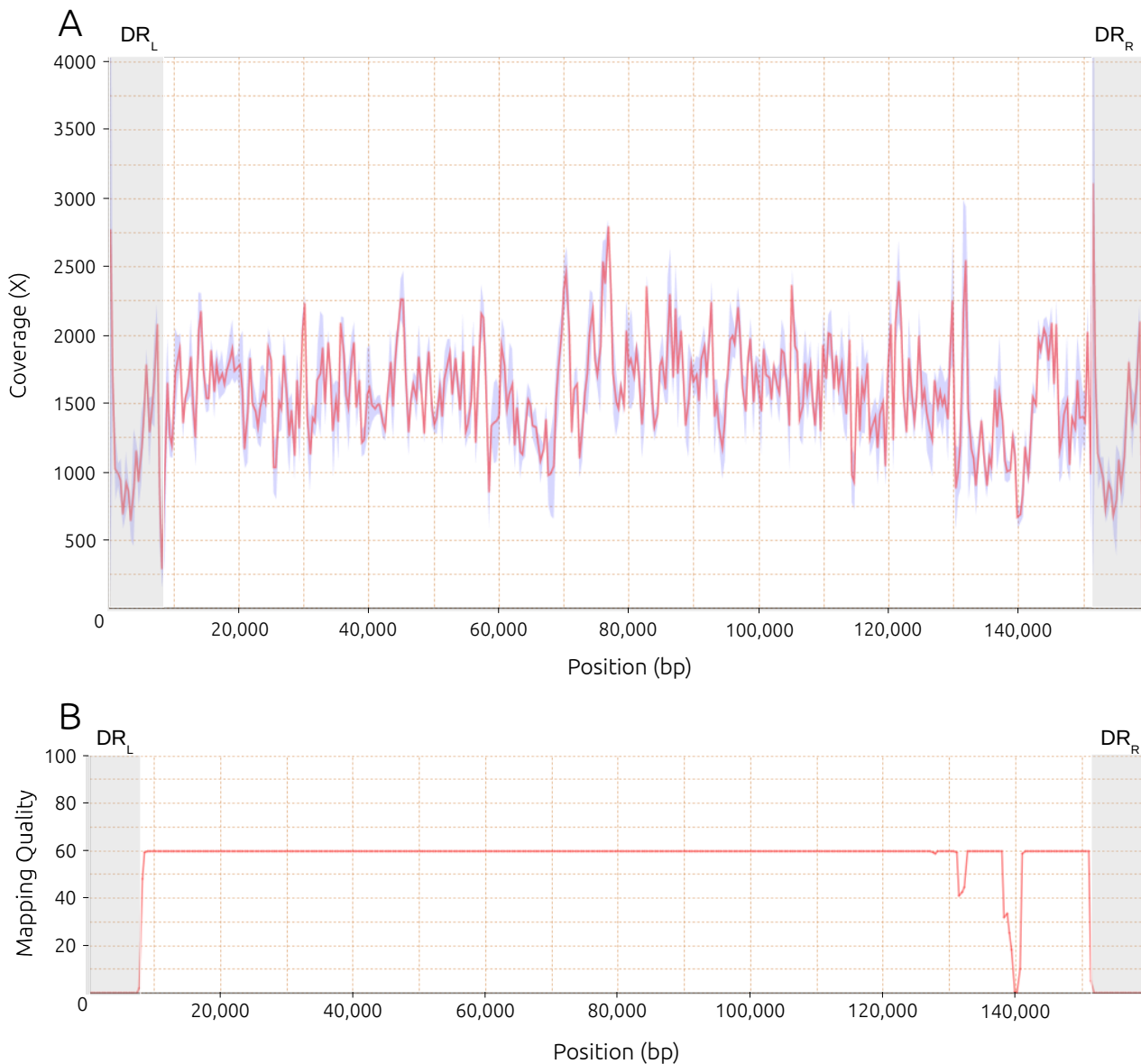


Figure 3.1 Coverage and mapping quality of SureSelect enriched HHV-6A strain U1102 sequence reads across the X83413 reference genome. (A) Coverage of SureSelect HHV-6A strain U1102 reads across the reference genome. SureSelect target enriched libraries from infected cell DNA were sequenced by Illumina MiSeq and mapped to the U1102 reference genome using the BWA MEM algorithm. (B) Mapping quality scores of sequence reads across the U1102 genome. Direct repeat regions of the genome are shaded and labelled DR_R and DR_L, accordingly.

The *de novo* assembly of read pairs resulted in contigs covering 95.24% (151743 bp) of the U1102 reference genome, with gaps occurring in repetitive sequences: a 1233 bp gap in R2, a 2287 bp gap in R3 and the remaining gaps in the DR around the T1 and T2 human telomeric repeat-like sequences. All previously predicted open reading frames were covered by the assembled contigs.

Consensus sequences generated from both the reference mapped and *de novo* assemblies were compared to each other and the NCBI genome reference sequence (HHV-6A strain U1102, accession X83413). The mapped assembly consensus sequence showed 99.98% identity with that of the NCBI reference sequence. The *de novo* assembly consensus

sequence shared an overall nucleotide identity of 99.62% with that derived from the mapping assembly. The main sequence variation was in the repetitive sequence regions, including the human telomere-like repeat regions of the DRs and in R2 and R3. Since these repetitive sequences confound assembly from reads shorter than the repeat units, these regions were removed to check assembly of unique sequence. Following this, the mapped and *de novo* assemblies shared 99.99% nucleotide identity, providing validation of this NGS method for the derivation of HHV-6A genomes. This consensus was then compared to the NCBI HHV-6A reference genome and showed 99.98% identity, some of this variation was expected due to the identification of previously defined or suggested corrections to DR6, U83 and U86 [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002] still present in the NCBI reference genome. However, these did not account for all the variation observed, additional SNPs were identified which are discussed further in section 3.3.

3.2.2 Long-range PCR amplicons

Since the SureSelect Target Enrichment System (Agilent Technologies) relies on previously determined sequence to bait new genomes, an additional method of enrichment was established to allow checks for further sequence variation in HHV-6A/-6B genomes. This was particularly important for the goal of sequencing CI-HHV-6A/-6B genomes, since at the outset of this thesis the genetic distance of these integrated viral forms to infectious virus genomes was not known. Therefore, an alternative method was also used which involved a whole genome long-range PCR (LPCR) amplicon based approach for the enrichment of viral genomic DNA. Both methods were compared to derive the complete genome of the previously unsequenced HHV-6A laboratory strain, AJ. The SureSelect Target Enrichment System was performed as described above. For the LPCR approach, enrichment of HHV-6A viral genomic DNA from JJhan infected cell DNA involved the use of primer sets, designed against the U1102 genome sequence, to generate 36 overlapping amplicons, ranging in size from 1 – 7 kbp, spanning the entire HHV-6A genome (methods section 2.3.1). These amplicons were then purified and pooled in equimolar ratios, prior to shearing and sequence library preparation. Again, cluster generation and sequencing was performed on an Illumina MiSeq. With consensus sequences generated by both mapping to the HHV-6A strain U1102 reference sequence and *de novo* assembly as described for the resequencing of strain U1102 above.

Paired-end sequencing of the SureSelect enriched AJ library generated a total of 1,937,791 reads, with a mean read length of 196.49 bp. For mapped assemblies, 1,914,498 (98.8%) of these reads mapped to the U1102 reference genome with a mean mapping quality of 52.52, 1,912,009 (98.7%) which were properly paired. Mean read coverage across the reference genome was 2,119.84 (SD 719.95), with 99.7% and 98.4% of the genome covered at read depths greater than 1 and 100, respectively (figure 3.2). *De novo* assembly of these reads generated contigs covering 90.2% of the U1102 reference genome.

The LPCR library generated 2,790,709 reads, with a mean read length of 224.31 bp. 2,787,945 (99.9%) of these reads mapped to the U1102 reference genome with a mapping quality of 52.67, where 2,787,689 (99.9%) of these reads were properly paired. The mean read coverage was 2,581.51 (SD 1,439.74), with 99.5% and 95.5% of the genome covered at read depths greater than 1 and 100, respectively (figure 3.3). Although the mapping quality scores across the genome from both the SureSelect and LPCR

methods were similar, there was greater read coverage variance across the genome with the LPCR generated library, predominantly coinciding with genomic amplicon locations. This appeared to be from variances in the equimolar pooling of the genomic amplicons prior to sequence library generation. However, despite these differences, coverage was still sufficient to allow determination of the genomic sequence. *De novo* assembly of these read pairs generated contigs covering 90.35% of the U1102 reference genome, with gaps primarily from the repeat regions described.

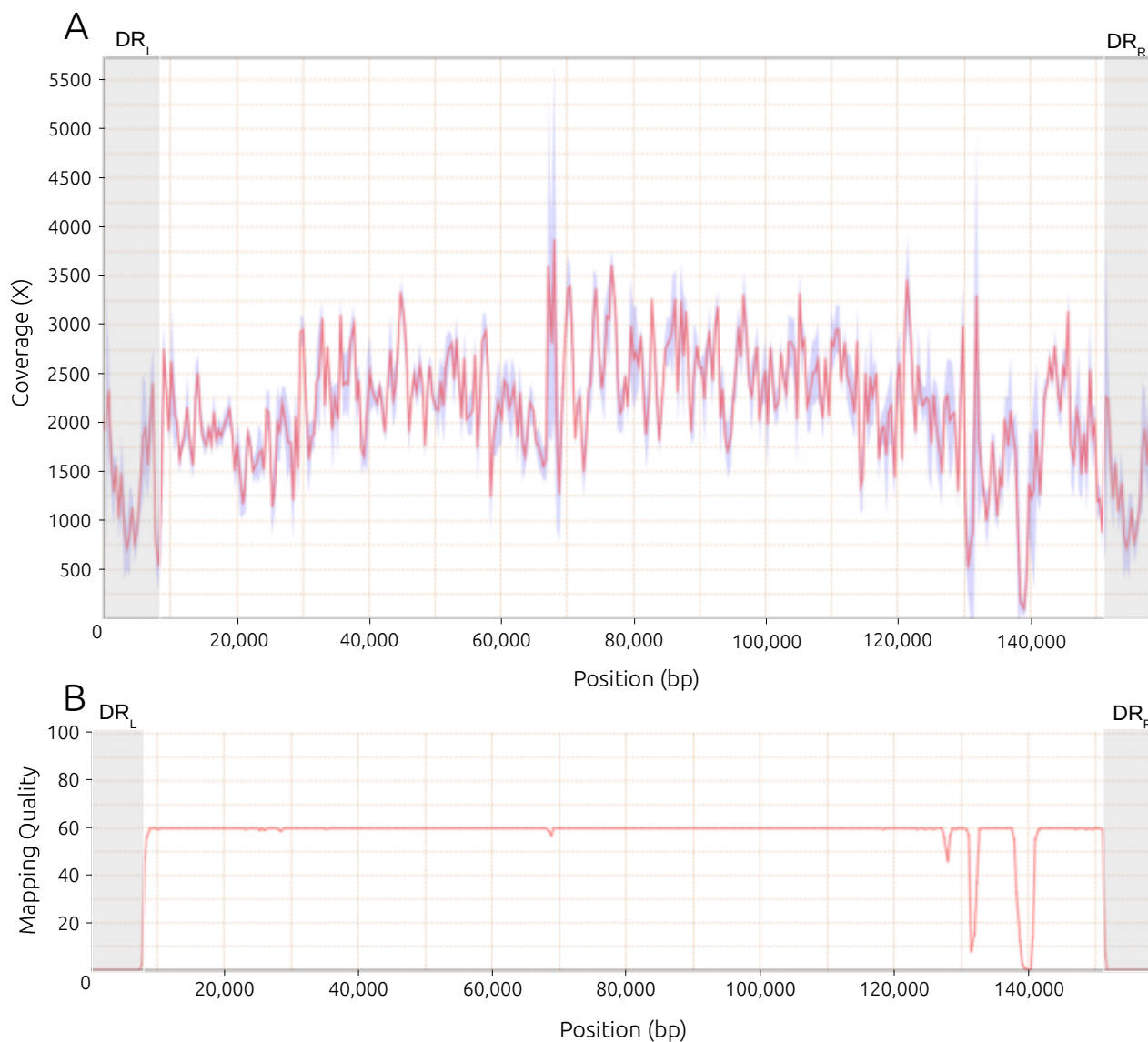


Figure 3.2 Coverage and mapping quality of SureSelect enriched HHV-6A strain AJ sequence reads across the HHV-6A strain U1102 reference genome. (A) Coverage of SureSelect HHV-6A strain AJ reads across the U1102 reference genome. SureSelect target enriched libraries from infected cell DNA were sequenced by Illumina MiSeq and mapped to the U1102 reference genome using the BWA MEM algorithm. (B) Mapping quality scores of reads across the U1102 genome. Direct repeat regions of the genome are shaded and labelled DR_R and DR_L, accordingly.

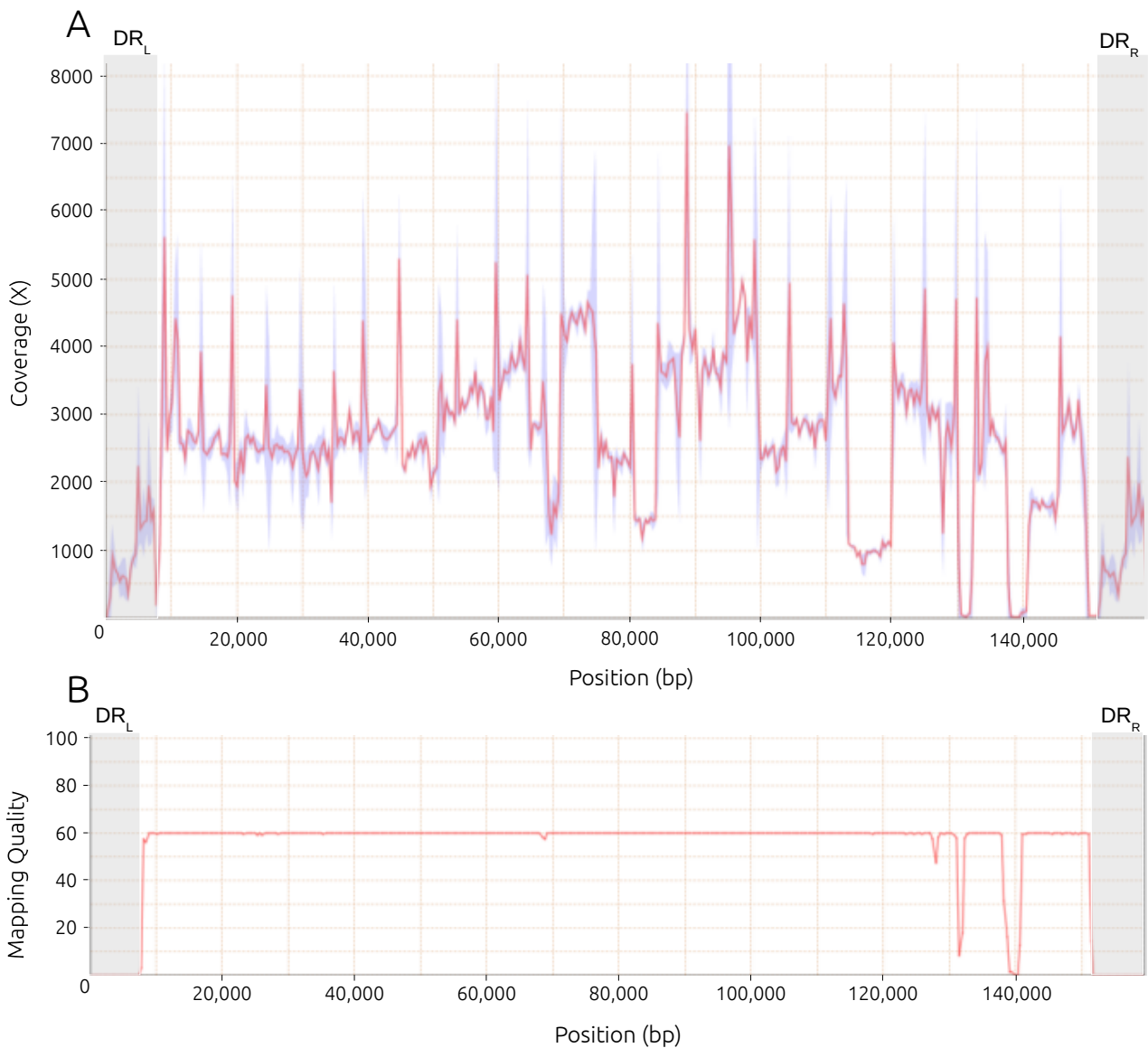


Figure 3.3 Coverage and mapping quality of LPCR enriched HHV-6A strain AJ sequence reads across the HHV-6A strain U1102 reference genome. (A) Coverage of LPCR HHV-6A strain AJ reads across the U1102 reference genome. LPCR derived libraries from infected cell DNA were sequenced by Illumina MiSeq and mapped to the U1102 reference genome using the BWA MEM algorithm. (B) Mapping quality scores of reads across the U1102 genome. Direct repeat regions of the genome are shaded and labelled DR_R and DR_L, accordingly.

The accuracy of both enrichment methods was compared. For this each method was compared separately to a combined consensus sequence determined using all available reads from both methods and additional Sanger sequencing verification of divergent or gap regions, as described in below in section 3.4. Using U1102 reference mapping, the consensus sequences derived from mapping of the SureSelect enriched and LPCR strain AJ sequencing reads to the U1102 genome shared 99.2% and 99.3% identity with the final consensus AJ genome, respectively. Using *de novo* assembly, the SureSelect and LPCR reads shared 98.1% and 98.3% identity, respectively. Again discrepancies between the mapped and *de novo* assemblies were confined to repetitive regions: the DRs and R3. With the exclusion of these repetitive regions, the SureSelect reference mapped and *de novo* assemblies shared 100% identity with each other. Similarly, the LPCR enrichment method

also shared 100% identity between the reference mapped and *de novo* assemblies. This showed both enrichment methods assembled equally well using both reference mapped and *de novo* assembly. However, in both cases the sequencing pipeline depended on Sanger sequencing of gaps and repetitive regions for completion of the full genome sequence.

Finally, to determine whether all previously defined genes could be resolved accurately by mapping assemblies, CDS nucleotide identity comparisons were made between the different reference mapped assemblies and the final resolved sequence of AJ. For this, all consensus sequences were again annotated to reflect recent experimentally defined annotation corrections, as described in section 3.2. Before pairwise comparisons and calculations of read coverage were made between all gene CDS, table 3.1. This indicated mapped assemblies from both SureSelect and LPCR libraries could accurately resolve all gene CDS across the UL region of the genome, even in genomic regions with increased variation, at the right end of the genome [Gompels & Kasolo, 2006]. As a further control, a number of these genes, including a set of defined variable genes, were compared to sequences determined by Sanger capillary sequencing (partly described in sections 5.2 and as subsequently published in full in Tweedy et al., 2015a). These also shared 100% identity with those generated by the mapped assemblies from both library generation methods. Therefore, SNP differences identified in the strain AJ sequence were confirmed using all three methods.

CDS	AJ LPCR mapped			AJ SureSelect mapped		
	% Identity	Mean Coverage	Minimum Coverage	% Identity	Mean Coverage	Minimum Coverage
DR1	95.75	1616.45	0	95.75	1494.26	0
DR6	95.81	4027.05	18	95.81	1464.84	1
U2	100	9505.95	5644	100	2416.51	1755
U3	100	9321.60	601	100	2224.82	219
U4	100	5820.74	4498	100	1884.06	1484
U7	100	5545.52	1823	100	1973.61	926
U10	100	4700.43	3084	100	1989.17	1629
U11	100	5076.88	2145	100	1586.60	928
U12	100	5042.46	4236	100	1690.05	1368
U13	100	4382.09	4003	100	1543.87	1270
U14	100	5388.61	2460	100	1868.92	1351
U15	100	4112.03	1157	100	1263.47	1110
U17	100	4625.00	3697	100	2084.87	1463
U18	100	4397.42	3878	100	1688.72	799
U19	100	5208.70	2517	100	2639.69	1698
U20	100	4491.90	2927	100	2003.99	1134
U21	100	4631.86	1814	100	2539.66	1276
U22	100	4785.69	4199	100	2563.95	2217
U23	100	5136.92	1320	100	2095.59	1030
U24	100	5417.96	5004	100	2509.66	1894
U24A	100	5514.00	5296	100	3266.36	3179
U25	100	5027.83	4400	100	2477.28	1928
U26	100	5521.15	4685	100	2726.60	1959
U27	100	5424.07	3986	100	2382.95	1547
U28	100	5805.10	2539	100	2130.84	1130
U29	100	5560.68	4887	100	2239.81	1759
U30	100	5845.46	4036	100	2489.21	1715
U31	100	4672.35	2744	100	2361.73	1468
U32	100	4939.60	3748	100	2185.78	1636
U33	100	6107.34	5187	100	2653.77	2037
U34	100	7167.45	5360	100	2685.18	2181
U35	100	5955.16	5483	100	1962.09	1702
U36	100	6071.62	5107	100	2227.00	1605
U37	100	6667.73	6297	100	2437.44	1685
U38	100	6001.78	2240	100	2166.27	694
U39	100	7785.20	3348	100	2272.09	1629
U40	100	7174.70	6335	100	1945.55	1372
U41	100	5667.65	2491	100	2093.45	1262
U42	100	7272.84	2549	100	2735.82	1503
U43	100	8212.30	7262	100	2050.77	1189
U44	100	8861.19	8134	100	2452.38	1981
U45	100	8024.43	3774	100	3065.64	2331
U46	100	4444.31	4139	100	2308.01	2149

CDS	AJ LPCR mapped			AJ SureSelect mapped		
	% Identity	Mean Coverage	Minimum Coverage	% Identity	Mean Coverage	Minimum Coverage
U47	100	4427.41	1490	100	3117.86	1467
U47A	100	3766.19	2954	100	2216.86	1758
U48	100	4502.69	3929	100	2440.31	1842
U49	100	4714.56	2430	100	2589.36	1852
U50	100	2540.71	1446	100	2376.88	1256
U51	100	2769.28	2427	100	2965.15	2005
U52	100	2589.15	2154	100	2055.14	1626
U53	100	7550.19	2361	100	2624.11	1921
U53.5	100	7509.09	6653	100	2825.59	2508
U54	100	6824.40	2798	100	2992.98	1821
U55	100	7732.69	977	100	2702.88	1540
U56	100	8619.75	5835	100	2512.97	1598
U57	100	7594.49	2720	100	2601.16	1282
U58	100	9268.91	5134	100	2211.52	1359
U59	100	9204.25	7576	100	2985.29	2281
U60	100	6411.49	3996	100	2348.60	1559
U62	100	7991.17	7236	100	2538.09	2258
U63	100	8359.30	7321	100	2479.87	1645
U64	100	5774.55	3553	100	2266.40	1318
U65	100	4425.32	4026	100	2560.70	2059
U67	100	4423.31	3832	100	2454.01	1962
U68	100	4956.52	4618	100	2897.55	2323
U69	100	6141.39	3140	100	2753.28	1556
U70	100	5276.62	4143	100	2461.10	1756
U71	100	5341.32	5079	100	2833.85	2555
U72	100	5366.70	4400	100	2486.93	1997
U73	100	5950.11	3671	100	2454.06	1665
U74	100	6995.05	5427	100	2729.20	2137
U75	100	7722.06	2058	100	2347.65	1849
U76	100	1929.96	1375	100	2030.75	1022
U77	100	1707.88	787	100	2190.08	1303
U79	100	6757.10	1652	100	2629.25	1268
U81	100	5944.96	5067	100	2691.77	1983
U82	100	6073.80	5451	100	2261.54	1725
U83	100	5477.55	5088	100	2307.49	2123
U84	100	5978.56	52	100	2209.75	1355
U85	100	7669.00	28	100	2048.19	109
U86	100	5193.77	4450	100	1881.77	128
U90	100	5287.02	3040	100	1331.07	644
U91	100	4419.95	2939	100	1737.33	1156
U94	100	2986.21	2250	100	1936.06	1052
U95	100	4080.97	2229	100	2528.45	1744
U100	100	4358.26	40	100	1882.68	16

Table 3.1 Percentage identity between HHV-6A CDS features from mapped assemblies methods with the complete sequence of HHV-6A strain AJ. Comparison of CDS with final AJ sequence resolved below. Emboldened denotes also Sanger sequenced with 100% identity.

Taken together, these results show that solution hybrid capture and LPCR both represent viable means for target enrichment of HHV-6A genomic DNA. As for assembly, the relatively low level of diversity observed to date amongst HHV-6A strains permits a reference based mapping assembly of the UL region of the genome. However, with this method issues are encountered with resolution of a number of repetitive regions. At these sites a *de novo* assembly method provides better, albeit incomplete, sequence resolution. However, currently determining the sequence of these variable and repetitive regions will require complementary Sanger sequencing to ensure accurate resolution. Because of these repetitive regions and the additional analyses required, this remains a bottleneck to the routine use of NGS for whole genome sequencing of HHV-6A strains.

3.3 Resequencing of HHV-6A Strain U1102 Sequence

During resequencing of the HHV-6A strain U1102 genome, section 3.2, it was noted a number of discrepancies existed between the NCBI archived U1102 reference sequence and the NGS derived consensus sequences. Some of these confirmed sequencing errors previously identified in DR6, U83 and U86 [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002]. These corrections were all identified during

the re-sequencing, as shown in figure 3.4. However, for DR6 the corrections reported by Schleimann et al., 2014, were only identified by the *de novo* assembly resulting from the issues already identified relating to the mapping of reads to repetitive regions as exemplified by the mapping quality scores across these regions, figures 3.2 and 3.3. This shows a combination of *de novo* and reference mapping is required to determine the sequence of Roseolovirus genomes.



Figure 3.4 Confirmation of previously identified corrections to the HHV-6A strain U1102 genome sequence. (A) DR6 - G addition at positions 6664 and 6666 resulting in frame shift early stop codon as seen in other HHV-6A strains and HHV-6B, could only be identified from the *de novo* assembly consensus sequence (B) U83 – GC inversion at position 123740. (C) U86 – G addition as position 128132, subsequently corrected in Refseq version (NC_001664), brings originally defined ORFs U86 and U87 into one long ORF. Encoded on minus strand, CDS shown 5' to 3' here.

A further five differences were noted between the archived nucleotide sequence and the NGS generated consensus sequences. These were located within the coding sequences of the U40, U42, U57, U58 and U100 open reading frames, figure 3.5. Blast searches [Camacho et al., 2009] of these re-sequenced open reading frames revealed sequences with 100% nucleotide identity to two of them, U40 and U42, were already present on the NCBI database. These were from genomic fragment sequencing of strain U1102 predating the publication of the complete genome sequence found under the accession numbers L20954 [Jones & Teo, 1993] and X92436 [Jones, unpublished], respectively. These were all nonsynonymous mutations and may represent mixed populations in the original isolate strain U1102, or sequencing errors.

A

```

X83413 U40      AAATATTTTGAAAAGGTGTGTGTTTGAGGCGTTGTGCTGAGACCTTGTCAATAAACATT
NC_001664 U40  AAATATTTTGAAAAGGTGTGTGTTTGAGGCGTTGTGCTGAGACCTTGTCAATAAACATT
Mapped U40     AAATATTTTGAAAAGGTGTGTGTTTGAGGCGTTGTGCTGAGACCTTGTCAATAAACATT
De novo U40    AAATATTTTGAAAAGGTGTGTGTTTGAGGCGTTGTGCTGAGACCTTGTCAATAAACATT
                *****.*****

Reference      KYFEKVLCLRRCAETLSINI
Re-sequence   KYFEKVLCLRRCAEHLINI
                *****

B

X83413 U42      TTTTCGAGGAAAACCTCGCACCTCTGAAACA-CTG--TGAATCCAGACTAGACGCACTAAGT
NC_001664 U42  TTTTCGAGGAAAACCTCGCACCTCTGAAACA-CTG--TGAATCCAGACTAGACGCACTAAGT
Mapped U42     TTTTCGAGGAAAACCTCGCACCTCTGAAACAACCTGGATGAATCCAGACTAGACGCACTAAGT
De novo U42    TTTTCGAGGAAAACCTCGCACCTCTGAAACAACCTGGATGAATCCAGACTAGACGCACTAAGT
                ***** **

Reference      FRGKLAPLKHK-ESRLDALS
Re-sequence   FRGKLAPLKQDESRLDALS
                *****:*****

C

X83413 U57      CATGCAAAGACGAACTGCCAAAAGTACGACTTCACCCCTCAAATTTGTCACCGAAAGCTTG
NC_001664 U57  CATGCAAAGACGAACTGCCAAAAGTACGACTTCACCCCTCAAATTTGTCACCGAAAGCTTG
Mapped U57     CATGCAAAGACGAACTGCCAAAAGTACGACTTCACCCCTCAAATTTGTCACCGAAAGCTTG
De novo U57    CATGCAAAGACGAACTGCCAAAAGTACGACTTCACCCCTCAAATTTGTCACCGAAAGCTTG
                *****

Reference      HAKTNCQKYDFTLKIVTESL
Re-sequence   HAKTNCQNYDFTLKIVTESL
                *****:*****

D

X83413 U58      GAGGAAACTCGTTTAGCGCGATCCATGCCGTGAGTTCTGGAAAAGATCGAGTCAATCATA
NC_001664 U58  GAGGAAACTCGTTTAGCGCGATCCATGCCGTGAGTTCTGGAAAAGATCGAGTCAATCATA
Mapped U58     GAGGAAACTCGTTTAGCGCGATCCCGTCCGTGAGTTCTGGAAAAGATCGAGTCAATCATA
De novo U58    GAGGAAACTCGTTTAGCGCGATCCGTGCCGTGAGTTCTGGAAAAGATCGAGTCAATCATA
                *****.*****

Reference      EETRLARSMPPVLEKIESII
Re-sequence   EETRLARSVPPVLEKIESII
                *****:*****

E

X83413 U100     AAACCACTGACCGCGATGACGGCGATACGTTTCTGTCCAATGACTACTAAATTAGAGCTA
NC_001664 U100 AAACCACTGACCGCGATGACGGCGATACGTTTCTGTCCAATGACTACTAAATTAGAGCTA
Mapped U100    AAACCACTGACCGCGATGACGGCGATAGCTTTCTGTCCAATGACTACTAAATTAGAGCTA
De novo U100   AAACCACTGACCGCGATGACGGCGATAGCTTTCTGTCCAATGACTACTAAATTAGAGCTA
                *****

Reference      KPLTAMTAIRFCPMTTKLEL
Re-sequence   KPLTAMTAIAFCPMTTKLEL
                *****

```

Figure 3.5 Further sites of variation identified by resequencing of HHV-6A strain U1102. (A) U40 - AC to CA at position 63873 (B) U42 – ACTGT to AACTGGAT starting at position 70451 (C) U57 – G to C at position 92149 (D) U58 – A to G at position 94068 (E) U100 – CG to GC at position 149552. All shown as CDS 5' to 3'. Amino acid sequence changes shown below nucleotide sequences.

3.4 Construction of the Complete Genome Sequence of HHV-6A Strain AJ

To generate the final consensus, sequence reads from both SureSelect and LPCR enrichment methods were combined to increase sequence read coverage across any low coverage regions, aiding *de novo* genome assembly. These combined sequence reads were

then used to generate consensus sequences by both *de novo* and reference mapped assembly, with comparisons made between the consensus sequences for validation. *De novo* assembly resulted in a total of 93.6% sequence resolution (as compared to strain U1102), with gaps only in the repetitive regions. The *de novo* consensus sequence of the UL region shared 100% identity with that of the reference mapped consensus with no gene insertions or deletions identified.

Remaining gaps and ambiguous regions were resolved by Sanger sequencing, using the primers described in section 2.3.2. Also, due to the locations of the primers used for generation of the long range PCR amplicons, noted difficulties with *de novo* assembly of repetitive sequences from short read generating NGS technologies and the inability to amplify a concatemeric junction sequence for strain AJ; the first 715 bp of direct repeat left (DR_L) and the final 1381 bp of direct repeat right (DR_R) could not be resolved using conventional methods. However, these could be determined due to the direct repeat nature of the genomic termini, as summarised in figure 1.4. The expected sequences of these regions were derived from the corresponding fully sequenced region from the opposite termini. The addition of these sequences to the *de novo* assembled draft genome allowed for resolution of the complete strain AJ genome sequence. In turn, the original sequencing read-pairs were mapped to the newly generated AJ genome sequence using BWA-MEM and SAMtools [Li & Durbin, 2009; Li et al., 2009], with BLAST searches [Camacho et al., 2009] of unmapped read-pairs revealing only human genomic and bacterial sequences and nothing with significant homology to herpesviral sequences. This demonstrated that no insertions had been missed during assembly. Annotations were derived as described in section 3.2, but also guided by GeneMark ORF predictions [Besemer et al., 2001]. The complete genome sequence of HHV-6A strain AJ, was subsequently reported in Tweedy et al., 2015b, and has been deposited in GenBank under the accession number KP257584 (a gene table is also in appendix 7.2).

3.5 Analyses of the Complete Genome of HHV-6A Strain AJ

The AJ genome is 156,714 bp in length, maintaining a typical class A herpesviral genomic organisation consisting of a 140,401 bp unique long region flanked by 8156 bp direct repeats (DR_R and DR_L). During determination of this genomic sequence, two reports of HHV-6A strain GS was published. These were included in the comparisons of HHV-6A/-6B strains. Strain AJ is the shortest of the HHV-6A strains sequenced to date (strain U1102 159,322 bp; strain GS_1 KC465951 156,864; strain GS_2 KJ123690 156,885 bp). Most of the sequence length disparity is from deletions of repetitive sequences in the R2 and R3 repeat regions of the genome. Phylogenetic analyses showed strain AJ shares the closest overall nucleotide identity with the low passage North American HHV-6A isolate GS (99.1%), followed closely by HHV-6A strain U1102 (98.42%). While the HHV-6B strains Z29 and HST share 92.41% and 91.79% identities, respectively, figure 3.6.

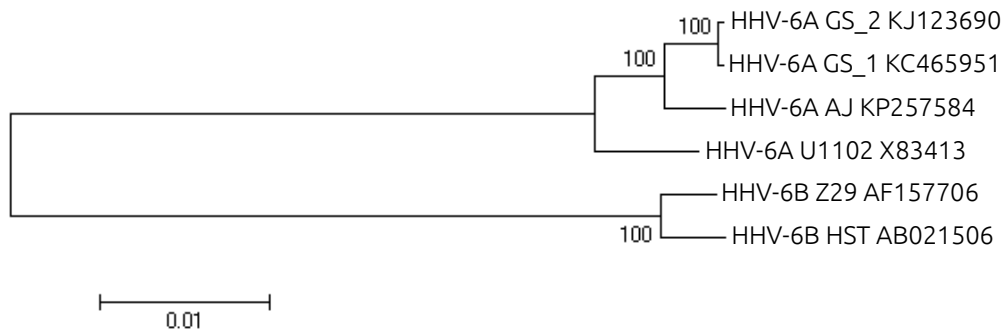


Figure 3.6 Phylogenetic analyses of HHV-6A strain AJ to compared to sequenced HHV-6A/-6B strains. Whole genome comparisons using maximum likelihood estimation and bootstrapping (1050 replicates). Accession numbers of all HHV-6A/-6B strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site.

All 85 genes previously identified in HHV-6A strains, U1102 and GS were present in AJ. ORF prediction software did not indicate the presence of additional genes, using a cut-off ORF size of 100 amino acids. Nucleotide comparisons between gene CDS of strains AJ and GS (table 3.2), revealed divergence at genes encoding glycoproteins, products involved in DNA replication and immunomodulation. However, a number of these, DR1 and DR6 [Borenstein et al., 2010], U21 and U24 [Jasirwan et al., 2015], are known to be dispensable for viral replication *in vitro*. Therefore effects resulting from passage of virus in tissue culture cannot be ruled out. Thus, despite their distinct geographic origins in the USA and Africa, HHV-6A strains AJ and GS are very closely related.

Pairwise comparisons to HHV-6A AJ							
	GS_1	GS_2	U1102		GS_1	GS_2	U1102
CDS	%	%	%		CDS	%	%
DR1	98.15	98.29	95.75		U47	99.59	99.59
DR6	97.88	98.13	95.81		U47A	100.00	100.00
U2	100.00	99.91	98.64		U48	99.95	99.95
U3	99.73	98.48	98.93		U49	99.87	99.87
U4	100.00	100.00	99.75		U50	99.76	99.76
U7	99.92	99.92	99.32		U51	99.67	99.67
U10	99.69	99.62	99.01		U52	99.87	99.87
U11	99.66	99.66	97.86		U53	99.75	99.75
U12	99.90	99.90	98.95		U53.5	99.86	99.86
U13	100.00	100.00	99.69		U54	98.84	98.84
U14	99.89	99.89	99.56		U55	99.38	99.38
U15	100.00	100.00	97.57		U56	100.00	100.00
U17	99.70	99.70	98.71		U57	99.80	99.78
U18	99.77	99.77	98.07		U58	99.48	99.48
U19	99.83	99.91	97.86		U59	99.81	99.81
U20	99.92	99.92	98.90		U60	99.85	99.85
U21	98.39	98.39	99.46		U62	99.61	99.61
U22	99.67	99.67	99.84		U63	99.69	99.69
U23	98.73	98.73	99.02		U64	99.17	99.17
U24	98.48	98.48	98.48		U65	99.70	99.70
U24A	100.00	100.00	100.00		U67	99.81	99.81
U25	99.58	99.58	99.47		U68	99.42	99.42
U26	99.55	99.55	99.55		U69	99.11	99.11
U27	99.07	99.07	99.07		U70	99.73	99.73
U28	99.79	99.79	99.30		U71	100.00	100.00
U29	99.89	99.89	100.00		U72	99.90	99.90
U30	99.85	99.85	99.38		U73	99.74	99.70
U31	99.82	99.82	99.49		U74	99.85	99.85
U32	100.00	100.00	100.00		U75	99.87	99.87
U33	99.65	99.65	99.93		U76	99.79	99.79
U34	99.04	98.92	99.76		U77	99.68	99.68
U35	100.00	100.00	100.00		U79	96.62	96.70
U36	99.52	99.52	99.66		U81	99.74	99.74
U37	99.87	99.87	99.50		U82	99.34	99.34
U38	99.74	99.74	99.05		U83	98.61	98.61
U39	99.60	99.60	99.36		U84	98.25	98.25
U40	99.22	99.22	99.17		U85	98.74	98.74
U41	99.91	99.91	99.32		U86	98.85	98.85
U42	99.94	99.94	98.71		U90	98.69	98.69
U43	99.81	99.81	99.65		U91	99.35	99.35
U44	99.84	99.84	99.35		U94	99.52	99.52
U45	99.82	99.82	99.56		U95	98.90	98.90
U46	99.61	99.61	98.04		U100	98.02	98.17

Table 3.2 Percentage identity between defined CDS of HHV-6A strain AJ and sequenced HHV-6A strains.

Of particular note, was variation observed in the DNA polymerase gene, U38. Comparisons with U38 from the other fully sequenced HHV-6A genomes, strains U1102 and GS, identified variation, figure 3.7, at a site which has been used to develop a HHV-6 subtype-specific, probe-based, real-time PCR (SSPBRT-PCR) species typing assay [Lou et al., 2011]. While the real-time PCR probe site was identical for all fully sequenced HHV-6A strains, variation was noted in both GS and AJ at both primer amplification sites. Thus the potential for reduced primer binding affinity, could confound the results of this assay through under diagnosis of HHV-6A infection.

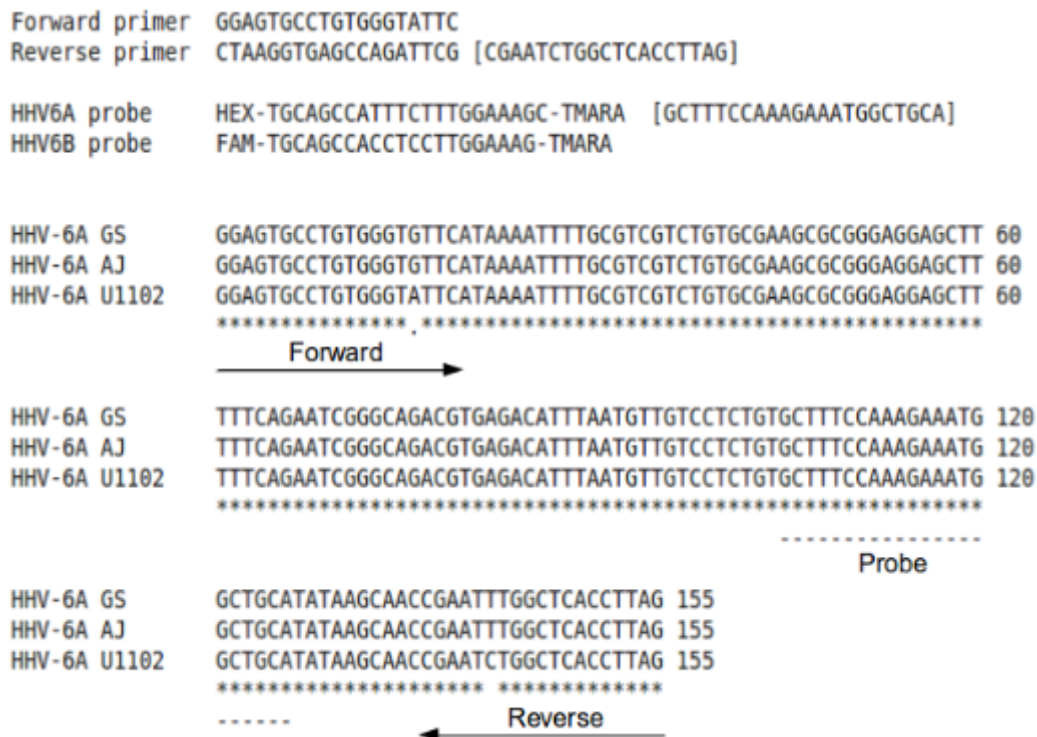


Figure 3.7 Variation in the DNA polymerase genes of HHV-6A strains and potential effects on the SSPBRT-PCR species typing assay. (A) SSPBRT-PCR primer and probe sequences (B) Alignment of U38 genes of fully sequenced HHV-6A strains at the SSPBRT-PCR expected amplification site. Primer locations are labelled with solid arrows, probe location is labelled as a dashed line.

3.6 Discussion

In recent years, the advances in second generation sequencing technologies have permitted its widespread use for the low cost (per base) generation of large volumes of sequence read data, due to its massively parallel throughput approach [Koboldt et al., 2013]. Along with advances in an array of areas of biological research, the availability of this data has begun to allow much greater consideration of the genetic variation of viral species at the whole genome level. The small size of viral genomes compared to the high levels of sequence read data achievable by even a single sequencing run by NGS technologies, suggests that whole genome sequencing should be a straight forward task. However, the obligate intracellular nature of viruses complicates such sequencing attempts due to the presence of host cell nucleic acid. While the viral genome copy number per millilitre in clinical samples can be in the millions for actively replicating virus, this is still in a minority when compared to levels of host nucleic acid. Therefore simply generating a NGS total DNA fragment library from such a sample results in high proportions of host nucleic acid and low proportions of viral nucleic acids; which is reflected in resultant sequencing reads. Thus issues can be encountered with achieving the levels of sequence read coverage across the viral genome required to permit whole genome assembly. Such problems can be overcome through *in vitro* viral culture to increase quantities of viral nucleic acid. However, *in vitro* culture may not be possible, and even if possible has the potential to rapidly alter the genetic content of a viral population, a feature which is observed extensively in a number of herpesviruses, including the closely related betaherpesvirus HCMV [Tyler et al., 2007; Dargan et al., 2010]. Thus potentially confounding the identification of genetic characteristics of the virus that define its

biological and health effects. The relative amounts of viral to host nucleic acid can be even further skewed towards the host in cases of restricted tissue infections or in the characterisation of latent infections, such as that observed in herpesviruses. Including the potentially latent germ-line integrated form of HHV-6A and -6B, CI-HHV-6A and -6B, that is the major subject of this thesis. In these cases methods for the enrichment of viral genomic DNA from cellular samples are often a prerequisite in order to achieve accurate viral sequence data, as well as helping to minimise reagent, computational resource and time costs. This is especially important where sample is limited or viral culture is not possible, again an issue with the characterisation of CI-HHV-6A/-6B from clinical samples from patients with integrated viral genomes. In order to accurately characterise these integrated genomes new methodology is required.

Here the use of two of the major approaches for target enrichment, in solution hybrid capture and long range PCR, were employed and assessed for use in the Illumina MiSeq mediated sequencing of HHV-6A strains. The aim being that these methods could then be used to characterise other Roseoloviruses directly from tissue samples, in particular to begin to allow characterisation of CI-HHV-6A and -6B. The results presented here demonstrate that both these methods could accurately determine HHV-6A genomic sequences. The accuracy was demonstrated with almost 100% identity of the resequenced U1102 genome with that of the NCBI reference genome [Gompels et al., 1995]. There were only 13/159,321 bp differences (99.99% identity). Three of the sites of difference agreed with previously reported base substitutions, two causing frame-shifts and one a GC inversion which currently remain in the NCBI archived reference genome sequence [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002]. Further differences were identified within 5 additional ORFs. These could represent sequencing errors from either the NGS or Sanger methods utilised. Or may be minor variants in the original isolate strain of U1102. Additional analyses of passaged U1102 would be required to verify this.

Difficulties in the resolution of repetitive regions, a common problem encountered in the assembly of short-read NGS sequence data [Treangen & Salzberg, 2011], meant that complete genome sequences could not be obtained by NGS methods alone. Genome completion required directed PCR amplification combined with nucleotide sequence determination using Sanger capillary sequencing. Therefore the pipeline for determining complete genome sequences still requires a combination of NGS combined with Sanger sequencing for HHV-6A/-6B strains. In contrast, genomes of the related betaherpesvirus, HCMV, while possessing a complex herpesviral class E genome structure have fewer repetitive regions sharing the complexity observed in the HHV-6A/-6B genomes, as such NGS methodologies have been rapidly applied to HCMV genomics [Sijmons et al., 2014]. Nonetheless, almost all of the 85 identified ORFs of HHV-6A can be rapidly determined by these NGS methodologies. Therefore these methods could be utilised to faithfully determine and characterise HHV-6A/-6B genomes directly from tissue samples. Importantly, these methods can now be applied to determine the genome of the germline chromosomally integrated form of these viruses.

There are some differences between these two enrichment approaches. Using the same amount of input DNA, the mean read coverage observed from sequencing of the LPCR generated library was greater than that seen with the SureSelect library (2,581.51 vs 2,119.84). However, there was much greater variance in the read coverage across the genome (SD 1,439.74 vs 719.95 and figure 3.2 vs 3.3), due to the genomic amplicon

locations. This was most likely due to minor differences in relative molarity of the individual amplicons, since NGS is extremely sensitive to DNA quantity. Prior to equimolar pooling of the LPCR genomic amplicons estimates of amplicon size were made based on U1102 with confirmation/adjustments made by comparison with standards following agarose gel electrophoretic separation. While DNA quantification of purified amplicons was estimated by UV absorbance on a spectrophotometer. Calculations for the equimolar pooling of the genomic amplicons for DNA sequencing library generation were based on these measurements. It seems likely that the inherent error levels in these measurement procedures, most notably the UV-absorbance based DNA quantification where contaminant absorbance at 260 nm can lead to inaccuracies in DNA quantification [Manchester, 1996], had a significant impact on sequence uniformity. In the case of the sequencing of HHV-6A strain AJ, this variability had no major consequences since variation amongst HHV-6A strains is relatively low and the level of coverage was still sufficiently high enough for genome assembly. In order to improve relative quantification and resulting sequence coverage, fluorometer based assays could be used which rely on the specific binding of dyes to dsDNA in order to achieve high levels of fluorescence which in turn can be more accurately quantified [Rye et al., 1993]. A number of commercially available applications of this technology are now available, the use of which should help to improve target DNA quantification.

The percentage of the genome which could be resolved was almost identical for both enrichment methods. Similarly, consensus sequences generated from both methodologies were again identical, when inaccuracies in the assembly of repetitive regions were taken into account. Thus, from a technical standpoint there is little to distinguish these enrichment methods for the NGS sequencing of HHV-6A strains for this application. Instead, the choice of method will likely rely on other factors such as time, cost and sample limitations. Given that both methods rely on Sanger sequencing to complete the genomic sequences, the LPCR amplicon method could show advantages in that an archived amplicon pool is created which can be used to complete the genomic sequence. In contrast, using the SureSelect system all starting material is used in library preparation.

The second aim here was to compare two different sequence read assembly methods. In many of the comparative studies performed to date the findings suggest that the exact choice of assembler has to be considered on a genome by genome basis [Ekblom & Wolf, 2014]. This comparison was performed through the use of two publicly available and most frequently used assembly programs, one utilising a reference genome mapping approach and the other a *de novo* assembly approach. The reference genome based mapping assembly used the BWA-MEM alignment algorithm [Li & Durbin, 2009] with subsequent SAMtools mediated variant calling for consensus sequence generation [Li et al., 2009]. While the *de novo* assembly utilised a De Bruijn graph mediated *de novo* assembly of contigs followed by reference genome based scaffolding, using a Velvet optimiser, Velvet, ABACAS pipeline [Zerbino & Birney, 2008; Assefa et al., 2009]. These particular programs are easily installed with low computational time and resource usage, are widely utilised and have extensive literature and community support surrounding their use, accuracy and limitations [Shang et al., 2014; Zhang et al., 2011]. As such the results here should be widely applicable to other roseolovirus genomes, in particular the chromosomally integrated forms of HHV-6A/-6B which are analysed in this thesis.

Comparisons between the consensus sequences generated from both assembly methods and with Sanger sequencing characterised regions suggested either target enrichment

method in conjunction with either assembly methods could be utilised for accurate sequencing of the complete UL region of the HHV-6A genome. Indicating the low level of diversity observed amongst HHV-6A strains, to date, permits a reference based mapping assembly of the UL region of the genome. This may be of use where time or computational resources required for a *de novo* assembly are not available. However, where possible the complementary information provided by both assembly methods should be used to guide genome assembly. This is especially important in cases where the target genome may be unknown, such as the integrated forms of HHV-6A/-6B. The resolution of repetitive regions outside of the UL region of the genome was confounded using the reference mapped assembly approach. This was particularly apparent in the DR regions of the genome where due to the multiple repeat iterations, sequence reads could map to multiple sites on the reference genome. This lowered the mapping quality score of a read to levels which would exclude it from consensus sequence assemblies due to cut-off limits imposed to prevent misaligned reads. For these regions a *de novo* assembly approach was best. Indeed, the previously identified DR6 substitution [Schleimann et al., 2014] was only identified by the *de novo* assembly method. However, although the use of paired end sequencing can be used to aid the resolution of these regions [Cahill et al., 2010], a feature incorporated into the Velvet assembler [Zerbino et al., 2008], errors were noted in the number of repeats when compared to sequence resolved in some of the Sanger sequencing used for gap filling. In this regard, while some alternative assembly methods utilise different assembly algorithms and thus have the potential to provide better estimates of repeat structures [Breseler et al., 2012; Ruan et al., 2013]; it seems that until Illumina based sequencing technology allows for improvements in read length, repetitive sequences still pose a problem to genome assembly. Here for HHV-6A, these gaps could be addressed using directed PCR amplifications combined with Sanger capillary sequencing. But this stage still represents a bottleneck and would need to be improved to enable high throughput HHV-6A/-6B whole genome sequencing. However, the bulk of the genome could be determined using the NGS methods alone, including all coding sequence, which would be sufficient for many purposes.

This work on identifying an approach for the NGS sequencing of HHV-6A led to the resequencing of the prototypical HHV-6A strain U1102, as well as determination of the complete genome sequence of the third HHV-6A strain to be sequenced, that of AJ. The resequencing of strain U1102 identified five new potential sites of variation, two of which were represented in previously reported sequence fragments. All newly identified sites of variation represented minor changes to the amino acid sequence. But were in genes involved in many crucial processes for the replicative cycle: U40 in DNA packaging [Jones & Teo, 1993], U42 in viral mRNA export from the nucleus [Zielke et al., 2011], U57 the major structural capsid protein [Littler et al., 1990], U100 (gQ) in virion attachment [Mori et al., 2003], and U58, the function of which is currently unknown in HHV-6A/-6B. These are differences to the HHV-6A strain U1102 genome sequence which has been utilised as the reference genome for two decades. These could be minor variants from mixed infection in the original isolated virus or result from sequencing errors due to the low genome coverage from many of the original Sanger shotgun sequencing projects. More strain comparisons would be required to determine the extent of divergence at these sites.

HHV-6A strain AJ was originally isolated from an adult HIV/AIDS patient from the Gambia [Tedder et al., 1987]. Along with strains U1102 and GS, strain AJ is one of the earliest isolates and subsequently most readily utilised laboratory strains of HHV-6A in the three decades since the discovery of the species. Thus the sequencing of this third HHV-6A strain

has provided more insight into genomic variation within the species, which should help to understand its role in pathogenesis. Strain AJ showed closest relation to the recently sequenced isolate of strain GS (Accession KC465951) [Gravel et al., 2013a]. This could be reflecting a highly evolved virus status or potentially a recent emergence. Both virus genomes were determined from tissue cultured virus, yet the high nucleotide identity observed between these two genomes, their different geographic isolation and culture, the retention in AJ of the full gene complement identified in strain GS and similarities in the overall architecture of the genome sequences between strains AJ and GS, suggest that *in vitro* passage of HHV-6A strains may not significantly alter the genetic content of the viruses in the same extent as observed amongst some other betaherpesviruses such as HCMV [Dargan et al., 2010]. Although it has been noted in HHV-6B that the genomes of laboratory adapted strains can be up to several kb shorter in length than wild strains, these are not due to the gene deletion and rearrangements shown in HCMV, but primarily due to the loss of repetitive sequences from the DR regions upon viral passage [Lindquister & Pellet, 1991]. Indeed these passage effects could contribute to the genome length disparity observed between strains AJ and GS, since the majority of the differences were in the DR region. Comparisons of the two GS strains reported, one early passage and one late, also showed limited differences [Gravel et al., 2013a].

The report on strain GS from a North American mixture of donors commented on the distinct differences to strain U1102, of African origin. The study suggested that these result from the different geographic origins [Gravel et al., 2013a]. However, strain AJ was also from African origin, yet it shares greatest similarity with the North American isolate. Additionally, sequence analyses noted variation in the DNA polymerase gene, U38. Due to its high level of conservation amongst the herpesviruses, motifs within this gene have been extensively utilised for diagnosis and in determination of relationships between herpesviral species [Minjolle et al., 1999; Tafreshi et al., 2005; McGeoch et al., 2006]. In the case of the HHV-6 species, amplification of regions of this gene is frequently used as a means of species typing. The variation noted here in strain AJ, but also present in strain GS potentially confounds one commonly utilised diagnostic assay, and where it has been applied, low prevalence of HHV-6A has been reported, potentially due to the lack of specificity in the primer sequences [Lou et al., 2011; Oakes et al., 2013; Tembo et al., 2015].

In summary, comparisons were made on methods for the enrichment and accurate NGS of HHV-6A genomes. These methods permitted enrichment and characterisation of herpesvirus genomes separate from the human host; methodology which can be used to allow accurate whole genome sequencing directly from tissue samples. Validation of methods resequenced the prototypical HHV-6A strain, U1102, identifying 5 new sites of variation. Application of the methods was then used to derive the complete genome sequence of the third HHV-6A strain to be sequenced, AJ. A robust NGS pipeline was established which could be applied to characterise the germline chromosomally integrated forms of HHV-6A/-6B.

Chapter 4: Genomic Analyses of CI-HHV-6A and Reactivation

4.1 Introduction

At the start of this thesis the genetic relationships between the germ-line integrated viral forms and circulating strains of HHV-6A/-6B was unknown. In order to understand their relationship to infectious viral strains, attempts were made to characterise the chromosomally integrated form of these viruses, using both Sanger sequencing and the next-generation sequencing pipeline established in the previous chapter.

Initially, PCR screens followed by Sanger sequencing were performed to characterise the junction of the viral integration site into the host chromosome from a panel of samples from patients with confirmed CI-HHV-6A and -6B. Subsequently, NGS was used to sequence three CI-HHV-6A genomes from clinical samples. This enabled comparisons of coding capacity of the CI-HHV-6A strains to that of exogenous viral strains, as well as the assessment of *cis* acting sequences likely important for viral reactivation/replication. Finally, deep sequencing with minor variant analysis was utilised to further investigate the reactivation potential of CI-HHV-6A.

4.2 Characterisation of a Common Integration Site at Chromosome 17p

A panel of 70 CI-HHV-6A/-6B patient DNA samples were collected for analyses. These were identified following screening of patients reporting to a diagnostic facility in Berlin, Germany, with cardiac complications and a hospital in Prague, Czech Republic with lymphoproliferative or inflammatory disease (see methods section 2.1.2). 44 samples were from the German diagnostic facility and the remaining 26 from the Czech Republic. These confirmed CI-HHV-6A/-6B samples were then assayed further in order to characterise the site of viral integration into the human genome. To do this a set of specific oligonucleotide primers were used to PCR amplify the sequence of the viral genome-human subtelomere integration site junctions.

CI-HHV-6A and -6B has been previously shown to insert into the sub-telomeric region of human chromosomes, in an orientation where U100 is located towards the centromere closest to subtelomeric sequences and U1 closest to the telomeric cap [Arbuckle et al., 2010, 2013; Huang et al., 2014]. As such a primer was utilised to specifically amplify from a HHV-6A and -6B conserved site in the DR_R of the viral genome (designated DR_R) [Arbuckle et al., 2010] in conjunction with primers previously utilised to specifically amplify from subtelomeric sequences from human chromosomes 10, 11, 17 and 18 (designated 10qF, 11q, 17p and 18q, respectively) [Britt-Compton et al., 2006; Arbuckle et al., 2010; Huang et al., 2014], which have been observed as CI-HHV-6A and/or -6B integration sites [Nacheva et al., 2008; Arbuckle et al., 2010, 2013; Huang et al., 2014]. Based on these previous analyses the primer pairs were predicted to amplify a products approximately 1-2 kbp in length. However, variation in the number of telomeric repeats between the viral integration site and the human subtelomere had the potential to alter this.

Trials with all primer sets only identified integration sites at chromosome 17. With no specific products amplified from chromosomes 10, 11 and 18, as determined by Sanger sequencing of minor amplification products. In 25/70 (36%) of the samples an amplicon of approximately 1.5 kbp, as previously observed, was generated utilising the 17p primer. 12/44 (27%) of these from the German myocarditis/cardiomyopathy patient set and 13/26

(50%) from the Czech inflammatory disease set. To confirm these represented CI-HHV-6A or -6B integration site junctions and obtain the nucleotide sequence of the junctions, a nested primer set was designed for use in sequencing reactions. For this, a ClustalW2 [Larkin et al., 2007] alignment of previously published CI-HHV-6A and -6B 17p integration sites (GenBank accession numbers: GU784872.1, KF366419.1 and KF366420.1, respectively) was used to design primers located at a conserved site within the subtelomeric region of the amplicon (designated 17Tel-DR_IntF) and a semi conserved region, 1 bp difference across 22 bp between CI-HHV-6A and -6B, of the DR_R (designated 17Tel-DR_IntR). From the 25 amplicons, the nucleotide sequence of the complete integration junction could be resolved for 12. Sequence comparisons confirmed the relationship of all of these with the HHV-6A DR regions, indicating all were CI-HHV-6A (complete sequences in appendix 7.3). While partial junctional sequences were resolved from a further 9 amplicons, 6 CI-HHV-6A and 2 CI-HHV-6B. This confirmed the ability of both CI-HHV-6A and -6B to integrate at this chromosomal subtelomeric site [Nacheva et al., 2008; Ward et al., 2006; Clark et al., 2006b; Arbuckle et al., 2010; Luppi et al., 1993; Morris et al., 1999; Torelli et al., 1995]. This is the largest group of integration sites characterised to date, showing a marked preference within these patient groups for integration on chromosome 17.

Analysis of the viral DR_R region of the junctional sequences in isolation showed that despite HHV-6A and -6B strains sharing 94-97% nucleotide homology with each other in this region (HHV-6B HST 97% with Z29, HHV-6A U1102 95% and 94% with GS and AJ respectively), the CI-HHV-6A sequences amplified here shared 100% homology with one another in this region which was distinct from that seen in available circulating strains (94% to U1102 and 92% to GS and AJ), figure 4.1A. The CI-HHV-6A sequences also shared 99% nucleotide identity with two other previously characterised CI-HHV-6A chromosomal integration site junctions, with evidence for integration at chromosome 17p13.3, figure 4.1A. The first of these an isolate (accession KF366419.1) of CI-HHV-6A from the Negev region of Israel [Cann et al., 2002; Huang et al., 2014] and the second an isolate (accession GU784872) of undefined, possibly US, origin [Arbuckle et al., 2010]. Yet they were distinct from the DR_R region of CI-HHV-6A samples from *in vitro* integration into the JJhan T cell line at chromosomes 11 and 18 (accessions GU784871 and GU784873, respectively) [Arbuckle et al., 2010]. Both CI-HHV-6B sequences amplified here also shared 100% nucleotide homology in this region, which was identical to another previously characterised CI-HHV-6B isolate, HGDP0628 from Sardinia [Cann et al., 2002; Huang et al., 2014] as well as HHV-6B strain HST, but distinct from strain Z29, figure 4.1B. Similarly, the sequences of the readily identifiable 17p subtelomeric region adjacent to the integration junctions of all the CI-HHV-6A and CI-HHV-6B samples shared 100% nucleotide identity with one another. As well as 100% identity with both the previously identified 17p CI-HHV-6A and -6B chromosomal integration sites (data not shown).

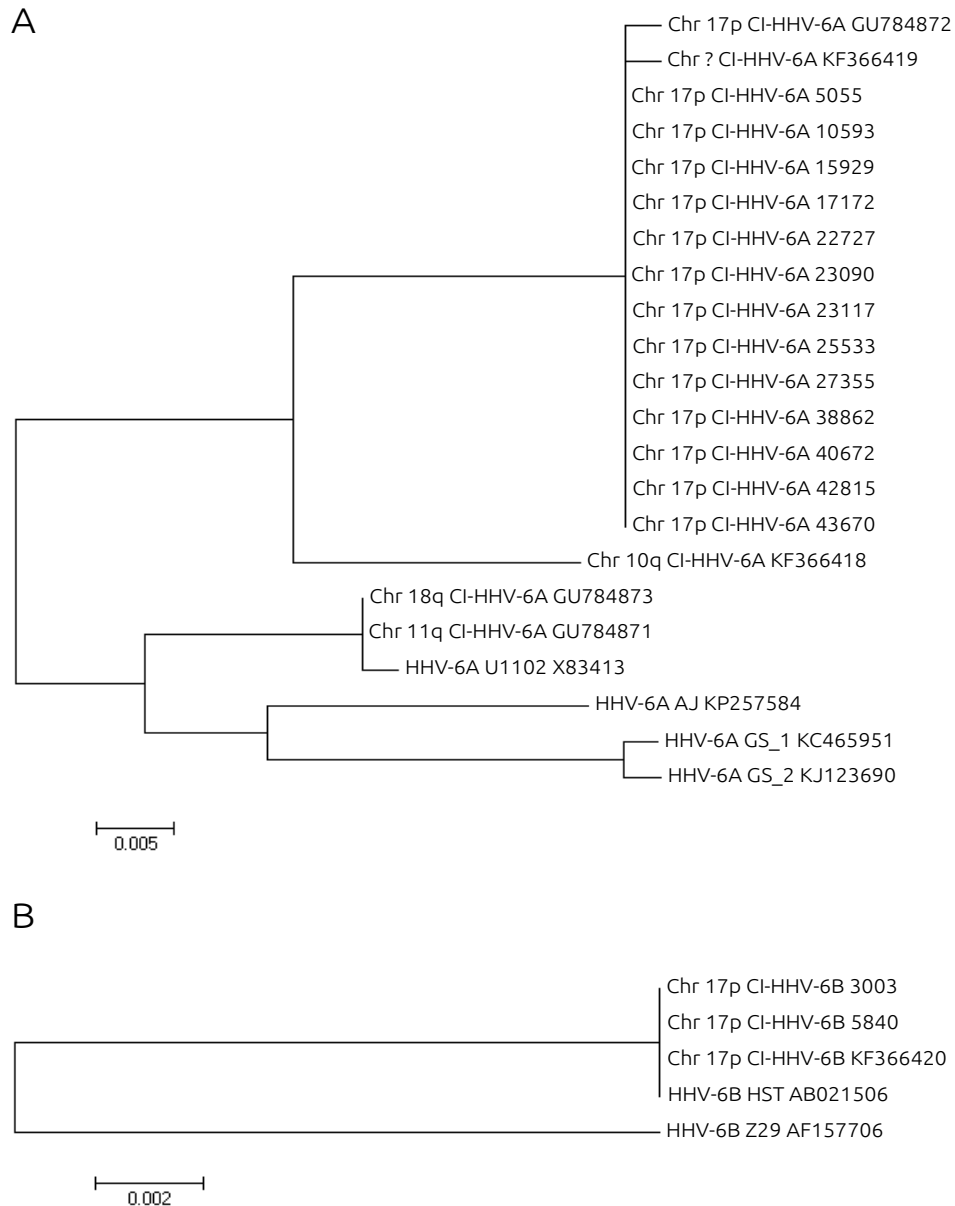


Figure 4.1 Phylogenetic analyses of viral DR_R sequence from viral chromosomal integration site junction amplifications. (A) CI-HHV-6A isolates compared to HHV-6A strains (B) CI-HHV-6B isolates compared to HHV-6B strains. Comparisons made using maximum likelihood estimation. Accession numbers of all HHV-6A/-6B strains and previous CI-HHV-6A/-6B integration site junctions, included in the figure. Branch length scale represents number of nucleotide substitutions per site. Complete junctional sequences of the CI-HHV-6A isolates amplified here, included in appendix 7.3.

Analysis of the complete junctions confirmed the loss of nucleotide sequence from the far right end of the genome, starting in the T2 perfect telomere-like repeat region of the DR_R, during the viral integration event [Arbuckle et al., 2010, 2013; Huang et al., 2014; Ohye et al., 2014]. Importantly, including loss of the highly conserved pac 2 DNA packaging motif. Analyses of the repeat structure showed these consisted of a short subterminal sequence, followed by a varying copy number of perfect telomeric repeats, telomeric repeat variants commonly found in human telomeres [Baird et al., 1995], before returning to perfect telomeric repeats of human or viral origin, then readily identifiable viral DR_R sequence. Therefore the integration site junctions did not consist of solely perfect telomeric repeat

subunits, but had a complex structure suggesting retention of host telomeric repeat variants.

In addition to this generalised architecture, large similarities could be observed in the repeat structures of the junctions from the samples examined. All the CI-HHV-6A junctions shared a characteristic pattern of perfect and imperfect repeats immediately after the readily identifiable 17p subtelomeric sequence. In addition to a pattern of imperfect repeats followed by 6 perfect telomeric repeats immediately prior to the viral DR_R sequence, figure 4.2. The only divergence was in the number of perfect repeats between these two signatures, which varied from 5 up to 67 here (mean 39, median 52). This indicated a specificity to the integration site, not previously observed.

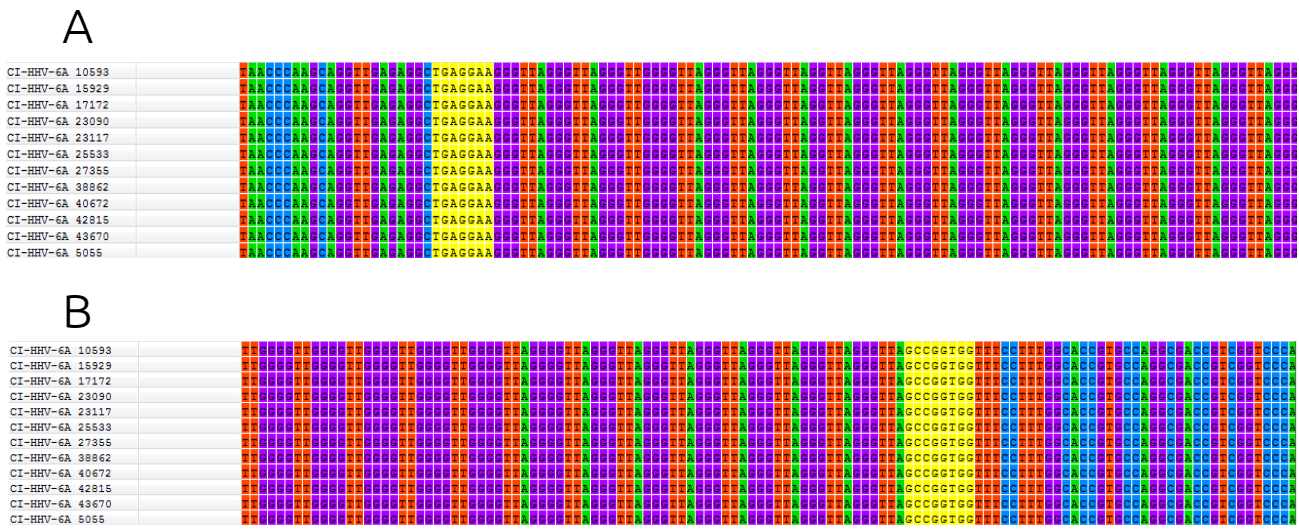


Figure 4.2 Alignments of CI-HHV-6A integration site junctions at chromosome 17p13.3 show common perfect and imperfect telomeric repeat structures. (A) Yellow highlighted sequence indicates the end of the readily identifiable chromosome 17p subtelomeric sequence. (B) Yellow highlighted sequence indicates start of readily identifiable viral DR_R sequence. Complete junctional sequences of the CI-HHV-6A isolates amplified here, included in appendix 7.3

This structure is distinct from the integration sites defined at other chromosomes, from both germ-line CI-HHV-6A (Chromosome 10; Accession KF366418) and *in vitro* cell line integrations, (Chromosome 11; accession GU78487 and chromosome 18; accession GU784873), summarised in figure 4.3. While still distinct, more substantial similarities could be observed between the repeat structure of chromosome 17p subtelomeric record from human genome sequencing (accession AC240565) as well as a previously defined integration junction (accession KF366419) from a distinct geographic origin, potentially also integrated at chromosome 17p. A further CI-HHV-6A isolate also integrated at chromosome 17p (accession GU784872), shared similar repeat structures adjacent to the viral DR_R, but the characteristic structure adjacent to the human subtelomeric region of the junction was absent. The common architecture of the human-viral integration site junction observed for isolates integrated at chromosome 17p indicated a shared genetic lineage.

HHV-6A DR_R

HHV-6A Strain U1102 DR_R telomeric repeats (X83413):

(GGGTTA)₅₉-DR_R

Human Chromosome (no integration)

Chromosome 17p (AC240565):

(GGGTTA)₂(GGGTT)₁(GGGGTT)₃(GGGGTTA)₁(GGGTTAimperfect)₆₅(GGGTTT)₁(GGGTTA)₉-Unique

Inherited CI-HHV-6A

Chromosome 17p CI-HHV-6A DR_R integration site (14 CI-HHV-6A isolates: 5055-43670 and KF366419):

(GGGTTA)₂(GGGTT)₁(GGGGTTA)₁(GGGTTA)₅₋₆₇(GGGTT)₁(GGGGTT)₄₋₈(GGGGTTA)₂(GGGTTA)₆-DR_R

Chromosome 17p CI-HHV-6A DR_R integration site (GU784872):

(GGGTTA)₆-DR_R

Chromosome 10q CI-HHV-6A DR_R integration site (KF366418):

(GGGTT)₁(GGGGTTA)₁(GGGTTTA)₂(GGGTTA)₂₇(GGCTTA)₂(GGGTTA)₁₈(GGGTT)₁(GGGGTT)₆(GGGGTA)₄(GGGTTA)₁₁(GGGGTTA)₄(GGGTTA)₁₇-DR_R

In vitro CI-HHV-6A

Chromosomes 11q in vitro integration (GU784871):

(GGGTTA)₄₁-DR_R

Chromosomes 18q in vitro integration (GU784873):

(GGGTTA)₅₄-DR_R

Figure 4.3 Comparison of sub-telomere adjacent repeat structure between CI-HHV-6A integrated at chromosome 17p and other defined chromosomal integration sites. Imperfect repeats derived from the human subtelomeric region are shaded. Number of repeat structures indicated in subscript. Accession numbers indicated in brackets. Full alignment of the junctional sequences of the CI-HHV-6A isolates amplified here, is included in appendix 7.3.

4.3 Genomic Analysis of CI-HHV-6A

4.3.1 CI-HHV-6A Genome Determination

Previous work on the phylogenetic relationships between CI-HHV-6A and -6B isolates through multiple loci genotyping, described in part in section 5.2 and fully in Tweedy et al., 2015a, suggested the greatest diversity of CI-HHV-6A and -6B was shown in CI-HHV-6A sequences. Therefore, these CI-HHV-6A genomes were investigated further. For this, the Illumina MiSeq NGS methodology determined in chapter 3 was applied for the whole genome sequencing of three CI-HHV-6A samples, isolates 2284/4305, 5055/1624 and 5814 from the German myocarditis/cardiomyopathy cohort. One of which, 5055, had evidence for integration into chromosome 17.

As detailed in the previous chapter, a long-range PCR method was used to amplify viral genomic DNA, described in section 3.2. The same 36 overlapping primer sets designed [against HHV-6A strain U1102, Accession X83413.1, RefSeq NC_001664.2] were used to produce amplicons, which were subsequently purified and pooled in equimolar ratios. These genomic DNA preparations were then sheared and DNA libraries generated for paired end sequencing. Sample multiplexing, cluster generation and sequencing was performed on an Illumina MiSeq sequencer. Read assembly, consensus sequence generation and genome annotation were all performed as previously described in section

3.2. For the CI-HHV-6A samples, a mapped assembly was performed, with comparisons made to *de novo* assembled contigs for validation purposes. As highlighted in section 3.2, resolution of repetitive regions of the genome again were confounded since read lengths were shorter than these repeat regions. Therefore, these repeat containing regions, R2, R3 and the DR's were initially excluded from the consensus sequences generated by the reference mapped assemblies. The resolution of these regions was instead performed by targeted *de novo* assembly. Again BLAST searches [Camacho et al., 2009] were performed on non-aligning read pairs with no indication of major insertions missed by the mapped assemblies, annotation of the genomes was performed as described in section 3.2.

CI-HHV-6A Genome ID	Total Reads	% Mapped Reads	Mean Read Coverage across UL	% UL Coverage >1 Read	% UL Coverage >25 Reads	% <i>De Novo</i> Coverage Across DR
2284/4305	2626303	96.40	1907.51 (SD 4,479)	99.92	91.67	40.13
5055/1624	20446080	98.31	19882.10 (SD 11,491)	98.37	66.72	83.10
5814	5165853	96.06	4023.99 (SD 4,545)	97.91	89.39	65.04

Table 4.1 Sequence read and coverage statistics for the CI-HHV-6A genomes subjected to NGS.

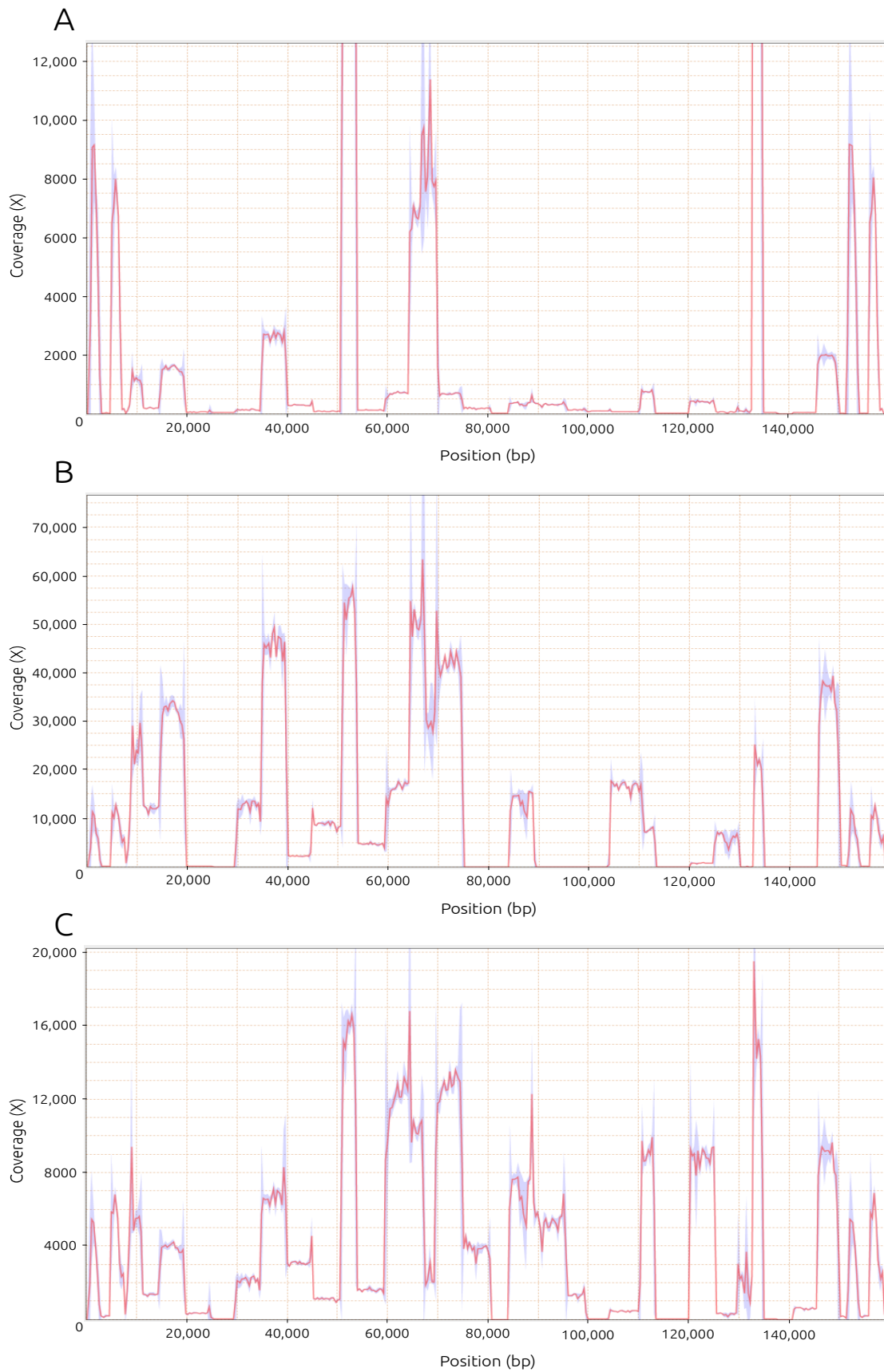


Figure 4.4 Sequence read coverage of CI-HHV-6A genomes across the HHV-6A strain U1102 reference genome following mapped assemblies. (A) CI-HHV-6A isolate 2284 (B) CI-HHV-6A isolate 5055 (C) CI-HHV-6A isolate 5814.

Sequence reads mapped to all currently defined HHV-6A CDS features, validating the use of this methodology for the enrichment of CI-HHV-6A genomic DNA. While also indicating that all currently defined CDS features are retained in CI-HHV-6A genomes. However, it was clear that variance existed in the coverage observed between the amplicons, as highlighted in section 4.3.1 and table 4.1. Some of the genomic amplicons were under-represented in the DNA pooling prior to sequencing library generation, resulting in low read coverage across CDS features within these genomic regions. Measures were taken to determine the practical level of read coverage required for accurate sequence resolution, described in appendix 7.4. This indicated a read depth of greater than 7 allowed for accurate sequence resolution across the CDS features. Therefore to minimise the potential for sequencing errors to confound downstream analyses, any genes where the read coverage from one or more of the CI-HHV-6A genomes was below this cut-off, indicated in table 4.2, were excluded from further analyses.

For sample 2284 coverage across the CDS regions was a minimum of 7 fold. Only two genes were at this minimal cut-off, U50, a capsid associated tegument protein and U76, the capsid portal protein. In total there were 11 nucleotide positions at a depth of 7 reads across these two conserved structural proteins. Thus, while at the limit of accurate sequence resolution this indicated that overall there were no major deletions or defects in the viral exome of CI-HHV-6A respective to HHV-6A, and all 85 genes could be identified. With analysis of non-aligning read pairs also indicating no major insertions. This represents the first report of whole genome sequencing from the germ-line integrated form of HHV-6A.

Taken together these results confirmed the methodology utilised for viral enrichment of HHV-6A genomic DNA from culture samples is also applicable for the enrichment of CI-HHV-6A genomic DNA from clinical samples. Furthermore, the viral genome as currently described for HHV-6A is also retained in CI-HHV-6A.

CDS	2284 U1102		5055 U1102		5814 U1102	
	Mean	Minimum	Mean	Minimum	Mean	Minimum
U2	1522.7	421	38397.16	19944	6648.13	4203
U3	1290.19	219	40268.08	16512	4917.9	1243
U4	234.55	180	18630.46	14476	1351.02	1046
U7	1401.69	187	40521.37	16686	3103.38	1183
U10	1938.56	1257	48657.89	32441	3722.55	2355
U11	437.79	18	10101.01	69	1093.68	166
U12	54.36	33	127.88	99	342.1	291
U13	57.33	47	130.8	112	328.95	309
U14	80.38	15	129.16	19	375.83	17
U15	41.15	22	42.16	13	41.92	23
U17	40.51	23	15.65	10	33.36	14
U18	43.66	31	16.53	12	34.1	24
U19	126.6	24	16781.17	23	1866.8	22
U20	135.02	15	20853.52	11655	2190.27	1135
U21	142.45	48	19912.85	8514	2160.75	951
U22	125.77	99	20568.85	18519	2182.35	1811
U23	1082.59	46	36582.05	6121	3768.7	722
U24	3444.47	3137	71347.85	68842	6322.49	5905
U24A	3725.95	3495	71901.76	66626	6776.1	6455
U25	3398.25	2878	70545.89	61687	6398.75	5490
U26	3609.93	3274	74937.3	65796	6763.7	6061
U27	3534.74	230	72295.38	54787	6682.31	394
U28	1329.09	286	22196.27	1696	4631.76	2759
U29	351.17	280	3512.05	3343	3081.17	2839
U30	352.58	89	5644.25	2735	3207.83	1158
U31	2229.94	47	19705.43	8636	2311.38	679
U32	25238.25	17559	76063.43	53823	14209.49	10163
U33	29798.2	26482	88526.41	80231	16266.52	14937
U34	22519.67	94	68676.93	6362	12926.78	1465
U35	125.54	104	7500.97	7098	1582.35	1482
U36	121.29	99	7343.97	6762	1609.29	1410
U37	126.28	94	7544.6	6615	1663.36	1411
U38	119.46	44	7710.53	2614	1566.01	628
U39	730.16	98	25650.22	7186	11675.13	1502
U40	774.02	686	26818.13	24015	12540.79	11502
U41	9994.14	659	80248.54	23788	10114.11	1451
Ori-lyt	9970.73	9180	41881.65	34014	1777.28	1649
U42	5806.72	568	65420.33	41001	8986.5	1568
U43	719.76	605	66903.68	61063	12820.57	11768
U44	766.15	689	70249.33	66744	13544.07	12298
U45	540.6	144	41477.45	35280	9497.74	3266
U46*	202.74	182	7.04	4	4124.52	3774
U47*	184.77	51	7.61	2	3707.86	1396

CDS	2284 U1102		5055 U1102		5814 U1102	
	Mean	Minimum	Mean	Minimum	Mean	Minimum
U47A	167.76	146	4.35	2	3817.66	3382
U48	187.9	153	8.02	1	3921.58	3427
U49	93.4	12	8.42	1	1194.16	0
U50	20.18	7	4.29	0	0.62	0
U51*	18.06	10	3.57	0	1.04	0
U52	18.3	9	4.05	1	0.38	0
U53	391.56	17	22324.03	1967	7913.16	6920
U53.5	378.58	326	22974.94	21408	7650.02	7040
U54	288.86	29	20436.73	2189	6094.24	817
U55	446.21	254	22986.71	3969	8426.56	4950
U56	392.12	302	2460.21	13	5745.84	4935
U57	322.11	96	22.51	13	5084.06	1754
U58	300.72	107	24.32	15	4402.83	1190
U59	133.89	100	23.24	16	1326.17	1130
U60	120.59	40	7.02	0	649.48	14
U62	140.57	122	13.06	9	1419.8	1345
U63	135.79	95	8.68	3	1499.62	1269
U64	121.49	65	8.89	0	231.46	24
U65*	100.41	72	5.64	0	40.03	26
U67	99.17	71	5.7	0	32.75	13
U68	105.46	81	1.98	0	35.73	21
U69	92.33	46	21999.51	21	411.69	20
U70	69.07	49	26742.35	23336	450.61	341
U71	65.42	51	27424.14	26209	454.56	386
U72	68.85	48	25926.13	24087	441.75	379
U73	161.68	35	26524.34	9937	1521.64	275
U74	860.85	640	12008.06	9761	8958.07	7808
U75	618.04	13	9935.16	0	6329.93	16
U76	19.55	7	2.35	0	23.66	10
U77	16.59	9	2.38	0	21.86	13
U79	437.68	56	1169.68	18	9097.3	180
U81	420.23	324	1149.03	949	8545.88	7475
U82	412.27	346	1225.59	1052	9232.35	7992
U83	407.04	336	1158.37	1031	8417.23	7788
U84	392.98	264	1184.52	904	8902.64	6920
U85	182.79	31	9530.33	958	3389.09	299
U86	76.87	10	9179.27	27	565.95	14
U90	17653.28	30	18810.7	14	7625.66	16
U91	40.42	23	21.96	13	18.25	13
U94	51.29	34	16.74	11	629.89	513
U95	538.39	21	8873.08	16	2169.54	13
U100	2096.9	14	50326.9	304	7600.19	90

Table 4.2 Read coverage across CI-HHV-6A CDS from the UL genomic region. Red colour denotes coverage across the CDS was at or fell below the defined cut-off for accurate sequence resolution, see appendix 7.4, for one or more CI-HHV-6A isolates. Stars denote sequenced obtained by Sanger capillary sequencing and utilised in further analysis. CDS of the DR regions are omitted here as they were determined by *de novo* assemblies.

4.4 CI-HHV-6A Variation

The utilisation of concatenations of highly conserved core genes has previously been utilised extensively to investigate the relationships between members of the *Herpesviridae* [McGeoch et al., 2006]. Therefore, initially the relationship between the CI-HHV-6A isolates and other human herpesviruses was examined using concatenations of a set of these homologues found in representative members of human herpesviral species from the three *Herpesviridae* sub-families, figure 4.5. These showed that the CI-HHV-6A strains clustered together, tightly linked but separate to the currently sequenced HHV-6A strains.

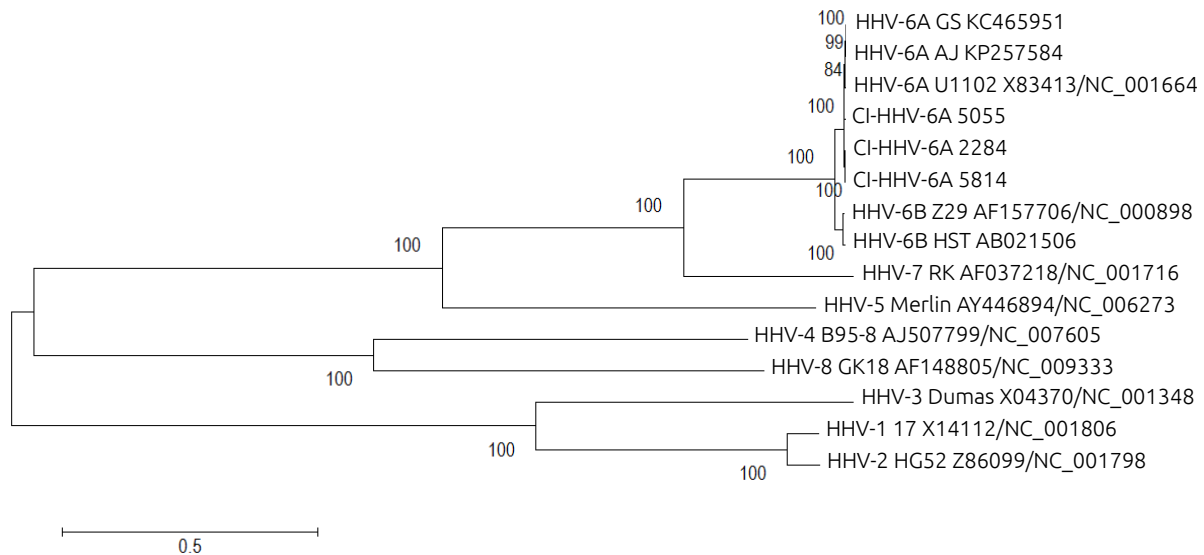


Figure 4.5 Phylogenetic analyses of CI-HHV-6A compared to prototypical human herpesvirus species from all three *Herpesviridae* sub-families. Maximum likelihood estimations, with bootstrapping (1050 replicates), from concatenations of *Herpesviridae* conserved genes: capsid triplex subunit 1 (HHV-6A/-6B U29), small capsid protein (U32), large tegument protein (U3), large tegument binding protein (U30), cytoplasmic egress tegument protein (U71), cytoplasmic egress facilitator-1b (U44), glycoproteins gB, gL, gM (U39, U82, U72), multifunctional expression regulator (U42), DNA polymerase catalytic subunit (U38), DNA polymerase processivity subunit (U27), helicase-primase RNA polymerase subunit (U43), helicase primase subunit (U74), single-stranded DNA-binding protein (U41), alkaline deoxyribonuclease (U70), uracil DNA glycolase (U81), ribonucleotide reductase large subunit (U28), capsid transport nuclear protein (U36), DNA packaging terminase subunit 2 (U40), terminase binding protein (U35), nuclear egress membrane protein (U34), nuclear egress lamina protein (U37). Accession numbers of all strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site.

Subsequently, relationships of the CI-HHV-6A isolates to HHV-6A strains were investigated further. To compare the relationships, classes of conserved and divergent genes were compared amongst CI-HHV-6A, HHV-6A and HHV-6B with the aim of subjecting these to phylogenetic analyses. This allowed the determination of the relationships, while taking into account the possibility of different evolutionary rates in highly conserved and non-conserved genes [Pond et al., 2010; Subramanian & Lambert, 2012]. To achieve this, the complete genome sequences of HHV-6A strains U1102 (modified version of accession X83413/NC_001664 with the corrections outlined in section 3.3), GS (accession KJ123690) and HHV-6B strains Z29 (accession AF157706) and HST (accession AB021506) were all re-annotated to reflect the experimentally defined corrections noted in the recent HHV-6A strain GS (accession KC465951) and AJ (accession KP257584) annotations. Thereby allowing for direct comparisons between genes of all the genomes using the same CDS definitions and splice donor/acceptor sites. Subsequently, pairwise comparisons were performed between all the defined CDS with all noted in relation to the prototypical HHV-6A strain U1102, table 4.3.

CDS	Pairwise comparisons to HHV-6A U1102								
	HHV-6A			CI-HHV-6A			HHV-6B		
	GS 1	GS 2	AJ	2284	5055	5814	Z29	HST	
%	%	%	%	%	%	%	%	%	%
DR1	96.98	96.59	95.75	95.41	96.18	95.75	89.90	89.90	
DR6	96.41	96.15	95.81	95.09	96.50	96.15	88.87	89.97	
U2	98.64	98.55	98.64	98.82	98.82	99.00	92.31	92.49	
U3	99.11	97.86	98.93	98.48	98.48	98.48	92.25	92.07	
U4	99.75	99.75	99.75	99.69	99.19	99.69	96.89	96.89	
U7	99.35	99.35	99.32	99.51	98.91	99.54	95.03	95.03	
U10	99.01	98.93	99.01	99.31	99.31	99.31	94.28	94.28	
U11	97.78	97.78	97.86	97.97	98.08	97.82	89.25	89.25	
U12	99.04	99.04	98.95	98.95	99.04	98.95	93.23	93.30	
U13	99.69	99.69	99.69	99.69	97.48	97.48	96.26	96.57	
U14	99.67	99.67	99.56	99.45	96.45	95.96	90.55	90.71	
U15	97.57	97.57	97.57	97.57	97.57	99.65	94.44	94.79	
U17	99.00	99.00	98.71	98.71	98.71	99.30	94.63	94.63	
U18	98.07	98.07	98.07	98.07	98.07	98.30	93.65	93.65	
U19	97.69	97.69	97.86	97.86	97.69	98.03	94.27	94.27	
U20	98.82	98.82	98.90	98.97	98.84	98.89	94.25	94.25	
U21	98.31	98.31	99.46	98.39	98.39	98.39	93.24	93.24	
U22	99.84	99.84	99.84	99.18	99.18	99.18	95.40	96.39	
U23	98.87	98.87	99.02	99.30	99.30	99.30	92.83	92.83	
U24	97.73	97.73	98.48	98.11	98.08	98.11	93.18	92.42	
U24A	100.00	100.00	100.00	100.00	98.85	98.85	93.68	93.68	
U25	99.68	99.68	99.47	99.16	99.89	99.26	97.37	97.37	
U26	99.55	99.55	99.55	98.87	99.32	98.87	95.15	95.05	
U27	99.49	99.49	99.07	98.47	99.75	98.47	95.55	95.83	
U28	99.50	99.50	99.30	99.09	99.59	99.09	95.82	95.78	
U29	99.89	99.89	100.00	99.22	99.89	99.22	96.00	96.00	
U30	99.48	99.48	99.38	98.68	99.72	98.68	94.86	94.49	
U31	99.47	99.47	99.49	99.37	99.37	99.23	94.83	94.69	
U32	100.00	100.00	100.00	99.25	99.25	99.25	95.88	95.88	
U33	99.58	99.58	99.93	99.01	99.29	99.01	97.10	97.24	
U34	99.04	98.92	99.76	99.40	99.16	99.40	95.19	94.71	
U35	100.00	100.00	100.00	99.38	99.07	99.38	97.82	97.17	
U36	99.59	99.59	99.66	99.11	99.18	99.04	96.63	96.41	
U37	99.37	99.37	99.50	99.25	99.37	99.25	96.98	97.11	
U38	99.05	99.05	99.05	99.11	99.08	99.11	96.18	96.12	
U39	99.20	99.20	99.36	99.48	99.44	99.48	95.67	95.75	
U40	99.04	99.04	99.17	99.08	98.95	99.08	96.88	96.84	
U41	99.29	99.29	99.32	99.29	99.24	99.29	97.35	97.32	
U42	98.77	98.77	98.71	99.29	99.29	99.29	95.47	95.47	
U43	99.46	99.46	99.65	99.61	99.65	99.61	97.33	97.33	
U44	99.51	99.51	99.35	99.35	99.35	99.35	95.95	95.95	
U45	99.56	99.56	99.56	99.29	99.29	99.29	95.76	95.76	
U46	98.43	98.43	98.04	100.00	99.61	100.00	95.29	95.29	

CDS	HHV-6A			CI-HHV-6A			HHV-6B		
	GS 1	GS 2	AJ	2284	5055	5814	Z29	HST	
	%	%	%	%	%	%	%	%	
U47	98.47	98.47	98.16	97.70	97.70	97.70	88.50	88.60	
U47A	100.00	100.00	100.00	100.00	100	100.00	90.91	90.91	
U48	99.95	99.95	99.90	99.81	100	99.81	95.30	95.40	
U49	99.74	99.74	99.60	99.74	99.74	100	96.97	96.84	
U50	99.52	99.52	99.40	99.4	99.58	99.89	96.28	96.28	
U51	99.56	99.56	99.45	99.45	99.45	99.45	99.78	94.92	
U52	99.87	99.87	99.74	99.74	99.74	99.74	96.65	96.65	
U53	99.50	99.50	99.62	98.36	98.36	98.36	96.22	96.16	
U53.5	99.46	99.46	99.59	97.96	97.96	97.96	95.92	95.92	
U54	98.77	98.77	99.27	97.24	97.24	97.24	88.60	88.82	
U55	99.69	99.69	99.38	98.31	98.31	98.31	95.07	95.07	
U56	100.00	100.00	100.00	99.55	100.00	99.55	98.09	97.98	
U57	99.11	99.13	99.01	99.11	99.11	99.11	96.61	96.61	
U58	99.57	99.57	99.22	99.09	99.27	99.09	95.77	95.69	
U59	99.15	99.15	99.15	99.15	99.15	98.67	95.63	95.54	
U60	99.80	99.80	99.75	99.55	99.8	99.50	97.75	97.70	
U62	98.45	98.45	98.45	98.84	98.45	98.84	98.06	97.67	
U63	99.08	99.08	98.77	98.77	98.77	99.23	97.54	96.93	
U64	99.40	99.40	99.17	98.42	99.4	98.72	95.56	95.56	
U65	99.60	99.60	99.50	97.42	97.52	97.52	92.86	92.86	
U67	99.62	99.62	99.81	99.53	99.81	99.53	96.70	96.89	
U68	99.13	99.13	99.71	99.42	99.71	99.13	95.94	95.65	
U69	99.76	99.76	99.23	99.47	98.93	98.82	94.85	95.03	
U70	99.39	99.39	99.52	99.52	99.39	99.11	95.91	95.98	
U71	99.15	99.15	99.15	99.15	95.30	95.30	92.74	92.31	
U72	99.52	99.52	99.61	99.61	99.61	99.42	96.62	96.71	
U73	99.57	99.53	99.57	99.53	99.62	99.45	97.27	97.06	
U74	99.45	99.45	99.40	99.04	98.99	99.04	96.13	96.28	
U75	99.87	99.87	99.73	99.33	99.33	99.07	96.13	96.00	
U76	99.69	99.69	99.79	99.79	99.95	99.38	97.17	97.58	
U77	99.72	99.72	99.72	99.72	99.84	99.68	97.29	97.29	
U79	96.55	96.62	98.46	96.63	96.85	96.12	90.74	90.67	
U81	99.48	99.48	99.22	98.96	98.96	98.96	95.05	95.05	
U82	99.73	99.73	99.34	98.94	99.47	98.94	95.48	95.48	
U83	98.96	98.96	98.30	98.64	98.64	98.64	92.52	92.52	
U84	99.32	99.32	98.15	99.22	99.42	99.22	94.27	94.56	
U85	98.97	98.97	98.40	98.97	99.20	99.20	94.73	94.62	
U86	98.14	98.09	98.58	97.57	95.45	94.94	85.70	85.35	
U90	97.66	97.66	97.81	97.38	97.59	97.35	85.67	85.42	
U91	97.19	97.19	97.19	97.19	97.40	97.62	77.27	75.76	
U94	99.39	99.39	99.05	99.05	99.25	99.05	96.54	96.54	
U95	97.59	97.53	97.77	97.23	96.46	95.71	83.96	84.14	
U100	97.87	98.02	98.07	97.02	97.12	97.02	88.65	88.65	

Table 4.3 Pairwise comparisons of the CDS features of HHV-6A/-6B strains and CI-HHV-6A isolates to HHV-6A strain U1102. 'i' denotes incomplete CDS sequence available. Red colour indicates read coverage across this CDS was below the defined cut-off for accurate sequence resolution.

Using these comparisons, mean percentage divergence was calculated across HHV-6A and CI-HHV-6A genes. The mean divergence observed between HHV-6A and CI-HHV-6A genes was 1.5%. Therefore to determine conserved genes, a cut-off of those with a mean divergence of lower than 1% was applied. These parameters defined a gene set comprised primarily of previously defined *Herpesviridae* core genes [McGeoch et al. 2006], with the addition of the Roseolovirus conserved genes U4, U22, U24A, U25, U51 and U73, table 4.4. Similarly, for the definition of divergent genes a cut-off of greater than 2% was applied, table 4.4, these included all of the genes which define the speciation of HHV-6A and HHV-6B [Dominguez et al., 1999; Gompels & Kasolo, 2006; Ablashi et al., 2013] in addition to genes U13 and U15 which remain to be fully characterised. This gene set includes abundant tegument proteins which are CD4 and CD8 T cell targets, as well as membrane envelope glycoproteins, targets for neutralising antibody. In addition to genes whose products are involved in gene expression, cell cycle and immune regulation. The immediate early gene, U90, in all CI-HHV-6A isolates also retained the deletion observed in HHV-6A strains which renders them susceptible to type I interferons [Jaworska et al., 2007; 2010]. Furthermore, the U54 of the three CI-HHV-6A isolates contained the divergent motif observed in HHV-6A strains, U54₂₉₃₋₂₉₇ GISTT, that distinguishes inhibition of transactivation of NFAT and subsequent IL-2 gene expression in HHV-6B from HHV-6A [Iampietro et al., 2014]. Greater diversity could be observed in CI-HHV-6A, with 6 genes

(DR1, DR6, U71, U79, U86 and U95) displaying greater than 3% divergence, compared to only two restricted to the DR region in HHV-6A. The sequences of both the conserved and divergent genes of CI-HHV-6A are in appendices 7.5 and 7.6.

Conserved		Variable	
CDS	Gene Function	CDS	Gene Function
U4*	Maribavir resistance	DR1	Tegument protein; Putative DNA directed RNA polymerase
U22*	Glycoprotein	DR6	Binds p41 DNA processivity factor U27, inhibits replication, G2/M arrest
U24A*	Glycoprotein	U11	Tegument phosphoprotein
U25*	Tegument protein	U13	
U27	Processivity subunit of DNA polymerase complex	U14	Virion tegument protein, p53 interaction cell cycle
U28	Large subunit of ribonucleotide reductase	U15	
U29	Component of intercapsomeric triplex between hexons and pentons	U19	IE-B protein
U30	Interacts with U31 protein	U47	Membrane glycoprotein gO complexes with gH/gL
U31	Huge virion protein; interacts with U30 protein; influences release of DNA from capsids during entry	U54	Virion transactivator, U54A transactivates NFAT
U32	Small capsid protein; Capsid transport	U65	Tegument protein
U34	Inner nuclear membrane protein; nuclear egress	U71	Myristylated tegument protein
U35	DNA packaging	U79	DNA replication
U36	Capsid transport	U86	IE2, IE-A protein; interacts with transcriptional material
U37	Nuclear matrix protein	U90	IE1, IE-A transactivator; enhances activation by IE2
U38	DNA polymerase catalytic subunit	U91	Membrane glycoprotein
U39	Membrane glycoprotein gB	U95	IE gene
U40	Terminase subunit	U100	Membrane glycoprotein gQ complexes with gH/gL binds CD46
U41	Single-stranded DNA-binding protein		
U42	Tegument protein		
U43	Component of DNA helicase-primase		
U44	Tegument protein		
U45	Deoxyuridine triphosphatase		
U46	Membrane glycoprotein gN complexes with gM		
U51*	Chemokine receptor		
U53	Protease; Capsid scaffold protein maturation		
U56	Capsid protein		
U57	Major capsid protein		
U69	Serine-threonine protein kinase		
U70	Deoxyribonuclease		
U72	Membrane glycoprotein gM complexes with gN		
U73*	Origin binding protein		
U74	Component of DNA helicase-primase		
U81	Uracil DNA glycosylase		
U82	Membrane glycoprotein gL complexes with gH and gO		

Table 4.4 CI-HHV-6A conserved and variable gene sets. Star denotes not part of defined *Herpesviridae* gene set, however, all are part Roseolovirus conserved genes [McGeoch et al. 2006].

Concatenations of these gene sets were multiply aligned. Subsequently, regions of sequence were removed and loci with gapping characters before the calculation of phylogenetic relationships amongst the CI-HHV-6A strains and their relationship to HHV-6A and HHV-6B strains, figure 4.6. Analysis of the conserved gene set showed that the CI-HHV-6A strains clustered independently from circulating viral strains, while there was evidence for recombination in some of the divergent genes, which showed mixed branching relative to the HHV-6A strains. Overall these analyses show that CI-HHV-6A genomes diverge at a set of genes which encode products involved in gene regulation, host infection, cellular tropism and immune regulation. Many of these genes are involved in the speciation of herpesviruses.

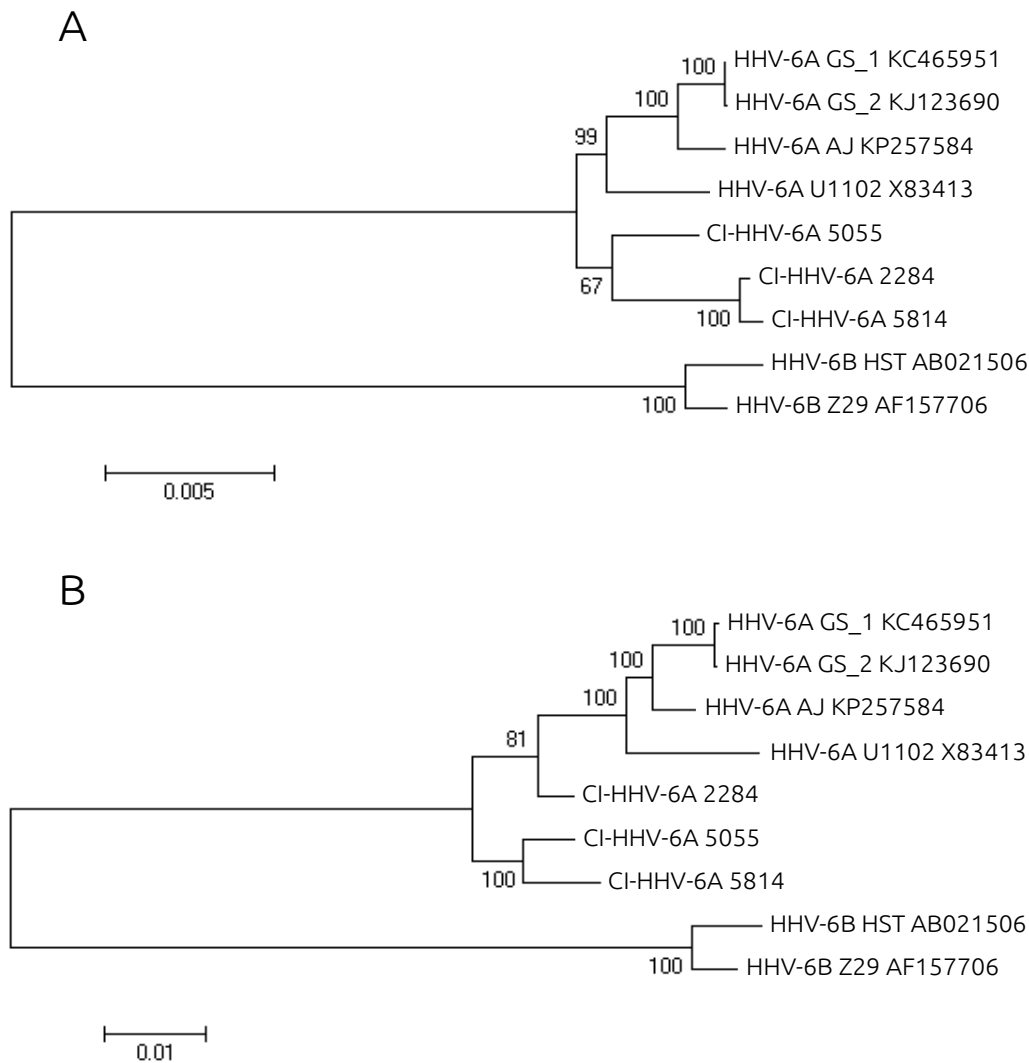


Figure 4.6 Phylogenetic analyses of CI-HHV-6A and HHV-6A/-6B conserved and variable gene sets. Maximum likelihood estimation on gene set concatenations with bootstrapping (1050 replicates). (A) CI-HHV-6A conserved genes (B) CI-HHV-6A variable genes. Gene sets as defined in Table 4.5. Accession numbers of all strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site.

4.5 Cis-acting Sites Required for DNA Packaging, Genome replication and Gene Regulation are Conserved in CI-HHV-6A

While not conclusive, evidence supporting the *in vivo* reactivation of CI-HHV-6A and -6B has been suggested by a number of means [Hall et al., 2010; Gravel et al., 2013a; Endo et al., 2014; Arbuckle et al., 2010; Prusty et al., 2013b; Huang et al., 2014]. Here, the reactivation potential of CI-HHV-6A was further assessed. As highlighted in section 4.3, all 85 genes previously defined for HHV-6A strains could all be identified in the CI-HHV-6A samples with no major indels. As the gene complement for genome reactivation/replication appeared to be present in the genomes, attempts were initially made to identify and characterise a number of the other highly conserved requirements for genome reactivation/replication and signatures of actively replicating genomes.

Cis-acting signals for genome replication were also conserved. The origin of lytic

replication (*oriLyt*), this motif, located between U40 and U41 ORFs in the UL region of HHV-6A and -6B, contains sequences bound by the viral origin binding protein, U73, to initiate the recruitment of the core DNA replication machinery [Inoue et al., 1994; Krug et al., 2001; Turner et al., 2002]. The 1.3kb *Mori oriLyt* [Turner et al., 2002] consists of two main regions, the first contains two origin binding sequences (OBP1,2) which share homologues with alpha- but not CMV herpesviruses [Dykes et al., 1997]. The second domain includes imperfect direct repeats (IDR). These are present as a tandem duplicated array in HHV-6B (IDR1,2). HHV-6A strains contain a tripartite array (IDR1,2,3), with a third IDR which is divergent from IDR1/2 and specific to HHV-6A [Turner et al., 2002]. Phylogenetic analyses of the CI-HHV-6A *oriLyts* with HHV-6A strains, figure 4.7, revealed all CI-HHV-6A *Mori* sequences clustered within the HHV-6A strains (5055 with U1102; 2284 and 5814 with GS/AJ). Closer inspection of alignments of this region, shown in appendix 7.7, demonstrate that all the CI-HHV-6A samples possessed OBP sequences which are identical to HHV-6A and -6B. In addition to the three IDR sequences of HHV-6A with only three SNPs across these sequences. Together this indicates the structure of the functional efficient origin of lytic replication is maintained and despite the repetitive structure observed here, it is among the most conserved regions of the genomes.

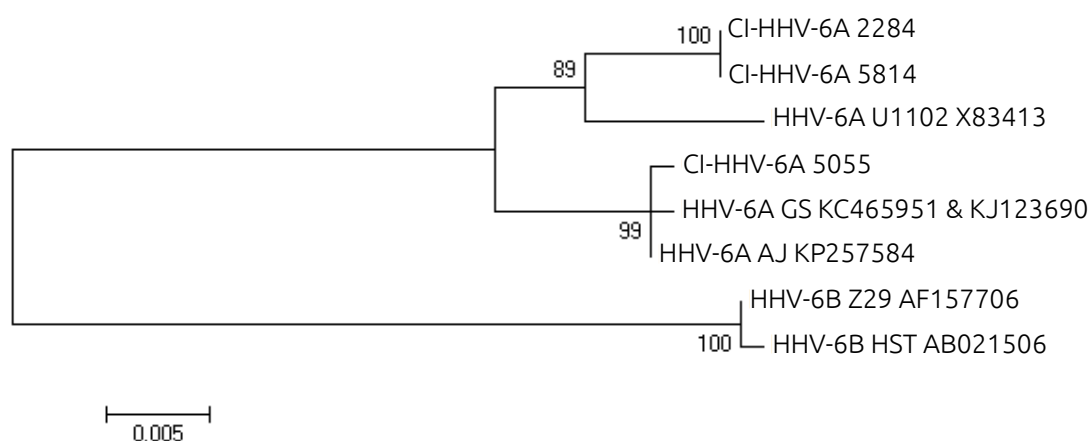


Figure 4.7 Phylogenetic analyses of the origin of lytic replication (*Mori*) of CI-HHV-6A isolates and HHV-6A/-6B strains. Maximum likelihood estimation with bootstrapping (1050 replicates) on the maximally efficient origin of replication defined for HHV-6A strain U1102 [Turner et al., 2002]. Accession numbers of all strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site. Full alignment shown in appendix 7.7.

The DNA packaging motifs, *pac1* and *pac2*, which flank the DR regions at the genomic termini are required for the cleavage of whole genome length subunits from rolling-circle generated concatemers and their subsequent capsid packaging [Deiss et al., 1986; Thomson et al., 1994a; Gompels & Macaulay, 1995; Deng & Dewhurst, 1998; Turner et al., 2002]. The *pac* sites closest to the UL genome region could be accurately resolved (*pac2* site from DR_L and *pac1* site from DR_R). But those located directly adjacent to the genomic termini could not be resolved. However, as found previously [Arbuckle et al., 2010; Huang et al., 2014] and confirmed for a number of CI-HHV-6A and -6B samples including 5055 (section 4.2), the *pac2* site at immediate end of the DR_R is absent in CI-HHV-6A and -6B genomes believed to result from the mechanism of chromosomal integration during the viral integration event. In addition, attempts at mapping the DR_L termini of CI-HHV-6A and -6B have indicated that the *pac1* site at the immediate end of the opposite side of the genome is also absent [Ohye et al., 2014; Huang et al., 2014]. The identifiable CI-HHV-6A

pac sequences were aligned with those of HHV-6A and -6B strains, figure 4.8. This showed both pac sites are retained in the CI-HHV-6A genomes with only a single SNP observed in the pac1 site, similar to that observed between circulating viral strains.

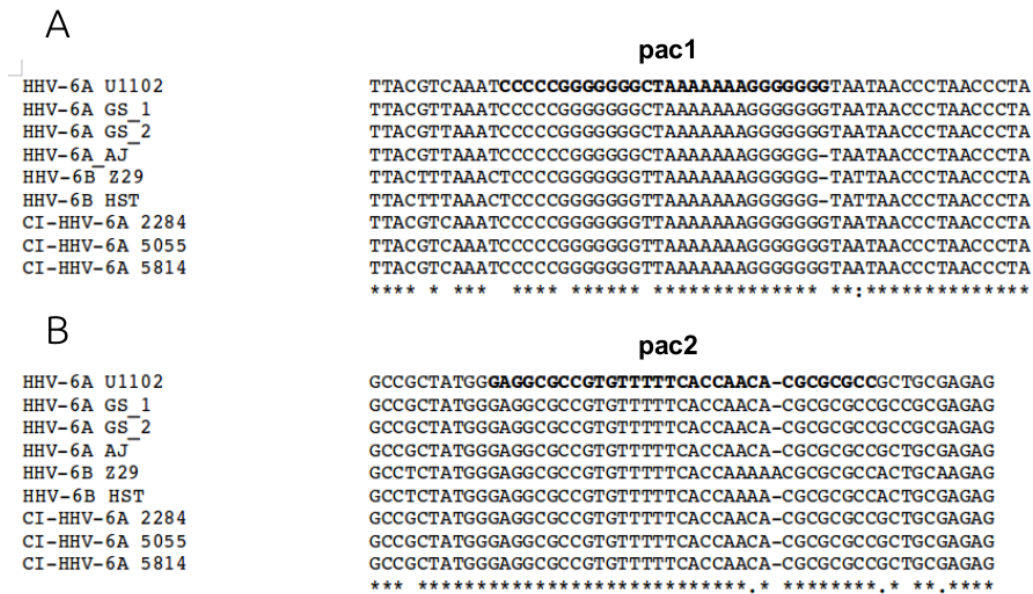


Figure 4.8 Alignment of CI-HHV-6A and HHV-6A/-6B pac sites with those of the prototypical strain HHV-6A U1102. (A) pac 1 site. (B) pac 2 site. Embolden sequence indicates the defined HHV-6A strain U1102 pac sites. HHV-6A strain GS genome sequences, GS_1 = accession KC465951, GS_2 = accession KJ123690.

Localised CpG suppression across the immediate early (IE) gene region, observed amongst a number of betaherpesviruses [Gompels et al., 1995], is also indicative of gene regulation by methylation [Honest et al., 1989; Martin et al., 1991b; Gompels et al., 1995]. CpG plots across the CI-HHV-6A genomes, figure 4.9, indicated that these were also present in all the CI-HHV-6A samples. While, all four of the recently defined microRNAs (miRNAs) of HHV-6B [Tuddenham et al., 2012], conserved at the structural and sequence levels in HHV-6A, also appear to be retained in CI-HHV-6A genomes, figure 4.10. However, issues with the assembly of the DRs due to the nature of their repeat sequences, as discussed in section 4.3, meant all the miRNAs could not be resolved for all samples. Taken together, the presence of these features imply that the CI-HHV-6A genomes may be competent for reactivation and viral replication.

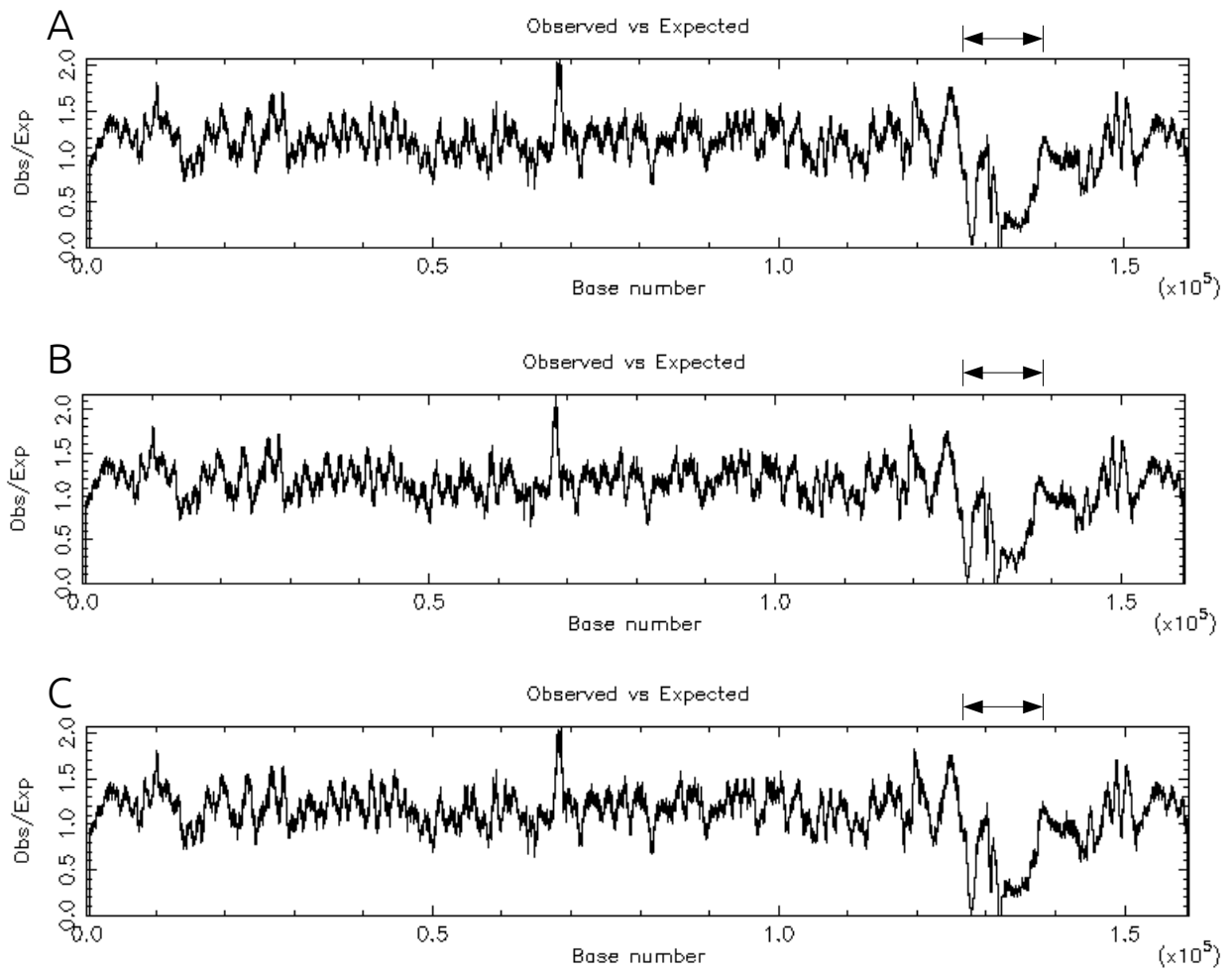


Figure 4.9 CpG frequency plots across the CI-HHV-6A genomes. (A) CI-HHV-6A isolate 2284 (B) CI-HHV-6A isolate 5055 (C) CI-HHV-6A isolate 5814. Arrows indicate CpG suppression at IE gene region corresponding to the genomic region from 125989 bp – 136829 bp.

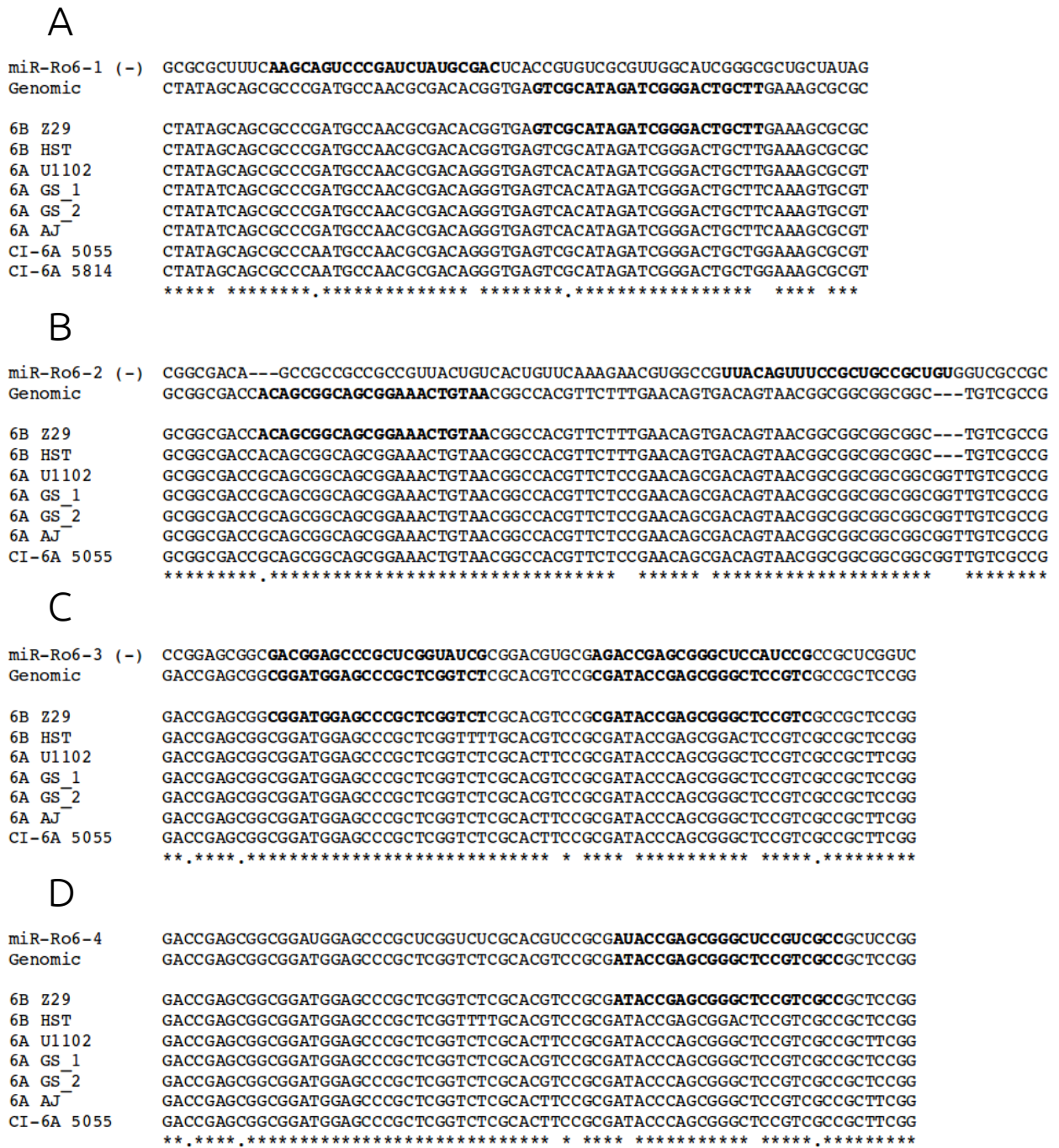


Figure 4.10 Conservation in CI-HHV-6A of the pre-miRNA sequences identified in HHV-6B [Tuddenham et al., 2012]. (-) denotes negative strand sequence. GS_1 = accession KC465951, GS_2 = accession KJ123690. Emboldened sequence indicates mature miRNA sequences.

4.6 Deep Sequencing Shows Minor Variants in CI-HHV-6A Genomes with Gene Expression

Previous work in this group, described fully in Tweedy et al., 2015a, made attempts to analyse cDNA samples from the German and Czech CI-HHV-6A and -6B patient cohorts, for gene expression from the integrated genomes. For this, seven patients from a German myocarditis/cardiomyopathy cohort were followed at different time points using primers

specific for a select group of HHV-6A, -6B genes, primer sequences also outlined in section 2.3.3. From this subset, four displayed evidence for gene expression, from multiple loci, at certain bleeds, but not at others. With sequencing showing these were identical to the integrated genotypes and distinct from known circulating HHV-6A and HHV-6B strains. The cDNA observed from one of these, the U83 gene, was the complete unspliced gene, which is only expressed after DNA replication [French et al., 1999].

Samples from two of these patients, CI-HHV-6A 2284 and 5055 (section 4.2), were also those which had been subjected to whole genome sequencing by NGS described above. With one patient, 2284, showing evidence for gene expression from integrated virus, while the other, 5055, appeared silent. This provided an opportunity to investigate the relationship between viral gene expression and reactivation, through deep sequencing analysis. For this minor variant analysis was performed on the CI-HHV-6A genomes and HHV-6A strain AJ (section 3.4) at two loci, U38 and U83, which had previously been sequenced by Sanger capillary sequencing [Tweedy et al., 2015a]. SNPs were called in comparison to the HHV-6A strain U1102 reference genome, with all SNPs identified reflecting those observed by Sanger capillary sequencing. To investigate minor variants a cut-off for SNP sensitivity was applied using an internal database control of a CG>GC inversion in the U83 gene CDS, identified by Dewin et al., 2006 and confirmed here in section 3.3, was still present in the U1102 reference sequence. Here, the original reference sequence was not called at read depths up to 6148, indicating a level of detection of at least 0.02%; a cut-off of 1 log above this was applied to verify minor variants, table 4.5.

SNPs		CI-HHV-6A						HHV-6A		
U38A Polymerase Position	SNP	5055 (17p)			2284 cDNA+			AJ		
		TR	SNP	REF	TR	SNP	REF	TR	SNP	REF
57357	G-A	7431	7422	8 (0.1%)	126	79	47 (37%)	6559	6559	-
57392	C-T	7537	7537	-	124	73	51 (41%)	6856	6856	-
57426	A-G	7541	7541	-	124	124	-	6959	6941	18 (0.3%)
57498	T-C	7811	7811	-	128	80	48 (31%)	7172	7172	-
57561	C-T	7125	7125	-	108	74	34 (32%)	7098	7098	-
57564	G-A	7070	7070	-	109	109	-	7063	7049	14 (0.2%)
57702	C-T	6973	6973	-	119	119	-	7369	7351	16 (0.2%)
57722	T-C	7053	7045	8 (0.1%)	125	125	-	7406	7406	-
57752	C-T	7133	7133	-	124	124	-	7300	7280	20 (0.3%)
U83 Chemokine										
123484	G-A	1281	1281	-	406	406	-	5273	5249	23 (0.4%)
123504	G-GTT-long / G-GT-short	1269	1105 G-GTT-long	0 (0%)	388	361 G-GT-short	6 (2%)	5088	4748 G-GTT-long	7 (0.1%)
123590	G-T	1101	1096	0 (0%)	347	346	0 (0%)	5162	5158	1 (0.02%)
123633	C-T	1063	1063	-	361	361	-	5420	5388	31 (0.6%)
123687	A-G	1093	1093	-	406	406	-	5683	5653	30 (0.5%)
123733	A-G	1101	1096	0 (0%)	429	395	34 (9%)	6419	6419	-
123740-41	CG-GC*	1108	1108	0 (0%)	411	411	0 (0%)	6148	6145	0 (0%)

Table 4.5 Minor variants in CI-HHV-6A isolates identified by deep sequencing. Table shows the position of SNPs in the CI-HHV-6A sequence reads. Total sequence read depth at the site (TR). The number of sequence reads calling a single nucleotide polymorphisms (SNPs) different from the reference strain HHV-6A U1102 (SNP). The number and percentage of sequence reads calling a SNP the same as the reference (REF). (17p) indicates chromosome 17p telomere integration. Shading shows minor variant SNPs above the defined cut off of 0.2 % (see methods section 2.8.6).

The results showed minor variants were identified in samples with evidence for virus gene expression, both strain AJ infected cells and the CI-HHV-6A sample positive for cDNA, isolate 2284. But there were no minor variants identified where cDNA was negative, in CI-HHV-6A isolate 5055. Although the minor variants shared similarities to the strain U1102 reference genome, SNP patterns across the entire CDS indicated they were distinct. Overall, these results indicated HHV-6A superinfection in CI-HHV-6A patients may lead to gene expression from the CI-HHV-6A genomes, as the detected cDNA was identical to the integrated genotype. Furthermore, the detection of the full length unspliced form of U83 in cDNA hinted a mechanism by which superinfection with HHV-6A may reactivate CI-HHV-6A.

4.7 Discussion

This chapter confirms the integration of both HHV-6A and -6B into the subtelomeric region of chromosome 17p13.3 and suggests this is a prevalent integration site. The observation of a conserved structure to the integration site junctions of the CI-HHV-6A samples at this chromosomal location suggests they likely originated from a single rare ancestral germ-line integration event which was subsequently expanded in the human populations studied here as a polymorphism. The apparent high prevalence of integration at this chromosomal location implicates either some preference for integration at this site and/or that the ancestral germ-line integration event pre-dates that seen at other chromosomal locations. Whole exome analyses of CI-HHV-6A genomes determined by NGS suggested divergence at a set of genes which encode products involved in gene regulation, host infection, cellular tropism and immune regulation, all markers previously used for herpesvirus speciation. Key genomic features required for viral replication, and likely any reactivation mechanism, were also retained in the CI-HHV-6A genomes. Finally, the identification of minor variants in CI-HHV-6A samples with previous evidence for gene expression, indicated a mechanism by which superinfection with circulating viral strains may permit gene expression and potentially reactivation from the chromosomally integrated form.

Initially attempts were made to identify CI-HHV-6A and -6B chromosomal integration site junctions. For this, a minority of the subtelomeric sites where viral integration has been observed to occur were chosen as potential targets (10q, 11q, 17p and 18q). These were chosen as either anecdotally the prevalence of integration at these sites appeared higher, or as they represented sites with well defined subtelomeric sequences for which validated primer sequences were available. Of those chromosomal locations trialled expected amplicons were only obtained utilising the primers designed to amplify integrated viral genomes from the subtelomeric region of chromosome 17p. Thus the integration sites of those not on chromosome 17 within this sample group, remains to be elucidated.

Similar to previous findings, the results here suggest both HHV-6A and -6B have integrated into the subtelomeric region of chromosome 17 (17p13.3) [Morris et al., 1999; Torelli et al., 1995; Morissette & Flamand, 2010; Nacheva et al., 2008; Ward et al., 2006; Clark et al., 2006b; Arbuckle et al., 2010; Luppi et al., 1993]. Furthermore, the results provided evidence that chromosome 17p13.3 was a prevalent site for HHV-6A integration. This was in agreement with anecdotal observations from previous studies, where chromosome 17p13.3 has often been identified as a site of CI-HHV-6 integration. Since chromosome 17p13.3 is frequently found to be one of the shortest telomeres in somatic cells [Martens et al., 1998], suggestions have been made that the apparent preference for this chromosomal location may result from an evolutionary advantage conferred by integration there, such as to viral reactivation potential. In this regard, *in vitro* integration experiments have noted integration at chromosome 17p during productive infection [Arbuckle et al., 2010]. However, at least 8 other chromosomal integration sites have been identified to date for inherited CI-HHV-6A [Watanabe et al., 2008; Nacheva et al., 2008; Ward et al., 2006; Huang et al., 2014; Clark et al., 2006b; Hubacek et al., 2009; Arbuckle et al., 2010; Tanaka-Taya et al., 2004], and at least two (11q and 18q) others during *in vitro* work [Arbuckle et al., 2010]; many of which are longer both in overall length and, despite heterogeneity, also in the telomeric region [Lander et al., 2001; Lansdorp et al., 1996]. This suggests that integration into the shortest chromosome may not be a requirement for the virus. However, observations have been made that the virally integrated chromosome is

frequently the shortest in somatic cells, even when compared to the length of that of chromosome 17p [Huang et al., 2014]. Thus the viral integration event itself may result in telomere shortening, regardless of chromosomal location either merely as a consequence of the integration or should such shortening aid viral reactivation. So while integration into this site on chromosome 17 may still provide some advantage or result from a characteristic of the still undefined integration mechanism, there may be other factors involved.

The identification of a common integration site junction architecture in the CI-HHV-6A samples appearing to be integrated at chromosome 17p13.3; sharing specific sub-telomeric repeat structures plus identical adjacent sequences from the human sub-telomere and virus DR-R termini; indicates an alternative, although not wholly incompatible, explanation for the apparent high prevalence of integration at this site observed in inherited CI-HHV-6A. Amongst HHV-6A strains the DR regions represent sites displaying some of the highest levels of variation, 5-8%, in the genome. Indeed it is variability in the repeat structures here which contribute greatly to the different genome sizes observed amongst HHV-6A strains. Furthermore, subtelomeres are known to possess complex degenerate repeat structures [Ambrosini et al., 2007] and in the limited information available regarding the chromosome 17p13.3 subtelomeric region, variation could be observed in these telomeric repeat structures. Thus the shared integration site junction structure in the CI-HHV-6A samples appears to indicate a common integration event which has given rise to some of the CI-HHV-6A genomes observed within the samples examined here. This could potentially be attributed to close ancestral relationship between the individuals in this sample panel, resulting from their similar geographic origins, Germany and the Czech Republic. The fact that the integration junction sequences shared notable similarities to one previously identified for CI-HHV-6A, also suspected to be at this chromosomal location and from distinct geographic origin, Israel [Cann et al., 2002; Huang et al., 2014]; suggested this viral integration event potentially occurred earlier in human history with inheritance through a number of generations. Thus the apparent high prevalence of integration at this chromosomal location in inherited CI-HHV-6A could result from an integration event into the germ-line which predates those seen at other chromosomal locations, followed by a subsequent expansion as a polymorphism by Mendelian inheritance. The acquisition of further genome and integration site junction data, either directly from CI-HHV-6A samples or alternatively as a by product of human genomic sequencing, will help to clarify this.

The sequence of the integration site junctions, confirmed the loss of nucleotide sequence from far right end of genome including the pac 2 site. This is consistent with models for a role of homologous recombination between the perfect telomeric-like repeats of the T2 region of the DR_R and chromosomal telomeric repeat sequences, in the viral integration event [Arbuckle et al., 2010, 2013; Huang et al., 2014]. In addition to this, there are also accounts of the loss of the pac 1 site from the far left end of the genome following integration [Ohye et al., 2014; Huang et al., 2014]. The loss of these specific signals for DNA cleavage and packaging from the ends of the genome have implications for the mechanism of viral integration as well as potential models of reactivation. Indeed models have been proposed in which two pac sites may represent the minimal complement required for reactivation via a t-loop formation mechanism [Huang et al., 2014; Prusty et al., 2013b]. Expansion of the T1 imperfect telomeric repeat regions, respective to that observed in circulating viral strains, has been noted for the CI-HHV-6A and -6B [Ohye et al., 2014; Huang et al., 2014]. While this could be a feature of the virus which resulted in the

germ-line integration events. Suggestions have also been made that this expansion results from erosion of viral sequence from the genomic termini, which activates a homology directed DNA repair mechanism as a prerequisite for the viral integration event [Ohye et al., 2014]. Therefore, any potential reactivation mechanisms must in some way account for the loss of these pac sites if replication competent virus is to be produced.

Preliminary work on characterisation of the these CI-HHV-6A and -6B samples, at 4 loci, discussed partly in section 5.2 and fully in Tweedy et al., 2015a, suggested CI-HHV-6A genomes may show greater diversity than those of CI-HHV-6B. This was investigated further through the NGS of three of CI-HHV-6A samples, for which sufficient template was available. An LPCR approach for viral genomic enrichment was employed as a result of sample limitations. However, the variability in coverage observed between the amplicons for HHV-6A, described in section 3.2, was much more pronounced for the CI-HHV-6A samples. As highlighted previously inaccuracies in the quantification and equimolar pooling of the amplicon DNA likely had an effect on sequence uniformity. However, the greater variance implicated additional factors contributed. The primer sets used for the CI-HHV-6A samples were the same used for HHV-6A, where optimisation had been performed for HHV-6A species. While results of the PCR reactions for CI-HHV-6A did not show clear signs of amplification difficulties; optimisation for each viral strain would be of benefit with this target enrichment approach adding further time consuming steps to the process in comparison to alternative methods.

However, through the use of both NGS and gap filling with Sanger sequencing, the complete coding complement of the genome could be determined for CI-HHV-6A isolate 2284, as well as sufficient genome sequence from CI-HHV-6A isolates 5055 and 5816 to assess conserved and divergent gene sets. This showed the integrated virus genomes possess the full gene complement required for viral replication. Analysis of the integrated genomes showed retention of many of the cis acting signals required for viral reactivation, with a high level of conservation. So in addition to the full genome complement, the integrated genomes also retain many of the genomic features likely required for any viral reactivation mechanism.

Analysis of the conserved gene set showed that the CI-HHV-6A strains clustered independently from circulating viral strains, while there was evidence for recombination in some of the divergent genes, which showed mixed branching with HHV-6A strains (a feature which is further investigated in Tweedy et al., 2015a). The divergent gene set consisted of 17 genes encoding products involved in roles in transcriptional regulation, virus infection, cell cycle and immune modulation. These are often features which are regarded as markers of speciation amongst herpesviruses. Indeed, the CI-HHV-6A divergent gene set included all of the genes which defined the recent speciation of HHV-6A and -6B [Dominguez et al., 1999; Gompels & Kasolo, 2006; Ablashi et al., 2014]. Together, the integration site junction and gene divergence analyses indicate integration of an ancestral HHV-6A/6B variant into the germline at this chromosomal location. While the reactivation/gene expression with possible recombination suggest that expansion of the integrated form in the human population could serve as source of emergent infection or contribute to gene divergence.

Finally, the depth of the NGS read coverage in some regions permitted analysis of minor variants in the CI-HHV-6A samples. This allowed for comparison between a sample with noted integrated gene expression, as identified by cDNA analysis, and one without. These

results suggested superinfection with HHV-6A strains, in the sample with evidence for integrated virus gene expression. This is consistent with a model where superinfection with HHV-6A or -6B has the potential to result in gene expression from/reactivation of the integrated virus. There is evidence of superinfections causing reactivation of latent and integrated virus. Reactivation of latent HHV-6B has been noted following superinfection with HHV-7 [Katsafanas et al., 1996; Tanaka-Taya et al., 2000]. Furthermore co-infection with *Chlamydia trachomatis*, has also been suggested to cause reactivation of CI-HHV-6A via a t-circle excision mechanism [Prusty et al., 2013a; Prusty et al., 2013b]. As new methods allowing for greater accuracy in distinguishing between the viral species, such as ddPCR [Sedlak et al., 2014; Leibovitch et al., 2014], and deep sequencing becoming a viable solution for detection of viral populations [Beerenwinkel & Zagordi, 2011; Giallonardo et al., 2014], the involvement of superinfection in integrated genome gene expression/reactivation should become clearer.

In summary, analysis of the integration site junctions of CI-HHV-6A at chromosome 17p indicated a shared genetic lineage between virally integrated genomes. The application of NGS methodologies investigated in the previous chapter allowed characterisation of three CI-HHV-6A genomes. These revealed CI-HHV-6A genomes are intact, possessing the full complement of genes identified for HHV-6A strains, in addition to key cis acting signals important for reactivation. Comparisons of these integrated genomes to circulating viral strains highlighted divergence at a set of genes which have been used previously to mark herpesvirus species. The minor variant analysis supported a mechanisms by which HHV-6A/-6B superinfection may result in gene expression from, or reactivation of the integrated CI-HHV-6A genome.

Chapter 5: Variation and Functional Analysis of the Viral Chemokine Receptor U51

5.1 Introduction

Both HHV-6A and -6B have been associated with a variety of inflammatory pathologies, such as myocarditis and encephalitis. Therefore, the encoded chemokine modulatory capabilities of these viruses may represent important virulence factors mediating viral replication and inflammatory pathology. Expression of these genes from the integrated genomes, could affect inflammatory pathology in general since every nucleated cell would have potential for expression.

These viruses encode two GPCRs, U12 and U51, which have been demonstrated to interact and signal in response to key inflammatory ligands [Gompels et al., 1995; Dominguez et al., 1999; Isegawa et al., 1999; Isegawa et al., 1998; Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008]. In cases of the germ-line integrated forms of these viruses, immunomodulatory effects may be more pronounced through expression in a greater range of cells or potentially even affects on immune system development. In this regard, other work in this laboratory on cDNA analysis from CI-HHV-6A/-6B samples presented evidence for gene expression from U51 and the viral chemokine, U83 [Tweedy et al., 2015a]. Furthermore, this work also showed that in the integrated form of the viruses, the viral chemokine, U83, was found to be predominantly in a long active form, less frequently observed in circulating viral strains. Indicating differences existed between inflammatory modulators in the integrated form of the virus when compared to that observed in circulating viral strains. Therefore, here the viral chemokine receptor, U51, from CI-HHV-6A/-6B was investigated further.

U51 represents a relatively conserved gene, sharing 99% homology amongst strains of each of the HHV-6 species and 95% homology between the species (accession numbers: X83413, KC465951, KP257584, AF157706 and AB021506). Despite this, the determination of the HHV-6A strain AJ genome sequence, described in chapter 3 of this thesis, and recent reports of the HHV-6A strain GS genome sequence [Gravel et al., 2013a; Bhattacharjee et al., 2014] indicated the presence of variation in the amino acid sequence of the amino-terminal region of the U51 gene (U51A) from that seen in the prototypical HHV-6A strain U1102. Moreover, there was similar variation identified in the U51 CDS of three CI-HHV-6A genomes subjected to NGS analysis, described in chapter 4.

According to current evidence, the amino-terminal domain of a chemokine receptor represents a key determinant in the chemokine-receptor interaction, receptor activation, and associated signalling properties. Therefore, any variation here could affect specificity and function of the receptor. Additionally, U51 also displays a number of unique features, when compared to other viral and human chemokine receptors. Most notably, an unusually short amino-terminal domain in comparison to other chemokine receptors; predicted to be 17 amino acids in U51 compared to a range from 26 amino acids, CX3CR1, to 65 amino acids, DARC, for human chemokine receptors. This combined with a broad ligand binding profile showing a combination of constitutive and ligand-inducible signalling [Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008], which outgroups the entire human CC chemokine receptor group. Therefore, changes to this domain as shown in the integrated genomes, could have large immunomodulatory effects. As such analyses were initiated at this gene to start to evaluate the consequences of the variation

observed in the integrated genomes and circulating virus.

To examine any functional differences, firstly, variation in the U51 gene was investigated further through the screening of a panel of samples from patients with CI-HHV-6A or -6B. Subsequently, molecular modelling software was employed to guide mutagenesis of the U51A amino-terminal domain and for the generation of chimeric receptor constructs with the human chemokine receptor CCR3. Generated constructs were expressed in mammalian expression systems and subjected to preliminary functional analysis. These included measurement of surface expression, as well as activation using chemotaxis and actin polymerisation assays.

5.2 U51 Variation in CI-HHV-6A and -6B

A panel of 57 CI-HHV-6A and -6B samples previously examined in this thesis (31 from patients reporting to a diagnostic facility in Berlin, Germany, with cardiac complications and 26 from patients reporting to a hospital in Prague, Czech Republic with lymphoproliferative or inflammatory diseases) were screened by PCR to amplify the complete CDS of the U51 gene from CI-HHV-6A and -6B strains, followed by nucleotide sequence analysis.

Firstly, conserved oligonucleotide primer sets were designed to amplify the U51 gene. Utilising a ClustalW2 [Larkin et al., 2007] alignment of the nucleotide sequences of the reference HHV-6A and HHV-6B genomes (strain U1102: accession X83413/RefSeq NC_001664 and strain Z29: accession AF157706/RefSeq NC_000898, respectively) species conserved sites flanking the protein coding sequence of the U51 gene were identified. Oligonucleotide primers, derived from these conserved sites, were used for both PCR amplification and Sanger sequencing of the complete protein coding sequence of the U51 gene from the CI-HHV-6A and -6B samples (methods section 2.3.3).

These primers amplified a predicted product of 1008 bp. Amplicons of the predicted size were amplified from 44/57 (77%) of the samples. The negative samples appeared mainly due to limited clinical sample since other regions had been previously amplified and sequenced in the laboratory in our subsequent publication which included the U51 data [Tweedy et al., 2015a]. Sanger sequencing was used to both confirm the correct identity of the PCR products and determine the complete CDS. 20/31 of the amplicons were from the German myocarditis/cardiomyopathy patient cohort and 24/26 from the Czech inflammatory disease set. BLAST searches [Camacho et al., 2009] of the nucleotide sequences, allowed species typing of the chromosomally integrated samples. Overall 26/44 (59%) of the amplicons were CI-HHV-6A and 18/44 (41%) of the amplicons represented CI-HHV-6B. For U51A, 19/26 (73%) were from the Czech cohort and 7/31 (23%) from the German cohort. For U51B, 5/26 (19%) were from the Czech cohort and 13/31 (42%) from the German cohort.

Subsequently, phylogenetic analyses were performed on these CI-HHV-6A and -6B sequences with comparisons made to available U51 sequences from circulating HHV-6A (U1102 accession X83413, GS accessions KC465951 and KJ123690 and AJ accession KP257584) and HHV-6B strains (Z29 accession AF157706 and HST accession AB021506). The CI-HHV-6A and -6B sequences shared 99-100% identity within each virus species and 95% between the two species. All CI-HHV-6B sequences were identical to HHV-6B strain

HST, and distinct from Z29. Whereas, all the CI-HHV-6A U51A sequences showed greater divergence than CI-HHV-6B, with all being distinct from known circulating HHV-6A strains.

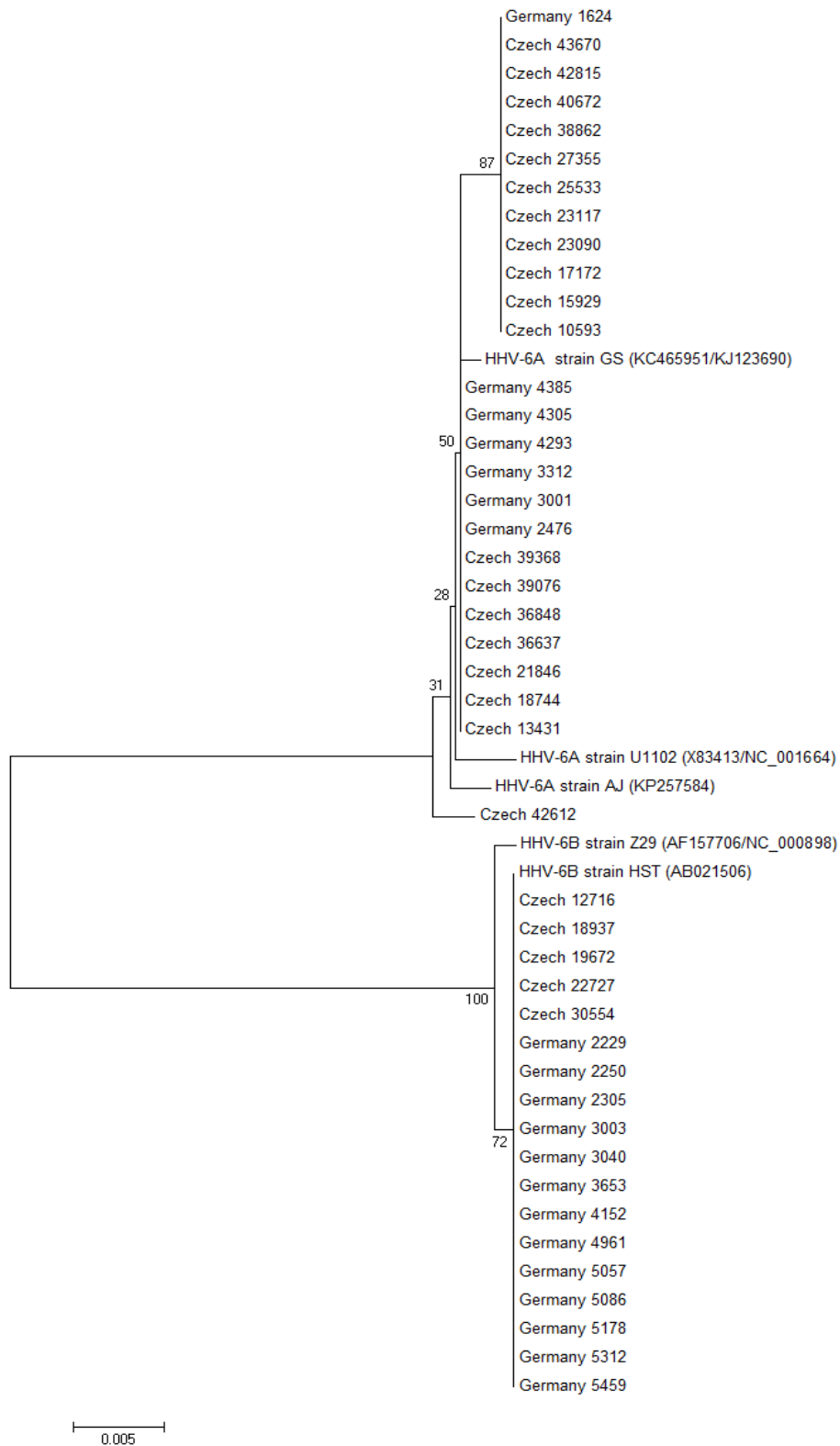


Figure 5.1 Phylogenetic analyses of U51 gene from CI-HHV-6A/-6B isolates with HHV-6A/-6B strains. Maximum likelihood estimation with bootstrapping (1050 replicates). Accession numbers of all strains used included in the figure. Branch length scale represents number of nucleotide substitutions per site.

To ascertain whether the observed divergence represented synonymous or non-synonymous substitutions, the translated amino acid sequences of U51 from HHV-6A, HHV-6B, CI-HHV-6A and CI-HHV-6B were aligned, figure 5.2. The coding region of U51B in the CI-HHV-6B samples were 100% identical to U51B of circulating HHV-6B strains, Z29 and HST. However, for the CI-HHV-6A strains there were 4 sites of amino acid variation in comparison to circulating HHV-6A strains, with the integrated forms differing from strain AJ at one site and strain U1102 at 3 sites. These non-synonymous SNPs encoded serine to proline (S156P) and tyrosine to histidine (Y167H) substitutions within the predicted second extracellular loop (ECL2), an isoleucine to leucine (I188L) substitution at the end of transmembrane domain five (TM5) and a glutamic acid to glycine (E2G) substitution in the amino-terminal domain. Of particular interest were the E2G and Y167H mutations as these represented charge changes in predicted extracellular domains. These small amino acid substitutions could have large effects, since these are external regions involved in receptor-ligand binding and the interactions involved in this binding are believed to rely heavily on charge-charge interactions [Blanpain et al., 1999; Szpakowska et al., 2012].

CI-HHV-6A	MEKETKSLAWPATAEFY	GWVFIFSSIQLCTMVLLTV	RFNSFKVGREY	AVF	50
HHV-6A GS	MEKETKSLAWPATAEFY	GWVFIFSSIQLCTMVLLTV	RFNSFKVGREY	AVF	50
HHV-6A AJ	MEKETKSLAWPATAEFY	GWVFIFSSIQLCTMVLLTV	RFNSFKVGREY	AVF	50
HHV-6A U1102	MEKETKSLAWPATAEFY	GWVFIFSSIQLCTMVLLTV	RFNSFKVGREY	AVF	50
CI-HHV-6B	MEKETKSLAWPATAEFY	GWVFIFSSIQLCTVVFLTV	RFNGFKVGREY	AVF	50
HHV-6B Z29	MEKETKSLAWPATAEFY	GWVFIFSSIQLCTVVFLTV	RFNGFKVGREY	AVF	50
HHV-6B HST	MEKETKSLAWPATAEFY	GWVFIFSSIQLCTVVFLTV	RFNGFKVGREY	AVF	50
	N-term	TM I	ICL I		
CI-HHV-6A	TFAGMSFNCFLLP	IKMGLLSGHWSLPRDFCA	ILLYIDDFS	SIYFSSWSL	VF 100
HHV-6A GS	TFAGMSFNCFLLP	IKMGLLSGHWSLPRDFCA	ILLYIDDFS	SIYFSSWSL	VF 100
HHV-6A AJ	TFAGMSFNCFLLP	IKMGLLSGHWSLPRDFCA	ILLYIDDFS	SIYFSSWSL	VF 100
HHV-6A U1102	TFAGMSFNCFLLP	IKMGLLSGHWSLPRDFCA	ILLYIDDFS	SIYFSSWSL	VF 100
CI-HHV-6B	TFAGMSFNCFLLP	IKMGLLSGHWSLPRDFCA	ILLYIDDFS	SAYFSSWSL	VF 100
HHV-6B Z29	TFAGMSFNCFLLP	IKMGLLSGHWSLPRDFCA	ILLYIDDFS	SAYFSSWSL	VF 100
HHV-6B HST	TFAGMSFNCFLLP	IKMGLLSGHWSLPRDFCA	ILLYIDDFS	SAYFSSWSL	VF 100
	TM II	ECL I	TM III		
CI-HHV-6A	MAIERINHF	CYSTPLLNENSKALAK	VCFP	IVWII	ISGVQALQMLNNYKATA 150
HHV-6A GS	MAIERINHF	CYSTPLLNENSKALAK	VCFP	IVWII	ISGVQALQMLNNYKATA 150
HHV-6A AJ	MAIERINHF	CYSTPLLNENSKALAK	VCFP	IVWII	ISGVQALQMLNNYKATA 150
HHV-6A U1102	MAIERINHF	CYSTPLLNENSKALAK	VCFP	IVWII	ISGVQALQMLNNYKATA 150
CI-HHV-6B	MAIERINHF	CYSTPLLNENSKALAK	VCFP	IVWV	ISGVQALQMLNNYKATA 150
HHV-6B Z29	MAIERINHF	CYSTPLLNENSKALAK	VCFP	IVWV	ISGVQALQMLNNYKATA 150
HHV-6B HST	MAIERINHF	CYSTPLLNENSKALAK	VCFP	IVWV	ISGVQALQMLNNYKATA 150
	ICL II	TM IV			
CI-HHV-6A	LQNETPQCFLAF	LRSGHDMWMLVYSVM	IPVMLVFI	YLY	SKNFMLLKDEL 200
HHV-6A GS	LQNETPQCFLAF	LRSGHDMWMLVYSVM	IPVMLVFI	YLY	SKNFMLLKDEL 200
HHV-6A AJ	LQNETPQCFLAF	LRSGHDMWMLVYSVM	IPVMLVFI	YLY	SKNFMLLKDEL 200
HHV-6A U1102	LQNETPQCFLAF	LRSGHDMWMLVYSVM	IPVMLVFI	YLY	SKNFMLLKDEL 200
CI-HHV-6B	LQNETPQCFLAF	LRSGHDMWMLVYSVM	IPVMLVFF	YLY	SKNFMLLKDEL 200
HHV-6B Z29	LQNETPQCFLAF	LRSGHDMWMLVYSVM	IPVMLVFF	YLY	SKNFMLLKDEL 200
HHV-6B HST	LQNETPQCFLAF	LRSGHDMWMLVYSVM	IPVMLVFF	YLY	SKNFMLLKDEL 200
	ECL II	TM V	ICL III		
CI-HHV-6A	STVTTYLCIYLL	LGTIAHLPKAGLSE	IESDKIFYGLR	DIFMALPVL	KVYY 250
HHV-6A GS	STVTTYLCIYLL	LGTIAHLPKAGLSE	IESDKIFYGLR	DIFMALPVL	KVYY 250
HHV-6A AJ	STVTTYLCIYLL	LGTIAHLPKAGLSE	IESDKIFYGLR	DIFMALPVL	KVYY 250
HHV-6A U1102	STVTTYLCIYLL	LGTIAHLPKAGLSE	IESDKIFYGLR	DIFMALPVL	KVYY 250
CI-HHV-6B	SSVTTYLCIYLL	LGTIAHLPKAALSE	IESDKIFYGLR	DIFMALPVL	KVYY 250
HHV-6B Z29	SSVTTYLCIYLL	LGTIAHLPKAALSE	IESDKIFYGLR	DIFMALPVL	KVYY 250
HHV-6B HST	SSVTTYLCIYLL	LGTIAHLPKAALSE	IESDKIFYGLR	DIFMALPVL	KVYY 250
	TM VI	ECL III	TM VII		
CI-HHV-6A	IPVMAYCMACD	DHTVPVRLCSIWLVN	LCKKCF	SCTRREKESD	LEVGIKML 300
HHV-6A GS	IPVMAYCMACD	DHTVPVRLCSIWLVN	LCKKCF	SCTRREKESD	LEVGIKML 300
HHV-6A AJ	IPVMAYCMACD	DHTVPVRLCSIWLVN	LCKKCF	SCTRREKESD	LEVGIKML 300
HHV-6A U1102	IPVMAYCMACD	DHTVPVRLCSIWLVN	LCKKCF	SCTRREKESD	LEVGIKML 300
CI-HHV-6B	ISAMAYCMACD	DHTVPVRLCSIWLVN	LCKKCF	SCTRREKESD	LEVGIKML 300
HHV-6B Z29	ISAMAYCMACD	DHTVPVRLCSIWLVN	LCKKCF	SCTRREKESD	LEVGIKML 300
HHV-6B HST	ISAMAYCMACD	DHTVPVRLCSIWLVN	LCKKCF	SCTRREKESD	LEVGIKML 300
				C-term	
CI-HHV-6A	K-	301			
HHV-6A GS	K-	301			
HHV-6A AJ	K-	301			
HHV-6A U1102	K-	301			
CI-HHV-6B	K-	301			
HHV-6B Z29	K-	301			
HHV-6B HST	K-	301			

Figure 5.2 U51A amino acid sequence variation. Alignment of U51 amino acid sequences for CI-HHV-6A/-6B and HHV-6A/-6B strains. All CI-HHV-6A amino acid sequences were identical, similarly all CI-HHV-6B sequences were identical, so both have been condensed to representative sequences here. Amino acid differences between HHV-6A strains and CI-HHV-6A have been highlighted yellow. Predicted locations of the extracellular domains have been coloured green: amino-terminal domain (N-term) and extracellular loops I – III (ECL I – III). Predicted transmembrane domains (TM I – VII) coloured black. Predicted intracellular domains (ICL I – III) and the C-terminal domain (C-term) have been coloured blue. Domain locations predicted from previous alignments by Milne et al., 2000, and updated using adjustments from GPCRHHM prediction software [Wistrand et al., 2006].

In the prevalent model of chemokine receptor activation, the “two-site” model, charge interactions are believed to play a key role in ligand binding specificity of a chemokine receptor [Szpakowska et al., 2012; Liou et al., 2014]. The E2G mutation observed in all the CI-HHV-6A genomes and a subset of circulating HHV-6A strains, represents the loss of a charged residue from the amino-terminal domain of pU51, thereby resulting in the loss of a net negative charge across the amino-terminal domain. As such, it was hypothesized that this substitution may be of functional significance to receptor-ligand interactions particularly in regard to ligand binding specificity. In addition, the variation in ECL2 could potentially play roles in receptor activation and the variation in TM5 may affect receptor activation and/or downstream signalling from the receptor. The functional effects of the N-terminal substitution were first investigated, since this could affect ligand specificity and alter cellular responses to chemokines.

5.3 Molecular Modelling of U51A Amino-terminal Domain-CCL2 Interaction

Molecular modelling was used initially to investigate the potential ligand specificity effects of the observed variation. To do this a model of the interaction of the receptor's amino-terminal domain with that of its most extensively characterised chemokine ligand, CCL2/MCP-1, was generated. The advantage of this was that both the tertiary structure of CCL2 and receptor interaction interfaces have previously been determined [Handel & Domaille, 1996; Lubkowski et al., 1997; Chakravarty et al., 1998; Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999; Lau et al., 2004]. In contrast, the structure of U51, like most human and viral chemokine receptors, has yet to be resolved resulting from the complexities of membrane association in the structural determination of GPCRs. In those cases where structural information is available for large protein binding GPCRs such as chemokine receptors, data on the amino-terminal domain is limited. It is often only present with a 5' truncation, if at all, which is believed to stem from either a disordered structure or the flexible nature of at least the immediate 5' end of this region of the protein [Wu et al., 2010; Park et al., 2012; Tan et al., 2013; Qin et al., 2015]. Indeed studies on the structures of N-terminal peptides of chemokine receptors complexed to chemokines also suggest a predominantly irregular loop structure to the amino terminal domain of the receptor, at least in the bound state, which binds a hydrophobic groove along the surface of the chemokine ligand [Skelton et al., 1999; Mizoue et al., 1999; Ye et al., 2000; Love et al., 2012].

To investigate possible similar structures in U51 of HHV-6A, a selection of bioinformatic software programs, utilising a variety of different prediction methods, were employed to model the secondary structure of amino-terminal domain. These predicted the 5' end of the domain had a disordered structure with an alpha-helical structure towards the 3' end of the N-terminal domain, figure 5.3A-C. Additionally a helical wheel plot [Schiffer & Edmundson, 1967] indicated that if modelled as an alpha helical structure the acidic residues of the domain would be clustered on one side of the helix with basic residues found on opposite face figure 5.3D. Since it was not possible to model a random coil structure for the U51 N-terminal domain and there was some indication of an alpha helical structure at least at the 3' end. An alpha helical structure was used to initiate a model investigating potential interactions with the chemokine. Wild type and mutant receptors were modelled as alpha helices using HyperChem software, figure 5.5.

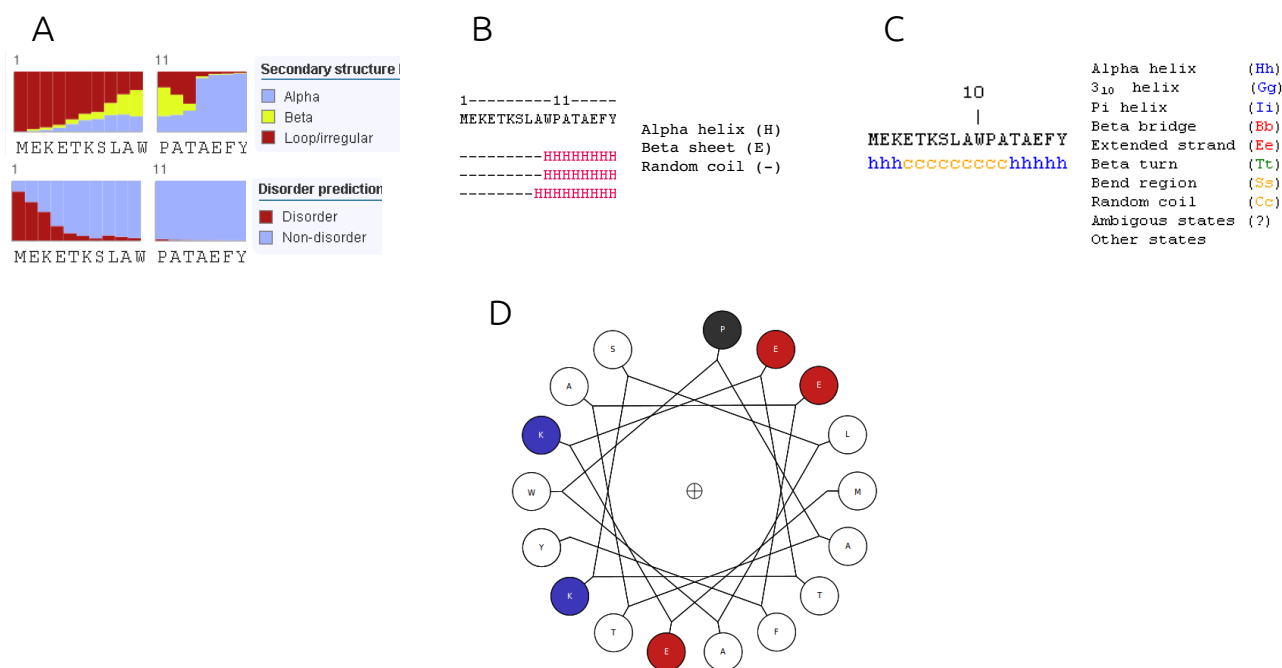


Figure 5.3 Secondary structure predictions of amino-terminal domain of U51A. (A) RaptorX web server [Kallberg et al., 2012] (B) JPred web server [Cole et al., 2008] (C) SOPMA web server [Geourjon & Deleage, 1995] (D) Helical wheel plot, red colouring indicates acidic residues and blue colouring indicates basic residues.

Docking was performed with CCL2 since this ligand of U51 has crystal structures resolved in a number of forms, in addition to extensive functional characterisation of interaction sites of the chemokine [Handel & Domaille, 1996; Lubkowski et al., 1997; Chakravarty et al., 1998; Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999; Lau et al., 2004]. These structures showed that like the majority of other chemokines CCL2 can form higher order oligomers. However, the obligate dimeric forms of CCL2 were unable to bind or activate CCR2 [Tan et al., 2012], implying the monomeric form is required for receptor interaction. Yet obligate monomeric mutants of CCL2 lacked the ability to recruit cells in an *in vivo* intraperitoneal recruitment assay [Proudfoot et al., 2003]. This Suggested that oligomerisation is essential for an aspect of chemokine function distinct from direct receptor binding. Together this data indicates that the dimerisation interface was unlikely to be involved in initial interactions with the N-terminus of chemokine receptors and therefore the modelling considered only the monomeric receptor binding interactions.

A number of the surface features of CCL2 identified from the structural data have also been investigated via mutational studies to characterise the effect of specific regions in its known functions: dimerisation, GAG binding and receptor interactions [Chakravarty et al., 1998; Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999; Lau et al., 2004]. With regards to the receptor binding determinants, Arg24, Lys49, Tyr13 and the ligands N-terminus have all been implicated to play a major role in interactions with the receptor. Of these Tyr13 and N-terminal truncations/mutants are predominantly linked to loss of receptor signalling or chemotactic activity while still retaining receptor binding abilities [Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999], implying they are involved more in the secondary interactions with the extracellular loops of the chemokine receptor that determine receptor activation in the prevailing 'two-site' model. In contrast, Arg24 and Lys49 are strongly linked to loss or reduction in receptor binding affinity

[Paavola et al., 1998; Jarnagin et al., 1999; Hemmerich et al., 1999], implying they are involved in the primary interactions with the amino-terminal domain of the chemokine receptor that are believed to determine the specificity of the interaction. Utilising this data allowed the search space used for the docking of the U51A N-terminal peptide to CCL2 to be narrowed down, concentrating around a hydrophobic groove running along the length of CCL2, figure 5.4A-B.

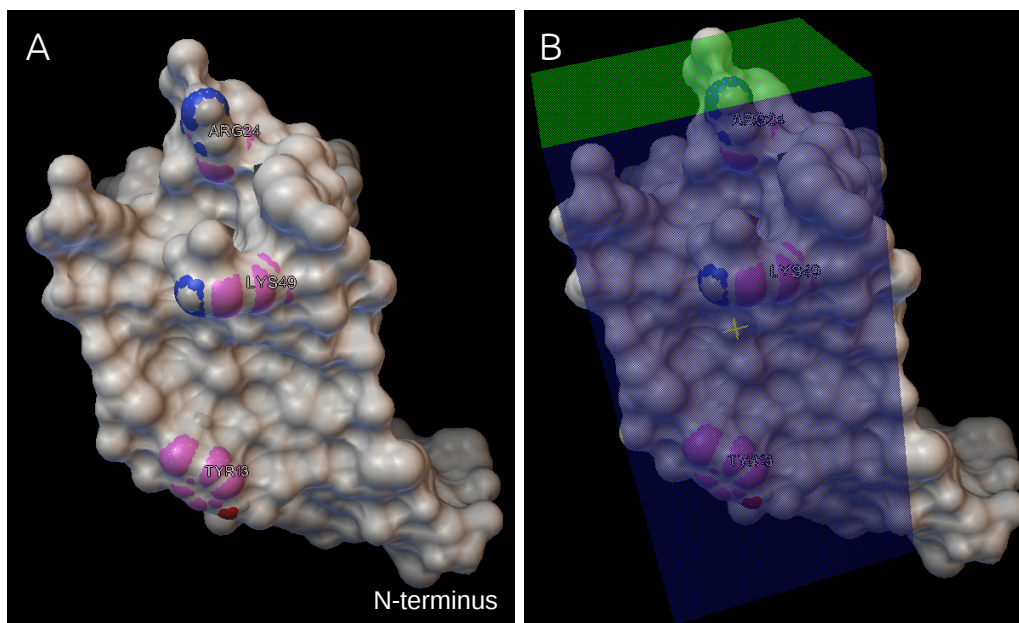


Figure 5.4 Search space parameters for ligand docking modelling. (A) Key interacting residues of CCL2 (B) Definition of search space parameters for amino-terminal domain docking.

Subsequently, docking of the helical viral receptor N-terminal peptides to the defined CCL2 search space were modelled using the AutoDock Vina software [Trott & Olson, 2010]; albeit with the actual roles reversed using the software such that the 'receptor' in the docking model was played by the crystal structure of CCL2 and the role of the 'ligand' by the viral receptor N-terminal domain peptides. The interaction was initially modelled with the wild type strain U1102 receptor N-terminal peptide strictly constrained to the alpha helical conformation by limiting the rotation of the bonds in the peptide backbone, figures 5.5A-B. The optimal predicted binding mode from this docking suggested that the viral receptor N-terminal domain would be orientated in a manner that would allow for further interactions of the N-terminal domain of CCL2 with the receptor binding pocket created by the extracellular loops. Additionally it gave some indication that the glutamic acid residue at position 2 of the viral receptors N-terminal domain had the potential to be an interacting partner with functionally important arginine residue at position 24 on the chemokine ligand. However, the predicted binding affinity was low, possibly from the structural constraints put on the rotation of the bonds. Additionally as the current evidence suggests a strict helical structure is unlikely to represent the natural form the amino-terminal domain of the receptor. The docking procedure was performed again, but this time with removal of the parameters constraining bond rotation in the peptide backbone, figure 5.5C-D. Again the optimal binding mode from this docking model predicted the receptor N-terminal peptide to be orientated in a manner which would be consistent with the 'two-site' model of chemokine receptor-ligand interaction, as well as

suggesting close association of the receptor's Glu2 residue and ligands Arg24 residue. This model had a much higher predicted binding affinity. Interestingly, this predicted binding model also suggested the glutamic acid residue at position 4 of the viral receptor peptide may also be interacting with Arg24 of the ligand, figure 5.5E. As well as placing the tryptophan residue at position 10 of the viral peptide in an orientation facing away from the interactions with the chemokine ligand. In structural studies of other chemokine receptors, most notably CXCR1, a tryptophan residue in this orientation found approximately half way along the length of the N-terminal domain has been implicated in anchoring the domain to the cell membrane, which would be accommodated in the U51 model [Szpakowska et al., 2012 & Park et al., 2011].

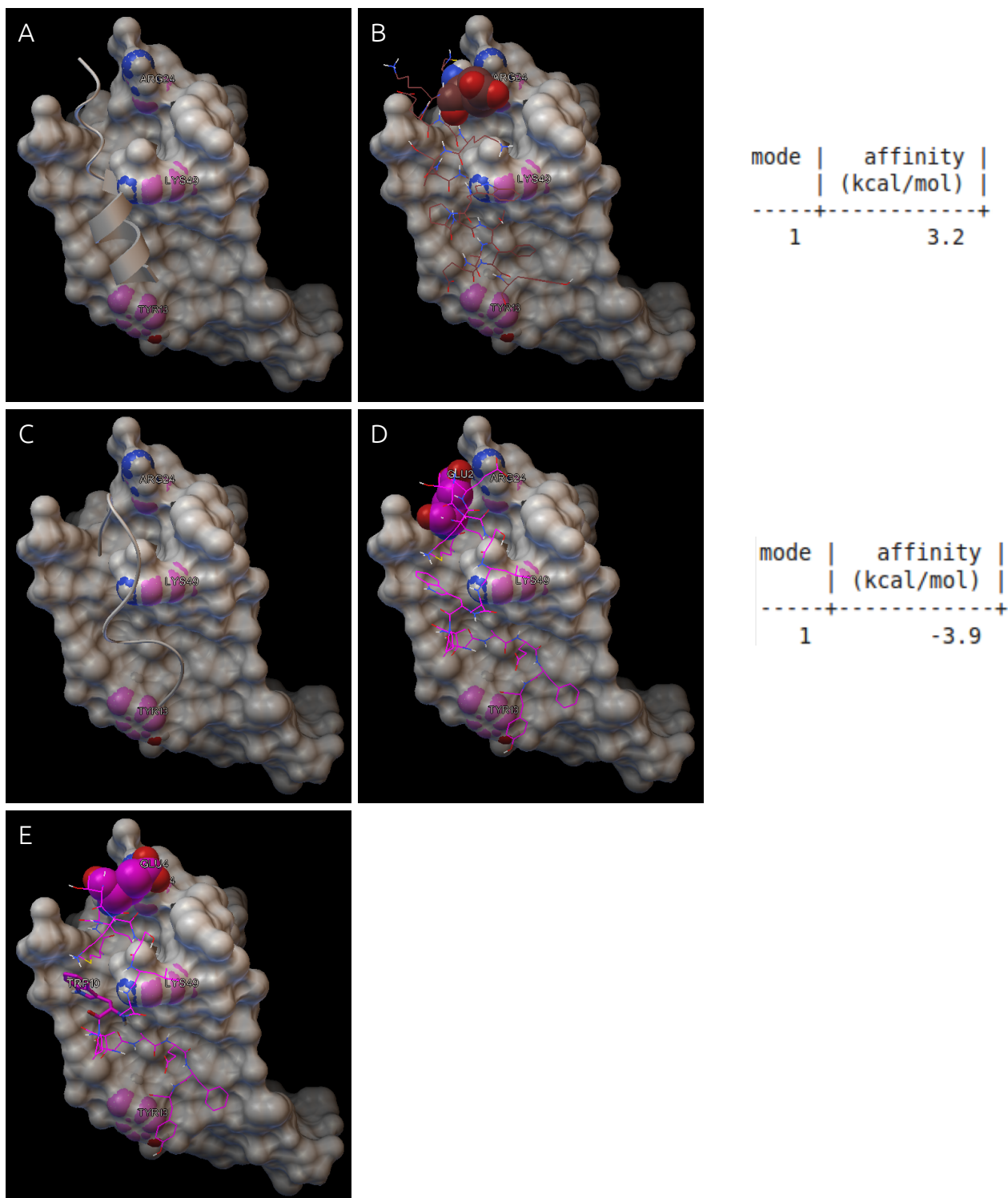


Figure 5.5 Modelling and predicted affinity of the interaction between the amino-terminal domain of HHV-6A strain U1102 U51 and the crystal structure of the chemokine ligand CCL2. (A) Ribbon structure representation of the N-terminus of U51 constrained in an alpha helical structure docked to a space filling representation of CCL2. (B) Molecular structure representation of the N-terminus of U51 constrained in an alpha helical structure docked to a space filling representation of CCL2. (C) Ribbon structure representation of the N-terminus of U51 with the torsional constraints on the peptide backbone removed, docked to a space filling representation of CCL2. (D) Molecular structure representation of the N-terminus of U51 with the torsional constraints on the peptide backbone removed, docked to a space filling representation of CCL2. (E) As (D) but also indicating the position of the tryptophan residue at position 10.

Next, the model was used to predict effects of the substitutions observed in the amino-terminal domain of the CI-HHV-6A samples. To do this docking was then performed with a viral peptide containing the glutamic acid to glycine mutation at position 2 ($\Delta E2G$) encoded in all CI-HHV-6A U51 genes, figure 5.6A-B. While the overall orientation of the viral peptide was the same as that seen for the strain U1102 peptide, the $\Delta E2G$ change led to distinct conformational differences at the immediate 5' end of the peptide, and notably a predicted reduction in binding affinity, figure 5.5C-D vs 5.6A-B.

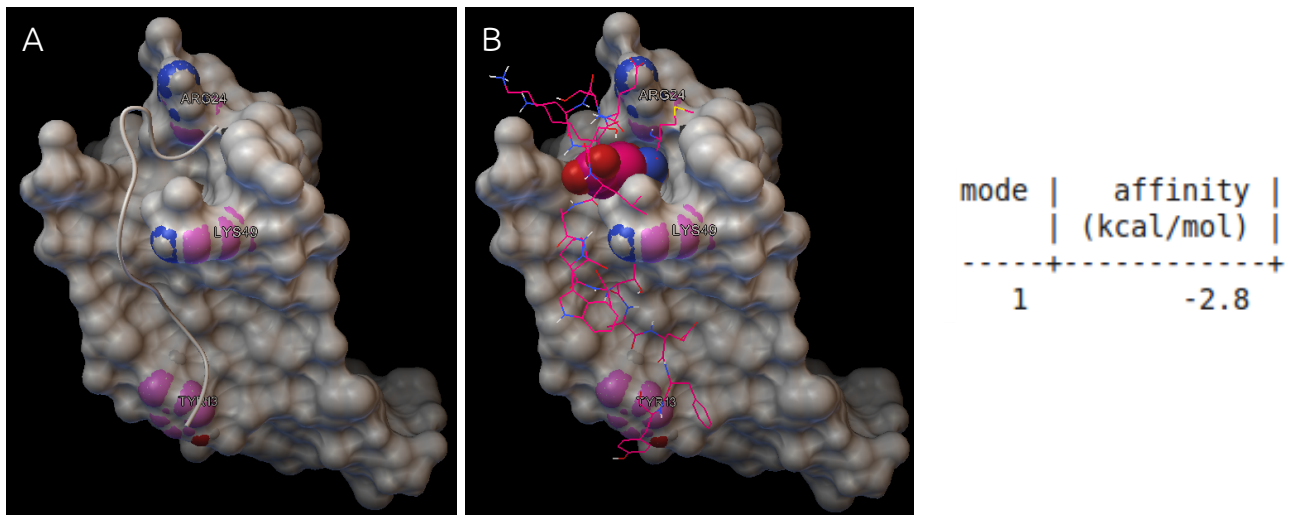


Figure 5.6 Modelling and predicted affinity of the interaction between the amino-terminal domain of the U51 from the CI-HHV-6A isolates and the crystal structure of the chemokine ligand CCL2. (A) Ribbon structure representation of the N-terminus of U51 docked to a space filling representation of CCL2. (B) Molecular structure representation of the N-terminus of U51 docked to a space filling representation of CCL2.

These models indicated that variation observed in this domain may alter the ligand binding dynamics of the receptor. Therefore the integrated U51 could have different function from that of some circulating viral strains. Notably, U51 from HHV-6A strain U1102 which has been utilised for all functional investigation to date. Based on these models, biological evidence for functional differences were next investigated.

5.4 Generation of Chimeric and Mutagenic Receptor Expression Vectors

To investigate the nature of the chemokine binding specificity derived from the amino-terminal domain of U51A, initially a chimeric chemokine receptor was generated. This swapped the amino-terminal domain of the human chemokine receptor CCR3 with the wild-type amino-terminal domain of U51A from the prototypical HHV-6A strain, U1102. This permitted greater investigation into the viral amino terminal domain in isolation from the rest of the receptor, while still being present in a naturally relevant environment of the cell membrane. This approach was based on chimeric human receptors similarly investigated for receptor specificity. These were human CCR1:CCR3 chimeric receptors used to demonstrate that specificity could be altered with changes in the amino-terminal domain [Pease et al., 1998]. Additionally, since U51 and CCR3 have overlapping but distinct chemokine ligand binding profiles, table 5.1, these could also be used to investigate the relationship between receptor binding specificity and functional activation. Central to this

was the interaction of U51A with the CCL2 ligand, distinct from CCR3 which does not recognise this chemokine ligand. This initial construct was generated by the gene splicing and mutagenesis by PCR-driven overlap extension protocol [Heckman & Pease, 2007] (see methods section 2.11.1 for a full description).

	CCL2	CCL5	CCL8	CCL11	CCL19	
U51A	+	+	-	+	+	Binding
	+	+	-	+	+	Chemotaxis
CCR3	-	+	+	+	-	Binding
	-	+	+	+	-	Chemotaxis

Table 5.1 Chemokine binding and chemotaxis properties of HHV-6A U51 and CCR3. Key ligands of U51 and CCR3 that have reciprocal specificities for investigating the interaction of binding and signalling.

To further investigate specificity of the amino-terminal domain of U51A, a number of other vector constructs were also generated, table 5.2. These utilised either a modified version of the gene splicing and mutagenesis by PCR-driven overlap extension protocol or mutagenic primers (methods sections 2.11.1 – 2.11.3). A vector was generated with the U51A amino-terminal cassette construct which could be utilised for the rapid addition of alternative receptor sequences to a HA tagged U51A amino terminal domain. As well as vectors based on an HA tagged wild-type U51A receptor, with alanine scanning mutagenesis of the entire amino-terminal domain. Allowing for further testing of the ligand specificity conferred by the amino-terminal domain of U51A.

U51NCCR3M	HA tagged receptor construct consisting of the human chemokine receptor CCR3 with the amino-terminal domain swapped for that of HHV-6A strain U1102 U51A. Allowing interogation of the amino-terminal domain of U51A in isolation from the remainder of the viral receptor.
HA-U51NCCR3	HA tagged receptor construct consisting of the complete human chemokine receptor CCR3 with the 5' addition of the amino-terminal domain of HHV-6A strain U1102. The nucleotide sequence at the 3' end of the U51 amino-terminal domain has been mutated to an EcoRI restriction site with maintenance of the amino acid sequence. Generated to act as a cassette for the rapid addition of alternative receptor sequences to the HA tagged amino-terminal domain of HHV-6A strain U1102 U51A.
HA-U51ΔE2G	HA tagged receptor construct consisting of HHV-6A strain U1102 U51A with a glutamic acid to glycine mutation at position 2. Reflecting the amino-terminal variation noted in U51A from CI-HHV-6A and certain HHV-6A strains.
HA-U51ΔM1A – HA-U51ΔY17A	A series of HA tagged receptor constructs consisting of the HHV-6A strain U1102 U51A with alanine scanning mutagenesis of the entire amino-terminal domain of U51A. For further interogation of specificity determinants found in the amino-terminal domain.

Table 5.2 Generated chimeric and mutant receptor constructs.

5.5 Surface Expression of U51NCCR3M Receptor Construct

Chemokine receptors must be trafficked to the cell surface for plasma membrane ligand interactions and subsequent receptor activation required to mediate signalling and chemotaxis. Induction of efficient chemotaxis requires sufficient receptor density on the cell surface [Szabo et al., 2001; Desmetz et al., 2006]. To detect surface expression of the generated receptors, all constructs were N-terminally tagged with an influenza haemagglutinin (HA) epitope. This enabled surface expression to be assayed by labelling with fluorescein isothiocyanate (FITC) labelled antibodies directed to the HA epitope and detection by flow cytometry. Expression was evaluated in two leukocyte cell lines, the human lymphoblastic K562 cell line and the murine pre-B lymphocyte L1.2 cell line, both previously used for surface expression of chemokine receptor constructs and chemotaxis assays [Catusse et al., 2008; Sabroe et al., 2005]. Therefore, a human leukocyte target of the HHV-6A virus was compared to a murine model system efficient for receptor expression.

The murine L1.2 cell line offered a significant advantage over the human K562 cell line, since it did not express competing human chemokine receptors, minimising the potential for interference. Additionally, high levels of surface expression following transfection of chemokine receptor constructs and efficient chemotaxis have been noted with this cell line [Vaidehi et al., 2009]. However, the use of this cell line with the closely related U51 receptor of HHV-7 failed to induce chemotaxis [Tadagaki et al., 2007] therefore it was important to consider the possibility that HHV-6 U51 may display limited chemotaxis in this cell line due to the requirement for host specific factors and look into alternative human cell lines. The human erythroleukemia K562 cell line was chosen as it does not express the chemokine receptors studied here under normal conditions, yet being of human origin the cells are potentially similar to the normal target cells of the virus with regards to intracellular factors required for receptor signalling. In addition a HA-tagged U51 construct has previously been shown to surface expressed and be able to induce chemotaxis in this cell line [Catusse et al., 2008].

Comparisons were made between different methods to optimise surface expression, appendix 7.8. As a result cells were subsequently electroporated when cultured cell density was $\sim 0.9 \times 10^6$ cells/ml, to ensure they were in early log phase since this optimised receptor gene expression. Electroporation was conducted with 2 μg of plasmid DNA/ 1×10^6 cells, then cells were cultured for a further 48hrs before downstream analysis, to permit maximum gene expression and cell surface transport. Electroporation of the U51NCCR3M chimeric receptor construct showed detectable but low surface expression in both cell lines, figure 5.7. Similar results were observed for transfection of the CCR3 and U51D constructs which have previously been utilised for functional analysis where similar transfection conditions had been utilised [Catusse et al., 2008; Sabroe et al., 2005].

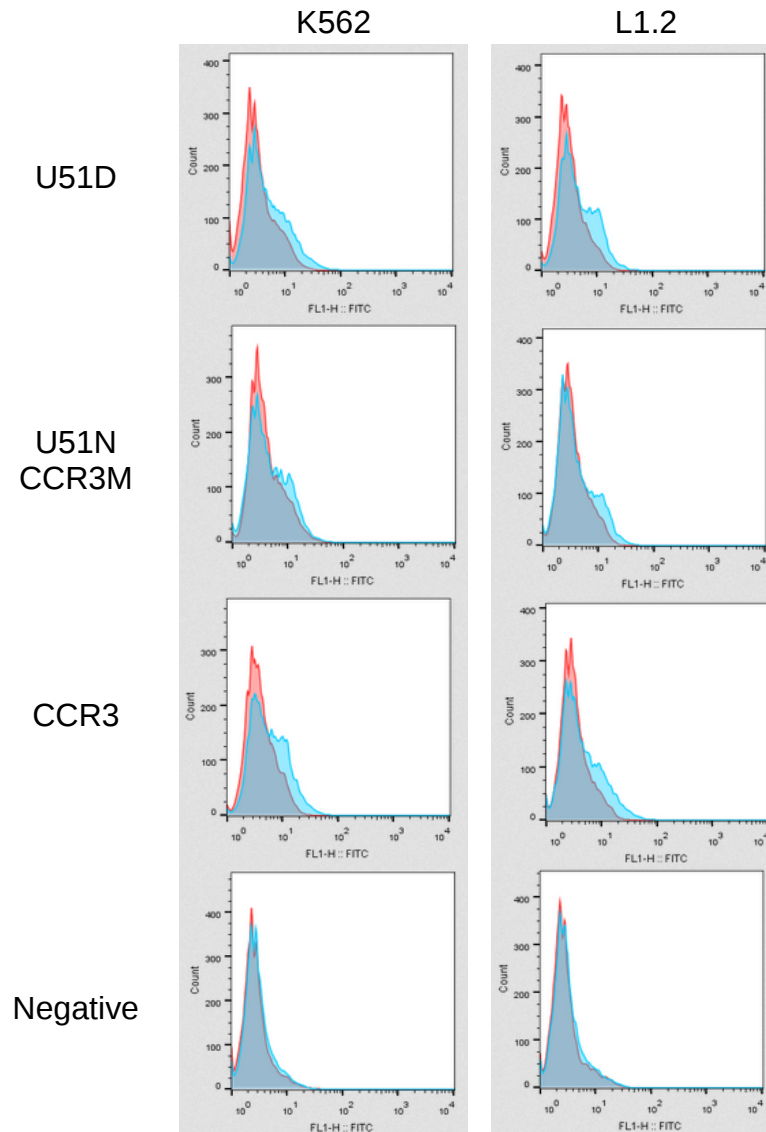


Figure 5.7 Surface expression of U51NCCR3M chemokine receptor construct in L1.2 cells. Histogram plots showing cell counts (y) against relative FITC fluorescence (x). Red shading indicates mock transfected cells and blue shading indicates experimental transfected cells.

5.6 Functional Analysis of U51NCCR3M Construct

Although surface expression of the receptor constructs was low, since chemotaxis has previously been observed at these expression levels, functional analysis by chemotaxis was investigated. This requires the specific transmigration of live cells across the chemotaxis filter in response to a chemotactic stimuli. Thus even if receptor expressing cells were in a minority of the total cell population, detection could be possible since only the receptor expressing cells could specifically migrate. However, this has to be measured against a background of random movements, chemokinesis and gravity flow effects. Experiments were performed in triplicate and preliminary results showed CCL2, a native ligand of the U51 receptor but not the CCR3 receptor, induced chemotaxis in cells expressing the chimeric receptor construct, figure 5.8. However, the chemotaxis levels were low and greater expression relative to background migration would be required to verify this result. This result indicated a change in both specificity and signalling of the CCR3 receptor

through the amino-terminal domain switch with the viral chemokine receptor.

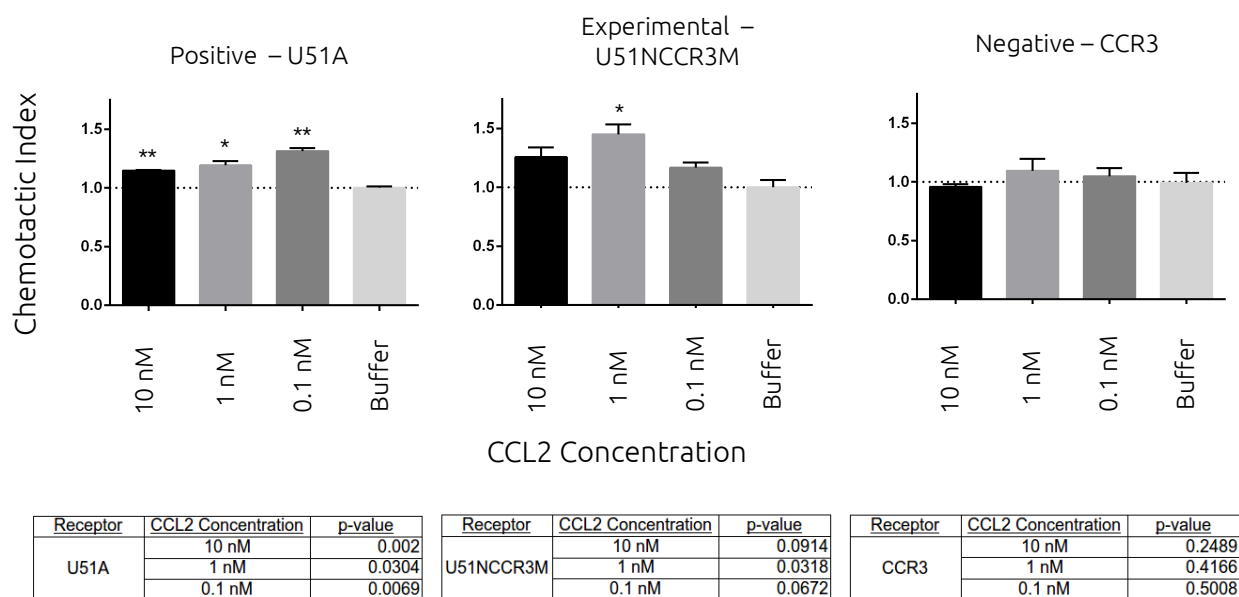


Figure 5.8 Chemotaxis of U51NCCR3M cells in response to the chemokine ligand CCL2. All ligand concentrations run in triplicate. Chemotactic responses are normalised as a chemotactic index, indicating the fraction of cells moving towards the chemokine ligand respective to buffer. One way analysis of variance (ANOVA) was used to determine significant chemotactic response (* = p value < 0.01 ** = p value < 0.05).

5.7 Discussion

The screening here of a key viral immunomodulatory gene from one of the largest groups of CI-HHV-6A and CI-HHV-6B samples collated to date shows that the integrated forms of these viral species have similar prevalences. This was supported by studies by others in this laboratory on three further loci, U38, U46 and U83 and subsequently published [Tweedy et al., 2015a]. As well as previous observations about the prevalence of CI-HHV-6A and -6B [Leong et al., 2007; Hubacek et al., 2009; Kuhl et al., 2015]. This is in contrast to the prevalence of circulating strains of HHV-6A and HHV-6B in the majority of countries studied to date, where HHV-6B is more frequently detected [Hall et al., 2006; 2008]. The exception to this being that of Zambia where HHV-6A and HHV-6B are detected at similar levels [Bates et al., 2009]. To date CI-HHV-6A has not been detected in screens in this southern African population, although one case of CI-HHV-6B has been reported in a north African population [Faten et al., 2012; Tembo et al., 2015]. Thus clear differences exist between the relative prevalences of the circulating and integrated forms of these viruses. If integration occurred recently or frequently, the relative prevalence of the integrated forms could be expected to reflect that of the circulating viral forms. While a number of factors, such as human genetics or differences in germ-line integration efficiency between HHV-6A and HHV-6B, could account for the differences in relative prevalence. An additional explanation may lie in the shared origins of the integrated samples and supported by data in the previous chapter. The higher prevalence of CI-HHV-6A could result from HHV-6A integration events, which predate those of CI-HHV-6B, allowing a greater length of time for expansion in the population. In this regard, the nucleotide sequences of the U51A gene from the samples where 17p13.3 chromosomal integration

was identified, were all identical (Germany sample ID 1624/5055 and Czech sample IDs 10593-43670, grouping together in figure 5.1). The same was also found with analyses of three further loci from CI-HHV-6A genomes integrated at this chromosomal location [Tweedy et al., 2015a]. Providing further support to the notion of shared origins of HHV-6A integrated at this site.

The prevalence of CI-HHV-6A in the lymphoproliferative/inflammatory disease cohort from the Czech Republic was higher than that observed in the German myocarditis/cardiomyopathy group. Suggesting the possibility of an association with this disease group. Again this was confirmed by studies in three further loci by others in the laboratory as included in our subsequent publication, Tweedy et al., 2015a. If such an association exists, key players in any viral contribution to disease could be from the virally encoded modulators of the chemokine system; which possess binding capabilities implying the ability to manipulate the majority of beta-chemokine inflammatory system [Dewin et al., 2006; Catusse et al., 2007; 2009; Isegawa et al., 1998; Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008]. Indeed, a recent study in the same German cohort examined here, from which gene expression from viral inflammatory mediators including U51 has been demonstrated [Tweedy et al., 2015a], found links between persistent inflammatory disease, in the form of myocardial complications, and CI-HHV-6A/-6B [Kuhl et al., 2015]. Furthermore, a recent large scale analysis has also found CI-HHV-6A/-6B to be a risk factor in the development of angina, with the prevalence more than three times greater in CI-HHV-6A/-6B than control groups [Gravel et al., 2015]. The implication being that expression from the integrated viral genome and/or reactivation leads to immune activation with cellular damage that promotes myocardial complications.

Other work in this laboratory has found that virally encoded chemokine, U83, is predominantly in a long active form, in CI-HHV-6A samples from this inflammatory disease cohort [Tweedy et al., 2015a]. Furthermore cDNA analysis indicated gene expression of this potent viral chemokine [Dewin et al., 2006; Catusse et al., 2007; Clark et al., 2013; Tweedy et al., 2015a]. Gene expression was also suggested from the chemokine receptor gene, U51, of these CI-HHV-6A samples [Tweedy et al., 2015a]. Here it was found that there is variation in the U51A gene in both HHV-6A and CI-HHV-6A, with amino acid sequence comparisons revealing four sites of variation. Of note, two of these represented charge changes in predicted extracellular domains, thus having a strong potential to affect viral receptor ligand specificity and receptor activation [Blanpain et al., 1999; Szpakowska et al., 2012]. As such it seems important to investigate possible functional significance, since the gene could be expressed in every cell and dramatically affect the inflammatory response giving dysfunctional cell mobilisation.

The amino-terminal domain was investigated further here, since this domain has a noted role in ligand specificity interactions in other chemokine receptors [Pease et al., 1998; Monteclaro & Charo, 1997; Blanpain et al., 1999; Wu et al., 1996; Gozansky et al., 2005; Mizoue et al., 1999]. Two chimeric receptor construct expression vectors have been generated, along with a selection of vectors containing constructs with alanine scanning mutations across the amino-terminal domain of the viral receptor. Preliminary attempts at functional analysis with the U51NCCR3M viral-human chimeric receptor construct have demonstrated transient surface expression, albeit at a limited level. Initial results showed that this amino-terminal domain switch may allow changes in both receptor specificity and activation. Future studies are needed to confirm this result and test the other constructs made (table 5.2), but there was not sufficient time to explore this further here. Of

interest, amino-terminal chimeric receptor constructs have previously shown to display specificity changes, but this did not correspond to chemotaxis activation [Monteclaro & Charo, 1996; Pease et al., 1998; Blanpain et al., 2003]. Therefore, these preliminary results indicate unusual differences in the viral receptor and mechanisms of chemokine receptor activation. Which may have implications for ligand binding specificity differences between the integrated and virus encoded forms of U51.

Higher levels of receptor surface expression would facilitate further analyses and many factors may contribute to this. It may be that amino-terminal changes in the current constructs affect cell surface transport of the receptor constructs or inhibit binding of the fluorescently labelled antibody utilised for surface expression analysis. However, chimeric CCR3 receptors have previously been expressed at levels sufficient for chemotactic analysis [Pease et al., 1998]. Amino-terminally HA tagged U51 receptor constructs have also been expressed in the K562 cell line at levels sufficient for chemotactic analysis [Milne et al., 2000]. Additionally, expression of the wild type CCR3 receptor construct in the L1.2 cell line is around half of that previously achieved and utilised for chemotactic analysis [Duchesnes et al., 2006]. Thus it seems an alternative explanation to mutation affecting the surface expression exists.

As highlighted, a number of factors affecting surface expression have been investigated (appendix 7.8), albeit with the human CCR5 chemokine receptor construct, which were applied to expression of the U51 and CCR3 constructs here with limited success. This could be due to different effects on the chimeric receptor. As an alternative means of achieving sufficient levels of surface expression, a stably expressing cell line was generated with the U51NCCR3M construct, described in appendix 7.9. This method has previously been utilised for functional analysis of U51 [Catusse et al., 2008]. While expression levels are typically lower in these cell lines, the advantage is that the proportion of expressing cells is greatly increased. There was not sufficient time to further investigate surface expression and functional analysis of these cell lines but this could be a subject for further study. Additionally, inconsistencies in the chemotaxis assay have led to considerations for alternative means of functional analysis of the receptor constructs. Exploratory investigations have been made with an actin polymerisation assay which would be a downstream marker for receptor activation [Burger et al., 2005; Udi et al., 2013], methods described in appendix 7.10. The actin polymerisation assay therefore gives an alternative means of assessing receptor activation, while ligand binding assays using the stable cell lines could give greater insight into ligand specificity and affinity changes.

In summary, a key site of amino acid variation was defined in an important immunomodulatory gene, U51A, encoded by the chromosomally integrated form of HHV-6A. This was investigated further by molecular modelling and functional analyses of a series of receptor construct expression vectors. Preliminary results showed altered roles of the amino-terminal domain of U51A in chemokine ligand binding specificity and receptor activation. With potential for changed affinity and response to chemokine CCL2 therefore restricting its activity. Further studies could confirm these functional changes and relation to inflammatory pathology.

Chapter 6: Discussion

6.1 Introduction

At the start of this thesis there was little understanding of the genetic relationship between the chromosomally integrated genomes of CI-HHV-6A/-6B and those of infectious HHV-6A/-6B. Evidence from *in vitro* cell line work has suggested that viral integration could represent a new mechanism of latency, which would be unique amongst the human herpesviruses [Arbuckle et al., 2010; 2013]. Although to date, *in vivo* evidence for this has not been presented. However, irrespective of the potential role as a latency mechanism, it is clear that both HHV-6A and -6B can and have integrated into the human germ-line. This has resulted in endogenous forms of these viruses, in approximately 1% of populations studied [Tanaka-Taya et al., 2004; Leong et al., 2007; Hubacek et al., 2009; Potenza et al., 2009; Pellett et al., 2012]. The pathological significance of these integrated forms are only beginning to be determined. Therefore the major intention of this thesis was to explore the integrated forms of these viral species to establish their relationship to circulating viral strains and begin to establish their biological significance.

The results presented here firstly established methodology which can be used to allow accurate genomic NGS directly from tissue samples. Validation of these methods was provided through resequencing of the prototypical HHV-6A strain, U1102. As well as by adding to the repertoire of fully sequenced HHV-6A strains through derivation of the complete genome sequence of strain AJ. This established a NGS pipeline which could be applied to characterisation of the germline chromosomally integrated forms of HHV-6A/-6B. Additional analyses of 12 viral integration site junctions indicated a shared genetic lineage of CI-HHV-6A integrated at chromosome 17p. Whole genome NGS of three CI-HHV-6A isolates revealed they are intact possessing the full complement of genes identified for HHV-6A strains, in addition to key cis acting signals required for replication/reactivation. In addition, comparisons of these integrated genomes to circulating viral strains highlighted divergence at a set of genes regarded as markers of speciation amongst herpesviruses. Minor variant analysis also supported a mechanism by which HHV-6A/-6B superinfection may result in gene expression from, and/or reactivation of the integrated virus. Finally, variation in the viral chemokine receptor, U51, observed in CI-HHV-6A was investigated due to a predicted role in ligand binding specificity affecting function.

6.2 Roseolovirus Next Generation Sequencing

Next generation sequencing technologies have been extensively applied for genomic analyses of the closely related betaherpesvirus HCMV [Reviewed in Sijmons et al., 2014]. However, at the start of this thesis these technologies had yet to be applied to the Roseoloviruses. Therefore, one of the initial aims of the work presented here was to investigate genomic NGS and subsequent assembly of HHV-6A/-6B strains, using the currently most prevalent commercially available platform. This was with a focus on establishing methods which would allow characterisation of the chromosomally integrated forms of these viral species, directly from clinical samples.

The work presented here suggests that both solution hybrid capture and long range PCR (LPCR) represent viable options for target enrichment of HHV-6A genomic DNA prior to Illumina MiSeq mediated sequencing. Both methods were capable of achieving sufficient

read coverage across the genome for whole exome sequencing. Indeed, the percentage of the genome which could be resolved, and generated consensus sequences, were almost identical for both enrichment methods. Thus, from a technical standpoint there is little to distinguish between these enrichment methods for the NGS sequencing of HHV-6A strains for this application. Instead, the choice of method will likely rely on other factors such as time, cost and sample limitations. In this regard, the overall time for the completion of both target enrichment methods is similar, ~48-72 hours. Although preliminary optimisation of amplicon primer and RNA bait library sequences could vary significantly. However, even with the possibility of multiplex PCR, the generation and purification of the genomic amplicons during this time period represent a greater work load than the hybridisation and capture steps required for solution hybrid capture based target enrichment. Yet LPCR based enrichment has the advantage of reagent cost at the scale performed here, approximately 50% cheaper than that of solution hybrid capture, although this could change with competition for commercial target enrichment strategies. LPCR based enrichment could in some regards be a more readily reproducible method since it does not rely on custom designed RNA bait libraries and theoretically can capture greater divergence. Moreover LPCR also provides an archived reference library of amplicons unlike solution hybridisation which depletes the original template. However, variability in the read coverage coinciding with the amplicon locations was an issue with LPCR enrichment, especially with the CI-HHV-6A samples. An issue where solution hybrid selection provided a clear advantage over LPCR enrichment. This variability not only had implications for resolution of the CI-HHV-6A genome sequences, but could also affect the detection of the signatures of genomic insertions/deletions or the use of software utilising read coverage to aid in the determination of repetitive regions [Xi et al., 2011; Nowak, 2015]. While amplicon primer optimisation for the CI-HHV-6A genomes may be beneficial there was no indication from the post amplification size and yield quantifications that this variability arose directly from PCR complications. Instead other factors most notably extended storage of the amplicons combined with minor inaccuracies in the quantification process seemed more likely to be the cause. Regardless, the experience here suggests focus should be placed on the crucial equimolar pooling step to ensure less variability thus aiding accurate genome sequencing.

Comparisons of different assembly methods, indicated that mapping of the read data to an appropriate HHV-6A reference genome with subsequent variant calling and consensus sequence generation was sufficient for accurate sequencing of the UL region of the genome. The most notable advantage of this being a reduction in the time and computational resources required for sequencing. However, where possible both a mapped and *de novo* assembly should be utilised as separate lines of evidence to inform the generation of the hypothesis for the underlying genome sequence. Problems were encountered with the assembly of reads across repetitive regions, which for HHV-6A/6B are in the R1-R3 and DR regions of the genome. The results here suggest that these regions cannot be accurately resolved by a mapped assembly resulting from read alignment failure at these more diverse regions or mis-alignment of reads complicated by repetitive regions. Targetted *de novo* assembly of these regions permitted incomplete sequencing. However, subsequent Sanger PCRs performed for gap filling which overlapped with the contigs generated for these regions, indicated that at least for the overlapping regions the *de novo* assembly was generating an accurate representation of the sequence. Thus, while *de novo* assemblies were able to partially sequence these regions, the need for Sanger sequencing for gap filling in repetitive sequence assessment represent a major bottleneck for the routine use of NGS for HHV-6A whole genome

sequencing. Currently the maximum read length achievable on the Illumina MiSeq platform is 2x300 bp, yet tandem repeat regions in HHV-6A genomes can extend to more than double this size often also surrounded also by imperfect repeats. Therefore, if read lengths are extended on the MiSeq platform this may allow a higher throughput by NGS for HHV-6A genome sequences..

The issue of repetitive sequences is also complicated by the presence of human telomeric-like repeat sequences within the genomic termini, thus the potential exists for sequenced reads of human origin to confound results. These repetitive areas of the genome are some of the major areas which define the sequence length variation observed between different HHV-6A species. To date the role of the R1-R3 repetitive sequences in the genome are largely undefined, however, their retention in all currently sequenced HHV-6A and -6B strains implies a functionally significant role. The exception to this is the human telomeric-like repeat sequences located in the DR at the genomic termini. These sequences have been proposed to play dual roles in HHV-6A replication both in integration of the viral genome into host chromosomal DNA and protection of the integrated viral DNA through neo-telomere formation following chromosomal integration [Arbuckle et al., 2010; Huang et al., 2014; Ohye et al., 2014].

6.3 Resequencing of HHV-6A Strain U1102

The establishment of a methods pipeline for Illumina MiSeq mediated HHV-6A genomic sequencing ultimately led to resequencing of the prototypical HHV-6A strain, U1102, which showed 99.9% identity with the original sequence. In addition to the three previously identified sequencing errors/sites of variation, in the CDS of the DR6, U83 and U86 genes [Schleimann et al., 2014; French et al., 1999; Dewin et al., 2006; Papanikolaou et al., 2002], confirmed by the resequencing here. Five further suggested corrections to the archived genome sequence were identified by both assembly methods: U40, U42, U57, U58, U100. These could affect the functions of these encoded proteins in viral replication. Little experimental information directly from HHV-6A/-6B strains is currently available on the majority of these genes. However, homologues of these HHV-6A genes are identified in other herpesviruses and indicate potential functions.

HHV-6A/-6B U40 is a homologue of HCMV UL56 and HSV-1 UL28 [Jones & Teo, 1993]. In both HCMV and HSV-1 the products of these genes form trimers, UL51-UL56-UL89 for HCMV [Borst et al., 2013] and UL15-UL28-UL33 for HSV-1 [Heming et al., 2014], collectively termed the terminase complex. The UL56/UL28 components possess a nuclear translocation signal and the ability to interact with a cellular importin [Giesen et al., 2000a] to permit transport of the trimeric complex to the nucleus whereupon the complex associates with viral DNA replication centres [Giesen et al., 2000b]. Additional sequences within the UL56/ UL28 component orchestrate binding to the highly conserved herpesviral pac sites, nuclease activity, ATP binding and interaction with viral capsid structures [Bogner et al., 1998; Tengelsen et al., 1993; Sheaffer et al., 2001; White et al., 2003; Scholz et al., 2003]. This is consistent with these complexes acting as molecular motors that cleave single unit length viral genomic DNA from its concatemeric form and package these into preformed capsid structures via the capsid portal. By inferred homology, HHV-6A and -6B U40 may interact with the products of U35 and U60 to form the terminase complex.

HHV-6A and -6B U42 is a homologue of HCMV UL69 and HSV-1 UL54/ICP27 [Winkler et al., 1994; Gompels et al., 1995], and in turn the ICP27 family of proteins due to the high level

of conservation within the *Herpesviridae*. These are expressed with immediate early/early kinetics [Winkler et al., 1994; Winkler & Stamminger, 1996; Sandri-Goldin, 2011], with some suggestion UL69 may also be present in the tegument of HCMV virions [Winkler & Stamminger, 1996]. ICP27 shuttles between the cytoplasm and nucleus [Soliman et al., 1997] and directly binds viral mRNAs [Corbin-Lickfett et al., 2009]. It acts as a transactivator of viral gene expression from the later kinetic classes, with most effect being mediated post-transcriptionally [Winkler et al., 1994; Rice & Knipe, 1990; Uprichard & Knipe, 1996]. These effects are mediated by a variety of different means, the best characterised being inhibition of cellular pre-mRNA splicing, the nuclear export of viral mRNAs through interactions with cellular mRNA export factors and stabilisation of viral mRNAs in the cytoplasmic compartment [Sandri-Goldin, 2011].

HHV-6A and -6B U57 is a homologue of HCMV UL86 and HSV-1 UL19/VP5, which has been defined as the major capsid protein in HSV-1 [Olshevsky & Becker, 1970; Brown & Newcomb, 2011]. Concurrent with this role as a virion structural protein, the gene is expressed in abundance with late kinetics in herpesviruses and as such is frequently used as a marker for the detection of viral replication.

HHV-6A and -6B U58 has homologues in the betaherpesviruses, HCMV UL87, and the gammaherpesviruses, EBV BcRF1, however, homologues do not appear to be present in the alphaherpesviruses. Deletion mutants of both UL87 and BcRF1 suggest these genes are expressed with early kinetics, and essential for growth due to their apparent involvement in a mechanism required for activation of late viral gene expression, which in EBV appears to be as a member of a viral 'preinitiation' complex [Yu et al., 2003; Dunn et al., 2003; Isomura et al., 2011; Aubry et al., 2014]. Additionally, recent evidence suggests an antisense transcript from the U87 gene region may be expressed during the late phase of replication, possessing two predicted reading frames, one of which shows high conservation among CMV strains [Ma et al., 2011]. However, currently the products or functions of these predicted ORFs functions have not been identified.

While HHV-6A and -6B U100 is unique to the Roseoloviruses. It encodes the envelope glycoprotein Q (gQ), two transcripts produce different products (gQ1 and gQ2) from the gene via differential splicing. These products associate with glycoproteins H and L (gH and gL) form complex which acts as the viral ligand for the cellular receptor CD46 and along with glycoprotein B (gB) mediate membrane fusion [Mori et al., 2003a; 2003b; Akkapaiboon et al., 2004; Tanaka et al., 2013].

Preliminary alignments of the HHV-6A strain U1102 CDS with HCMV homologues indicated that the sequence variations fell outside of any of the currently defined functional domains. However, the crucial role of many of these gene products in herpesviral replication, may mean that investigation of these components in HHV-6A and -6B may be of interest to produce intervention strategies. In this regard, a number of small organic molecules have been found to be interfere with HCMV replication via inhibition of the terminase complex [Krosky et al., 1998; Buerger et al., 2001; Goldner et al., 2011]. The existence of drugs which abrogate HSV-1 replication, through an indirect action affecting ICP27 function [Murata et al., 2001; Park et al., 2013], implies that the members of this gene family could represent viral intervention targets. RNA interference (RNAi) of the capsid proteins UL18 and UL19 in HSV-1 and -2 has been found to significantly affect the replication of acyclovir resistant clinical isolates of these viral species. Suggesting these could be a target for RNAi based intervention strategies, especially in the case of strains

resistant to the most prevalent herpes antiviral drugs [Jin et al., 2014]. Studies with mutant BACs suggest gQ is essential for virus production [Tang et al., 2011]. Also in addition to gB and gH, gQ also contains epitopes which are recognised by neutralising antibodies [Pfeiffer et al., 1993; 1995], which in combination with significant sequence variation between the HHV-6 species at these genes raises the possibility of an intervention via variant specific neutralising antibodies. Thus these corrections or sites of variation may be of importance for such work.

6.4 Sequencing of HHV-6A Strain AJ

The established Illumina MiSeq NGS methods were further validated by genomic sequencing of HHV-6A strain AJ, adding a third viral strain to the repertoire of fully sequenced HHV-6A strains. The recent sequencing and analysis of an isolate of HHV-6A strain GS [Gravel et al., 2013a] led the authors to propose that the observed divergence between strain GS and the prototypical HHV-6A strain U1102 [Gompels et al., 1995] may stem from their distinct geographic origins. Strain GS was first isolated from patients with lymphoproliferative disorders, including one with HIV/AIDS, from the United States/Jamaica [Salahuddin et al., 1986] and strain U1102, while originally isolated in the United Kingdom, was from a Ugandan HIV/AIDS patient [Downing et al., 1987]. However, the high degree of relation observed here between strain AJ, originally isolated from an adult HIV/AIDS patient from Gambia [Tedder et al., 1987], and strain GS from the USA imply alternative explanations must exist for the variation observed between strains U1102, GS and AJ. Indeed, despite their distinct geographic origins, strains AJ and GS showed closer relations to one another than either did to strain U1102. This may reflect convergent evolution of the strains or alternatively a more recent emergence event, possibly related to chromosomal integration. Current evidence cannot provide a definitive answer to this question. However, the presence of a number of features in the HHV-6A genome may favour the latter of these explanations. For example, in contrast to the product of HHV-6B IE1/U90, which efficiently evades type I interferon responses, HHV-6A remains sensitive resulting from the lack of a specific gene insertion [Jaworska et al., 2007; 2010]. U20A also displays signs of unregulated immune surveillance since it downregulates MHC class I more efficiently than HHV-6B U20 [Glosson & Hudson, 2007]. Additionally, while U54B inhibits nuclear factor of activated T-cells (NFAT) signalling, through improper dephosphorylation and nuclear translocation, U54A actually acts as a transactivator of NFAT [Iamietro et al., 2014].

In depth sequence analyses of AJ also noted variation in the DNA polymerase gene, U38. Due to its high level of conservation, regions of U38 are often used for HHV-6 species typing. However, this variation noted in strain AJ, but also present in strain GS, fell within primer site locations used in a real-time PCR based diagnostic assay for the HHV-6 species [Lou et al., 2011]. This assay has been readily utilised [Lou et al., 2011; Oakes et al., 2013; Tembo et al., 2015] and could confound results through under-detection of HHV-6A species, through reduced primer affinity. Such issues can easily be overcome through the use of multiple genomic sites for diagnosis in the currently established real-time PCR based assays. A number of newer technologies such as ddPCR, which has already been employed for the detection of HHV-6A and -6B [Sedlak et al., 2014; Leibovitch et al., 2014], and ultra-deep sequencing with variant analysis [Barzon et al., 2013; Beerenwinkel et al., 2011] display promise in the ease of quantification of infections as well as the detection of multiple or co-infections. However, these may still require a PCR amplification based approach, such that the identification of multiple genomic targets from all available

sequence available remains a crucial consideration.

During annotation of HHV-6A strain AJ, ORF prediction software was utilised to confirm the location of previously defined genes. These indicated a number of short ORFs (<100 amino acids) may be present in the AJ, however, due to the lack of evidence indicating any transcription and/or translation from these sites these were not been included in the current annotation. But short reading frames are increasingly found to be a feature of many genomes [Andrews & Rothnagel, 2014]. Indeed, two recent reports utilising ribosome profiling in the gammaherpesvirus KSHV and the betaherpesvirus HCMV have revealed a large number of new ORFs [Arias et al., 2014; Stern-Ginossar et al., 2012]. These were from both novel genomic loci and overlapping currently annotated ORFs, including numerous examples of leaky scanning and non-AUG initiation. While the cellular function of many of these remain to be determined, this implies a currently underappreciated level of complexity to expression from herpesviral genomes. This is in addition to recognition of the increasing role virally encoded microRNAs and long non-coding RNA play in herpesviral replication [Hook et al., 2014; Tuddenham et al., 2012; Nukui et al., 2015]. Thus currently defined annotations will likely require extensive updates as the status of many of these features become further defined.

6.5 Germ-line Chromosomal Integration of HHV-6A and -6B

Chromosome 17p Integration

To date, germ-line CI-HHV-6A/B genomes have been identified at 10 different chromosomal locations. This indicates there has been a minimum of two separate integration events into the germ-line for HHV-6 species, one for HHV-6A and another for HHV-6B. The presence of integrated viral genomes from these species at multiple chromosomal locations in the germ-line could potentially have arisen from additional independent germ-line integration events, interchromosomal translocation of integrated viral genomes, or excision and re-integration events within germ-line cells. Currently, further study is required to ascertain the relevance of these different scenarios. Vertical transmission of CI-HHV-6A and -6B is well established [Mori et al., 2009b; Arbuckle et al., 2010; Huang et al., 2014; Kuhl et al., 2015]. Indeed integrated genomes account for most cases of congenital infections by these viruses [Hall et al., 2010]. This vertical inheritance implies that there is some level of expansion of the integrated genomes within populations by Mendelian inheritance. However, mechanisms governing the frequency of integration at a given chromosomal location need to be established. For example, the integration could represent a singular event with extensive expansion in the population or frequent separate integration events at preferential integration location each with limited expansion.

Analysis of the integration site junctions of CI-HHV-6A amplified utilising a chromosome 17p13.3 specific primer sequence, suggested a shared genetic lineage between the integrated viral genomes at this chromosomal location, within the European population under investigation here. This was further supported by phylogenetic clustering of isolates integrated at this chromosomal location in the U51 gene. Indeed, additional work in this laboratory utilising multiple loci analysis of CI-HHV-6A isolates from this cohort with evidence for integration at chromosome 17p, showed similar phylogenetic clustering, as well as retention of a long form of the viral chemokine gene, seen rarely in clinical strains [Tweedy et al., 2015a]. This data implies germ-line integration of HHV-6A/-6B may be a

series of rare events which have subsequently been expanded as a polymorphism. In line with this, recent results analysing CI-HHV-6B in non-familial Japanese patients has also noted similarities between integration sites at the same chromosomal location, adding support to the notion of ancient origins of these integration events [Ohye et al., 2014].

Analyses of the CI-HHV-6A integration site junctions showed a shared structure of combined perfect and imperfect telomeric repeats. These shared characteristic repeat patterns close to readily identifiable sequence from both the chromosomal subtelomeric region and the viral DR_R. There was some variance in the number of repeats between these two signatures. However, two isolates from independent donors, 25533 and 27355, shared identical junctional sequences implying close familial relationship. The variance observed between other isolates suggested replication slippage, or homologous recombination has occurred, perhaps reflecting accumulative changes since the integration event or that viral integration has effects on telomere biology. In this regard, a recent report on the length of CI-HHV-6A/-6B associated telomeres in somatic and sperm cells suggests elongation of virally associated telomeres can occur in germ-line cells, implying active maintenance in these cells, although this is presumed to occur by a telomerase mediated mechanism. While virally associated telomeres were frequently the shortest in the somatic cell, the implication being that integration may disrupt telomere function [Huang et al., 2014].

The possibility of more than a single CI-HHV-6A integration event at this chromosomal location remains, however. CI-HHV-6B was also noted at chromosome 17 in this work, and in a number of previous reports [Luppi et al., 1993; Torelli et al., 1995; Morris et al., 1999; Ward et al., 2006; Nacheva et al., 2008]. So at least two independent integration events at this chromosomal location have occurred, one for each viral species. The previously identified 17p CI-HHV-6A junctional sequence of undefined, possibly US, origin (accession GU784872) shared similarities to those identified here but lacked the characteristic imperfect telomeric repeats identified in the CI-HHV-6A samples here. As suggested above, accumulative effects of replication slippage or homologous recombination could be a possible explanation for this. Another possibility is that *in vitro* passage of the leukocytes from that patient affected the integration site or its cloning into plasmids prior to sequencing. The integration sites investigated in this thesis were amplified directly from clinical samples. However, it is also possible that this could represent a separate lineage of CI-HHV-6A integrated in the germ-line at this chromosomal location. Indeed, all of the samples assessed here were from a clustered region in central Europe (Czech Republic and Germany). But similarities were also noted to a previously defined CI-HHV-6A integration site from Sardinia (accession KF366419). Therefore, it would be of interest to investigate additional populations, as well as the ethnicity of the individuals with CI-HHV-6A/-6B assessed here, to give further insight into the extent and origins of these integration events.

Furthermore, while the viral integration site junction sequences generated here appear to share common architectural features and were amplified utilising primer sequences designed to amplify viral genomes integrated into the subtelomeric region of chromosome 17p; the differences in the number of pure TTAGGG telomeric repeats between these common features could also be indicative of integration events in the subtelomeric regions of more than one chromosome. In this regard, it is well established that the subtelomeric regions of a number of human chromosomes share regions of significant homology [Riethman et al., 2004; Ambrosini et al., 2007]. In particular, the

distal, telomeric adjacent, regions of the subtelomeres from these different human chromosomes contain short repetitive sequences of high homology, which have arisen from segmental duplication and interchromosomal translocation events [Mefford & Trask, 2002; Mewborn et al., 2005; Linardopoulou et al., 2005]. It is also true that the subtelomeric and telomeric regions of all human chromosomes remain poorly defined when compared to more central regions of the genome, resulting largely from the repetitive nature and high levels of sequence homology, which have hampered their inclusion in genome sequencing projects [Lander et al., 2001; Riethman et al., 2004]. Such that numerous polymorphisms in the subtelomeric regions of the human chromosomes are likely yet to be defined, which in combination with segmental duplication and chromatid exchange could allow for similar sequences to be present in the subtelomeric regions of multiple chromosomes. Thus, there is the possibility that the 17p primer sequence utilised in this work could also amplify from regions of significant homology located on chromosomes other than at 17p13.3. For true clarification of the chromosomal location of these integrated genomes, additional techniques such as that of FISH analysis would be of merit. Indeed a defined protocol for the generation of FISH probes which permit the high sensitivity detection of CI-HHV-6A/B has recently been published [Ohye et al., 2016]. However, owing to the requirement for interphase or metaphase chromosome preparations for FISH analysis and the nature of the CI-HHV-6A/B samples available in this work, FISH analysis on the individuals included in this work would not be possible without the acquisition of further samples. Of note, there was also a distinct clustering of CI-HHV-6A located on chromosome 22q in seven of the thirteen Japanese patients examined in the work generating the CI-HHV-6A/B FISH probes [Ohye et al., 2016]. Which could support either the notion of a minority of integration events with expansion in the population, as argued for chromosome 17p in the European populations studied here or perhaps that chromosome 22q represents an additional preferential site of integration. However, sequences of the integration site junctions at chromosome 22q or genomic data from the CI-HHV-6A samples which could determine the relationship of these integrated viral genomes has yet to be determined.

As such, it does remain a possibility that the integration sites amplified here could originate from multiple integration events at the same chromosomal location or different homologous subtelomeric locations. However, a number of distinct features support the interpretation here of a shared genetic lineage between CI-HHV-6A genomes integrated at chromosome 17p in the populations studied here. The observation that the readily identifiable viral DR sequences from all of these amplified integration site junctions was identical, despite the DRs being amongst the most variable regions of HHV-6A/-6B genomes. The shared degenerate repeat structures preceding both readily identifiable subtelomeric and viral DR sequences. Plus the distinct phylogentic grouping of nucleotide sequences from these CI-HHV-6A genomes believed to be integrated at 17p, when compared to CI-HHV-6A/-6B samples where integration site junctions could not be amplified utilising the 17p primer; shown here for U51 and further in more recent work utilising concatenations of both conserved and variable genes (U38, U46, U51 and U83) from throughout these CI-HHV-6A genomes [Tweedy et al., 2016]. However, without confirmation of the chromosomal locations of integration or further genomic data from these CI-HHV-6A samples, the possibility of multiple integration events remains. As greater characterisation of these integrated viral forms emerges, relationships between the isolates integrated at different chromosomal locations should become clearer.

Prevalence of Germ-line Integration

Integration of CI-HHV-6A at chromosome 17p appears prevalent in the samples examined here, in agreement with similar previous anecdotal suggestions of a high prevalence of integration at this chromosomal location [Morissette & Flamand, 2010]. As above, CI-HHV-6B was found to be integrated at this chromosomal location both here and previously. This suggests there may be some preference for integration at this chromosomal location. There are also differences in the relative prevalence of CI-HHV-6A and -6B with respect to their circulating counterparts. In American, European and Japanese populations, epidemiological studies have identified HHV-6B as the initial infection responsible for the fever and exanthem subitum disease typically associated with infection with a HHV-6 species [Yamanishi et al., 1988; Dewhurst et al., 1993; Zerr et al., 2005a; Hall et al., 2006]. While estimates vary, the prevalence of HHV-6B in these populations is much higher than HHV-6A [Hall et al., 2006; 2008]. However, the screening here of a key viral immunomodulatory gene from one of the largest groups of CI-HHV-6A/-6B samples collated to date suggested that the prevalence of the integrated forms are in fact similar to one another. A feature which was further confirmed in this laboratory through the screening of additional loci from in these samples [Tweedy et al., 2015a] and previous observations about the prevalence of CI-HHV-6A and -6B [Hall et al., 2006; 2008; Leong et al., 2007; Hubacek et al., 2009].

While effects on detection may play a role in prevalence differences observed between HHV-6A and -6B [Leibovitch et al., 2014]. There still appears to be clear distinctions between the relative prevalence of the germ-line CI-HHV-6A/-6B compared to acute HHV-6A/-6B infections. If HHV-6A and -6B possess the same integration capabilities, the prevalence of the integrated forms would be expected to reflect those of the circulating forms. These relative prevalence distinctions may reflect differences in germ-line integration efficiency between the two species. In this regard, HHV-6A has been shown to be the predominant cause of symptomatic primary infection in an African population yet to date CI-HHV-6A has yet to be detected in screens in African populations, although one CI-HHV-6B case has been defined in North Africa [Bates et al., 2009 ; Faten et al., 2012; Tembo et al., 2015]. While a number of factors, including human genetics could account for this, it may also have implications for germ-line integration efficiency.

An additional explanation may lie in the shared origins of integrated samples implicated above. The higher relative prevalence of CI-HHV-6A could result from HHV-6A integration events which predate those of CI-HHV-6B, allowing a greater length of time for expansion in the population. Additionally since integration at chromosome 17p seems to be prevalent for CI-HHV-6A, at least in Europe, this may represent an early HHV-6A integration event. The multiple loci analysis performed previously suggested greater divergence for CI-HHV-6A from HHV-6A than that seen for CI-HHV-6B [Tweedy et al., 2015a]. Furthermore, the pattern of gene divergence, including all of the genes which define the speciation of HHV-6A and HHV-6B [Dominguez et al., 1999; Gompels & Kasolo, 2006; Ablashi et al., 2014], observed in the chromosome 17p integrated isolate, 5055, could also be seen to support the idea of integration of an ancestral HHV-6A strain. Further screening in this, and additional geographic regions will allow differences to be identified. Additionally, attempts to date the integration events at chromosome 17p could also provide insight into the origins of this integration.

CI-HHV-6A genome

Utilising the established NGS methods, genomic analysis of three CI-HHV-6A isolates was performed, representing the first characterisation of these chromosomally integrated viral genomes. The complete UL region was derived for one strain, 2284, and conserved and divergent sets of genes for comparison in two further CI-HHV-6A genomes. Completion of the DR regions of the genomes was confounded by repetitive sequences, but the two genes DR1 and DR6, encoded in these regions were characterised. While these genome sequences remain to be fully completed, all CDS features of HHV-6A strain, as currently defined, are retained in CI-HHV-6A isolates with no evidence for indels outside of the previously identified loss of the distal pac sites from both genomic termini [Ohye et al., 2014; Huang et al., 2014]. Additionally, retention of a number of *cis* acting signals required for lytic viral replication, and likely any potential reactivation mechanism, was also confirmed. Together these suggest that CI-HHV-6A isolates remain intact, thus alongside gene expression from the integrated genomes, lytic replication of the viral genomes is likely possible if a reactivation/excision mechanism exists. With one of these isolates, 5055, confirmed to be integrated at chromosome 17p and the implications above that integration at this site may represent an early CI-HHV-6A integration event; there is a possibility that many if not all of these integrated HHV-6A/-6B genomes remain intact and replication competent. During the work towards this thesis, the sequencing of a CI-HHV-6B isolate was reported, similarly indicating the integrated viral genomes remain intact with no disruption to defined ORFs [Huang et al., 2014]. However, this genome shared much closer relation to circulating HHV-6B strains, with overall less than 0.2% divergence, while CI-HHV-6A genomes had 3-6% divergence in some genes. The methods established here should be useful to further characterise CI-HHV-6A/-6B genomes directly from clinical samples, in order to understand their genetic content and pathological associations.

Comparisons of the integrated genomes to circulating viral strains highlighted increased divergence at 16 genes. This was in marked contrast to the sequenced CI-HHV-6B genome, where less than 300 SNPs were observed across the genome [Huang et al., 2014]. These CI-HHV-6A divergent genes included all the genes which have recently been used to define the speciation of HHV-6A from HHV-6B [Dominguez et al., 1999; Gompels & Kasolo, 2006; Ablashi et al., 2014]. Including genes with roles in transcriptional regulation, virus infection, as well as cell cycle and immune modulation. The phylogenetic grouping of these genes, ancestral to HHV-6A with divergence towards that observed in sequenced HHV-6B strains, supports the interpretation here that the CI-HHV-6A genomes could represent integration events of ancestral HHV-6A isolates.

Minor variant analysis indicated the possibility of superinfection with circulating viral strains in two of the CI-HHV-6A samples examined. Genomic DNA from the integrated genomes would be in all cells, while lytic replication will produce high viral titres but this would only be in a minority of cells. As such the predominant sequence should be that of the integrated genome. However, the potential to confound sequence resolution remains, especially in the case of low genome coverage. Sample choice should be considered in further characterisation of the integrated forms of these viral species.

Role of Germ-line Chromosomal Integration

The analysis of these CI-HHV-6A genomes suggests a series of rare integration events in the germline. In a number of ways, these germ-line integration events of HHV-6A/-6B can

be seen as analogous to the integration which has occurred amongst certain members of the *Retroviridae*, known as the human endogenous retroviruses (HERVs). Within the *Retroviridae* family, viral genome integration represents an obligatory step in viral replication. Following host cell entry of infectious virions, reverse transcription of the ssRNA(+) genome produces a linear dsDNA copy. A virally encoded integrase, then mediates random integration of the dsDNA copy into the host chromosome [Li et al., 2006]. Unlike integration events seen in a number of other viruses, excision of the provirus is not required for continuation of the viral life cycle, instead the provirus remains integrated into the cellular chromosome. This somatic integration with horizontal transmission via infectious virions, represents the exogenous form of retroviruses.

However, retroviruses can also be seen to exist in a vertically transmissible endogenous form, resulting from the integration of ancient retroviruses into the germ-line [Weiss, 2006]. Indeed, in humans it is believed that ~8% of the human genome consists of the remnants of germ-line integration of ancient retroviruses [Li et al., 2001]. The majority of these HERVs are defective resulting from the accumulative effects of nonsense mutations, insertions, deletions and recombination events since their original integration. However, a number of HERVs, primarily those representing more recent integration events such as the HERV-K family, retain intact ORFs with active gene expression and some even forming viral-like particles in certain tissues and pathologies including MS [Dewannieux et al., 2005; Boller et al., 2008; Contreras-Galindo et al., 2015; Kurth & Bannert, 2010; Balada et al., 2010]. Upregulation of expression from HERVs has also been noted following infection by other viruses, such as exogenous retroviruses [van der Kuyl, 2012; Toufaily et al., 2011; Perzova et al., 2011], but also notably following infection with herpesviruses, including HHV6A and -6B [Bergallo et al., 2015; Turcanova et al., 2009; Tai et al., 2009]. Indeed the simultaneous presence of HHV-6A and HERV antigens has been linked to increased cellular immune responses, the implication being that such responses may have a role in MS progression [Brudek et al., 2004].

Unlike the HERVs, the integrated CI-HHV-6A genomes, at least in the case of those examined here, appear to be intact with regard to currently defined coding content and many of the *cis* acting signals required for viral replication, and likely any reactivation mechanism. The only loss being the pac 2 site from DRR, and likely the pac 1 site from DRL as previously observed [Ohye et al., 2014; Huang et al., 2014], which have implications for any proposed mechanisms of integration and possible reactivation.

Like the HERVs it appears that expression from CI-HHV-6A/-6B genomes is possible, there are now multiple reports of the detection of transcripts from integrated genomes including studies from this laboratory concerning the patients analysed in this thesis [Daibata et al., 1998a; Clark et al., 2006a; Strenger et al., 2014; Huang et al., 2014; Tweedy et al., 2015a]. Similarly, other pathogens may be able to provide a trans-activating function to activate the integrated HHV-6A/-6B genomes [Katsafanas et al., 1996; Tanaka-Taya et al., 2000; Prusty et al., 2013a; Prusty et al., 2013b]. Since vertical transmission leads to the presence of the viral genome in every nucleated cell, gene expression and/or reactivation from the integrated virus has the potential to affect early development. Recent reports on the length of the CI-HHV-6A/-6B associated telomere suggest it is frequently amongst the shortest in somatic cells, with the implication the integrated viral genome may be perturbing telomeric function resulting in rapid telomeric shortening [Huang et al., 2014]. Such events may have implications for reactivation of integrated genomes, discussed further in section 6.7. But they could also have effects on expression from the integrated

viral genome. For example, it has been shown that the transcription of telomeric repeat containing non-coding RNAs (TERRA) which start in the subtelomeric region [Azzalin et al., 2007] can be induced by telomeric shortening [Cusanelli et al., 2013]. Which could mediate expression from the viral genome. Furthermore, the ubiquitous presence of the integrated viral genome implies gene expression may also be possible in tissues normally non-permissive for viral replication *in vivo*. These and other implications of CI-HHV-6A/-6B for disease is considered in more detail in section 6.8.

In this context the findings here support the notion that integration of CI-HHV-6A/-6B into the germ-line represent a series of rare events. If reactivation of these integrated forms is possible, this vertical transmission could be viewed as providing a replicative advantage through an additional transmission route. However, the evidence points towards these being aberrant or artifactual events. The recent findings suggestive of somatic integration of HHV-6A/-6B as a mechanism of latency [Arbuckle et al., 2010; 2013], alongside those identifying the ability of HHV-6A/-6B to bind to the sperm acrosome [Kaspersen et al., 2012], provide a mechanism for how these germ-line integration events may have occurred.

6.6 Somatic Chromosomal Integration of HHV-6A and -6B

While the focus of this work has been to characterise the inherited forms of HHV-6A and -6B, the mechanism underlying the germ-line integration events that gave rise to these inherited forms may stem from a naturally occurring mechanism utilising homologous recombination between cognate telomeric repeat sequences and employed during infection of somatic cells. As the two are likely intrinsically linked, the potential for viral integration during somatic infection will be considered here.

Currently, *in vivo* evidence for the somatic cell chromosomal integration of HHV-6A/-6B does not exist. However, the recent work on integration and reactivation of HHV-6A/-6B *in vitro* has led to suggestions that in contrast to the formation of nuclear episomes observed amongst many members of the *Herpesviridae*, chromosomal integration represents an alternative mechanism for the establishment of latency by HHV-6A and -6B [Arbuckle et al., 2010; 2013]. This would not represent a unique feature, as similar strategies of host chromosomal integration for the establishment of a dormant form can be observed in a number of other viral groups. Indeed, as a result of early work on bacteriophages during the 1950s, it has long been recognised that viral genetic material can integrate into the genome of host cells [Lederberg & Lederberg, 1953]. Upon infection of host cells, the viral reproduction cycle of temperate phages can proceed in two major ways: the lytic cycle or the lysogenic cycle. In contrast to the active production of progeny phages in the lytic cycle; during the lysogenic cycle integration of the viral genome into the bacterial chromosome results in a quiescent form of the phage genome known as a prophage. Periodically environmental cues act as signals of favourable viral replication conditions resulting in induction, whereby the integrated phage genome is excised from the host cell chromosome and re-enters the lytic cycle [Hendrix, 2013]. Additionally reproduction of a bacterium following phage integration, results in the presence of the prophage in both daughter cells. Theoretically allowing for indefinite replication of the prophage by prokaryotic replication. Thus the lysogenic cycle can be seen to provide a replicative advantage to the virus through the induction of a latent state until adequate conditions for viral replication occur. As well as an alternative route of transmission to new host cells.

Since the discovery of genomic integration in bacteriophages, there has been recognition that a number of animal viruses can also integrate into cellular genomic DNA [Choo et al., 1987; Feng et al., 2008; Murakami et al., 2004; Belyi et al., 2010; Delecluse & Hammerschmidt, 1993; Kripalani-Joshi & Law 1994; Pellet et al., 2012]. As exemplified by the Human Papillomaviruses, many of these represent occasional aberrant integration, with no defined role in viral replication, noted mainly through roles in oncogenesis [Choo et al., 1987; Moody & Laimins, 2010]. However, for a number of viral groups cellular integration represents a prerequisite for viral replication or an occasional feature which appears to play a role in the viral replication cycle. Two in particular, Gallid herpesvirus 2 (GaHV-2), more commonly known as Marek's Disease Virus (MDV), and the Adeno-associated viruses (AAV), share notable similarities to the integration observed in HHV-6A and -6B.

Gallid herpesvirus 2 /Marek's Disease Virus

MDV is responsible for Marek's disease in chickens, which is characterised by the rapid development of malignant T-cell lymphomas. Due to its oncogenic nature it was originally classified within the *Gammaherpesvirinae*, however, sequencing of the viral genome revealed it was in fact closely related to HSV-1 and VZV and led to the classification as the type species of a new genus, *Mardivirus*, within the *Alphaherpesvirinae* [Tullman et al., 2000; Osterrieder et al., 2006]. The current model of MDV pathogenesis proposes inhalation of virus containing dander leads to the infection of phagocytic cells in lung tissue and transport to main lymphoid organs [Baaten et al., 2009]. At these sites infection of B cells leads to transfer of the virus to activated T cells, the targets for lytic replication and latency, respectively. Reactivation and lytic replication of the virus following transport of infected cells to feather follicles permits horizontal transmission of the virus via shedding [Calnek 2001; Osterrieder et al., 2006]. However, in a minority of the latently infected CD4+ T cells transformation occurs, which are the source of the lymphoma. During the course of latent infection, integration of the full length genomic DNA into the host cell chromosome has been observed to be a common occurrence, with suggestion that it is potentially a requirement for the establishment of latency in MDV [Delecluse & Hammerschmidt, 1993; Delecluse et al., 1993; Kaufer et al., 2011]. The latent integrated state in turn appears to be a prerequisite for cellular transformation, as the number of cells harboring latent genome directly correlates with efficiency of lymphoma formation [Kaufer et al., 2011]. Integration preferentially occurs in the telomeric region of the host chromosome, although the exact location with the telomere i.e. subtelomeric, intratelomeric or distal remain to be determined, with no observed preference for a particular chromosome [Delecluse & Hammerschmidt, 1993; Delecluse et al., 1993; Kaufer et al., 2011]. While the mechanism still remains elusive the presence of host telomeric repeat sequences in the genomic terminal repeat structures [Kishi et al., 1991] and a virally encoded telomerase subunit [Fragnet et al., 2003] likely play a role. The genomic termini of MDV contain arrays of telomeric repeats. In MDV, two arrays of repeats have been identified within the terminal repeat regions found at the ends of the genome, a short telomeric repeat region invariably consisting of 6 repeat units and a long telomeric repeat unit consisting of a variable number of repeats up to 100. Indeed, recent work utilising MDV telomeric repeat mutants found that the shorter of the repeat arrays was not essential for integration, although did affect integration efficiency. However, the longer variable copy number array is essential for directed insertion into host telomeres *in vivo*, [Kaufer et al., 2011; Greco et al., 2014]. In the same studies, complete deletion of the viral

telomeric repeat sequences still permitted integration, albeit less efficiently, into the host genome at non-telomeric sites suggesting an importance to this mechanism rather than merely a consequence of the presence of telomeric repeats. Furthermore, the evidence available suggests that this integration does not represent a replicative dead end with viral replication leading to production of linear genomic DNA [Delecluse et al., 1993; Kaufer et al., 2011; Schermuly et al., 2015]. Again, with the viral telomeric repeats implicated for a key role in this process as viral repeat mutants displayed reduced reactivation efficiency [Kaufer et al., 2011; Greco et al., 2014]. However, the mechanism for this reactivation from the latent integrated form remains unknown.

Adeno-Associated Viruses

The ubiquitous human parvoviruses, the adeno-associated viruses (AAVs), have also been shown to integrate into the host chromosome [Cheung et al., 1980; Calcedo et al., 2009]. These small single-stranded, linear DNA genome containing viruses possess two major ORFs, *rep* encoding four replication proteins and *cap* encoding three capsid protein [Srivastava et al., 1983]. In isolation AAVs are replication incompetent, instead for the production of infectious virions they require a helper virus, typically an adenovirus, although other viruses including the herpesviruses, HSV-1 and HHV-6A/-6B can also provide helper functions [Geoffroy & Salvetti, 2005; Stutika et al., 2015; Thomson et al., 1994b]. In the absence of a helper virus, AAVs establish a latent infection, permitting long-term persistence of the viral genome [Berns et al., 1975]. There are differing reports on the primary mode of how this persistence is achieved, with studies in culture indicating efficient integration of the viral genome into host cellular DNA, while others suggest that *in vivo* AAV genomes persist predominantly in an episomal form [Kotin et al., 1990; Samulski et al., 1991; Schnepf et al., 2005]. Whichever is the case, it is clear that AAV can establish latency through integration into the host genome following natural infection [Mehrlle et al., 2004]. This integration has largely thought to have been highly site-specific, with preferential integration into a site within the protein phosphatase 1 regulatory subunit 12C gene found on chromosome 19q13.42, a site known as AAVS1 [Kotin et al., 1990; Giraud et al., 1994; Linden et al., 1996]. However, more recently additional AAV integration sites have been noted on different chromosomes, 1q31.1, 3p24.3 and 5p13.3 [Schnepf et al., 2005; Huser et al., 2010]. The integration of AAV is dependent upon two major non-cellular factors. The two largest of the replication proteins, Rep68 and Rep78 [Weitzman et al., 1994]. In addition to Rep binding and endonuclease sites, consisting of GAGC repeats, found in the inverted terminal repeats located at the viral genomic termini, as well as present at sites in the human genome [Giraud et al., 1994; Feng et al., 2006]. The two large Rep proteins possess a range of functions, including DNA binding, ATPase, endonuclease and helicase activities [McCarty et al., 1994; Wonderling et al., 1995; Im & Muzyckza, 1990], playing integral roles in many other aspects of viral replication. During the process of integration, the large Rep proteins form a complex through the binding of Rep binding sites present in both the viral and the cellular genome sequences [Weitzman et al., 1994]. Rep68 through its helicase and endonuclease activity, mediates the specific introduction of nick at homologous motifs found within the viral and cellular Rep binding sites [Lamartina et al., 2000]. The remainder of the integration mechanism has yet to be elucidated but has been proposed to occur through a series of DNA synthesis and strand switching steps involving components of the non-homologous end joining pathway [Henckaerts et al., 2009; Daya et al., 2009]. Following integration, certain Rep transcripts act to suppress viral gene expression, permitting persistence of the viral genome until helper virus is present [Tratschin et al., 1986]. In the presence of helper virus Rep proteins

switch roles to act as activators of viral gene expression, with Rep expression likely playing a role in release of the AAV genome from the cellular chromosome allowing active replication to proceed [Pereira et al., 1997; Samulski et al., 1982; Ward et al., 1994]. While the mechanics of this rescue mechanism remain to be determined there is some evidence that it may occur through a Rep mediated nick in an inverted terminal repeat of the integrated AAV genome followed by DNA replication [Ward et al., 2003]. Additionally, while not specifically associated with any human disease states, it is clear that AAV infection has pronounced effects on host cells, including cell cycle arrest, induction of apoptosis and cellular gene repression, many of which involve the viral Rep proteins [Berthet et al., 2005; Schmidt et al., 2000; Dutheil et al., 2014].

HHV-6A and HHV-6B

Like MDV, HHV-6A and -6B contains telomeric repeat arrays within repeat regions at their genomic termini and genomic integration occurs at telomeric regions of the host chromosome. This integration occurs at a location close to the subtelomere, with suggestions that it potentially occurs through a homologous recombination event between the perfect TTAGGG telomeric repeat arrays encoded in the T2 region of the DR_R of the viral genome and the host chromosomal telomeric repeats. The integration event results in insertion of the HHV-6A or -6B genome in an orientation, where U100 is located towards the centromere and U1 towards the telomeric cap, as well as the loss of the pac2 site from the DR_R [Arbuckle et al., 2010; Ohyne et al., 2014; Huang et al., 2014]. The pac1 site of the DR_L is also lost during the integration event, suggesting either a second recombination event in the T1 region of the DR_L or erosion of the pac site followed by telomere elongation. [Ohye et al., 2014; Huang et al., 2014]. Therefore as demonstrated for MDV, this implies an importance for the virally encoded telomeric repeat arrays in HHV-6A/-6B chromosomal integration.

However, to date evidence suggests that the presence of human telomeric repeat sequences alone does not promote integration. Indeed genomic termini-like telomeric repeat sequences are also retained in a number of other distantly related herpesviruses, including members of the *alphaherpesvirinae* [Kishi et al., 1991; Tyler et al., 2011; Telford et al., 1995]. The conservation of these sequences could suggest an important role in viral function outside of integration, such as protection of the linear viral genome from cellular nucleases or linkage of episomal forms to the cellular chromosome. Although additionally, it could mean that telomeric integration events may be a feature in additional herpesviruses besides MDV, HHV-6A and HHV-6B. The other human roseolovirus, HHV-7, also possesses similar telomeric repeat sequences but to date, integration into host chromosomal DNA has not been noted [Hall et al., 2004]. However, the narrow tropism of HHV-7, preventing germ-line integration, and the minority of cells which become latently infected, or even just a minority of latently infected cells if multiple latency strategies can be employed, would make identification of possible integrated cells challenging. Such is the case for the identification of somatically integrated HHV-6A/-6B from clinical samples.

Observation of the insertion of a telomeric repeat lacking MDV mutants, albeit inefficiently, into the host chromosome [Kaufer et al., 2011] implies other factors play a role in the viral integration events. Likewise, there is a report of CI-HHV-6B integration at a non-telomeric region of chromosome 12 (12q14) [Goel et al., 2013], which could also suggest factors outside of the presence of telomeric repeats also play a role in HHV-6A/-6B integration events. One likely possibility is the product of the U94 ORF, an ORF

encoded by both HHV-6A and -6B yet absent amongst other human herpesviruses, which may play a role in the promotion of integration [Thomson et al., 1991; Rapp et al., 2000]. U94 encodes a protein with homology (~24% amino acid identity) to the integrase/replicative AAV-2 Rep 68/78 proteins [Thomson et al., 1991]. The ability of HHV-6A/-6B to act as a helper virus for AAV replication [Thomson et al., 1994b], in addition to a number of cases of suspected gene piracy, notably the viral chemokine and chemokine receptors, U83, U12 and U51 [Gompels et al. 1995; Vischer et al., 2006]. Indicate a mechanism by which an ancestral HHV-6 member could have acquired a rep gene homologue. With the U94 ORF being highly conserved among HHV-6A and -B strains [Rapp et al., 2000], expressed at high levels during latency [Rotola et al., 1998], and pU94 possessing a role in the inhibition of viral replication [Rotola et al., 1998; Caselli et al., 2006], it has been suggested that it may play an essential role in the establishment of latency. Indeed, U94 retains many of the activities of the AAV Rep proteins, namely single stranded DNA binding activities to telomeric repeat containing regions [Mori et al., 2000; Dhepakson et al., 2002; Caselli et al., 2006; Trempe et al., 2015], 3'-5' exonuclease, ATPase and helicase activities [Trempe et al., 2015] and can functionally complement a AAV-2 Rep deletion mutant [Thomson et al., 1994b]. Additionally, there is some indication that these activities of U94 can also disrupt the host chromosome telomeric D-loop [Trempe et al., 2015], the displacement loop created by strand invasion of the G-overhang found at the chromosomal terminus, to allow formation the T-loop structures that protect the chromosome ends [O'Sullivan & Karlseder, 2010]. Thus U94 may play a similar role in the HHV-6A and HHV-6B replication cycle, including the possibility of involvement in integration of viral genomic DNA.

Thus like the integration observed for MDV and AAV, somatic integration could play a role in HHV-6A/-6B latency. Chromosomal integration could be seen to be particularly advantageous in HHV-6A/-6B due to the high turnover rate of many of productively infected cell types. Ensuring maintenance of the viral genome during cell division. The site of integration of the telomeres has also been suggested to be important for latency. For example, silencing of genes in close proximity to telomeres has been noted [Baur et al., 2001]. Aiding to limit expression of viral transcripts which could trigger immune responses. However, telomeres also represent dynamic structures and the transcription of telomeric repeat containing non-coding RNAs (TERRA) is a feature of this [Azzalin et al., 2007]. Additionally U94 is expressed at high levels during latency and inhibits viral replication [Rotola et al., 1998; Caselli et al., 2006]. So analogous to the AAV Rep transcript mediated suppression of viral gene expression, U94 may play an important role. Finally if HHV-6A/-6B integration does represent a bona fide mechanism of latency, a reactivation mechanism must exist for excision of the viral genome from the cellular chromosome in a form that would permit lytic replication to proceed.

6.7 Reactivation of Integrated HHV-6A/-6B

In addition to the gene expression shown in the cohort of CI-HHV-6A/-6B patients analysed in this laboratory, there is increasing evidence emerging with regard to the reactivation of CI-HHV-6A and -6B. Treatment of PBMCs from CI-HHV-6 patients and a CI-HHV-6 cell line with trichostatin A (TSA) and to a lesser extent 12-O-tetradecanoyl-13 acetate (TPA), compounds known to reactivate latent herpesviruses, could reactivate both CI-HHV-6A and -B to produce infectious virions [Arbuckle et al., 2010; Arbuckle & Medveczky, 2011; Arbuckle et al., 2013]. The sequence of the gB gene in transplacentally acquired HHV-6A cases was shown to be identical to that of the mothers CI-HHV-6A yet divergent from

other known HHV-6 isolates, implicating transmission of reactivated CI-HHV-6A [Gravel et al., 2013b]. Similarly, in the case of a patient with X-linked severe combined immunodeficiency (X-SCID) and inherited CI-HHV-6A on chromosome 22, both RT-PCR detection of viral RNA and isolated virus were identical to the integrated virus from both the patient and father, yet distinct from known circulating strains [Endo et al., 2014]. Extra-chromosomal HHV-6A/-6B DNA has also been noted following *Chlamydia trachomatis* infection of CI-HHV-6A/-6B cells lines and PBMCs from patients with CI-HHV-6A/-6B [Prusty et al., 2013a; Prusty et al., 2013b]. While frequent detection of transcripts from an inhibitor of viral replication, U94, during latency [Rotola et al., 1998; Caselli et al., 2006], as well as spliced transcripts from the the IE transactivator, U90 [Kondo et al., 2002; Kondo et al., 2003; Huang et al., 2014], also support a role for integration as a mechanism of latency.

Thus like MDV and AAV, there appears to be a mechanism for excision of the viral genome from the cellular chromosome, and circularisation of the genome that would permit active viral replication in permissive cells. This is similar to the involvement of the Rep proteins of AAV in reactivation. HHV-6A/-6B U94 could potentially play a role in excision of the viral genome. There is evidence that U94 can disrupt the host chromosome telomeric D-loop [Trempe et al., 2015]. However, Rep mediated excision requires endonuclease activity, which was not detected for U94 [Ward et al., 2003; Trempe et al., 2015]. Also with AAV possessing a ssDNA genome, the proposed mechanism does not seem to fit for the excision of the dsDNA HHV-6A/-6B genome. Similarly, the dsDNA genome of MDV is excised from a telomeric region in the apparent absence of a Rep homologue [Delecluse et al., 1993; Kaufer et al., 2011; Schermuly et al., 2015]. While this mechanism of MDV reactivation is yet to be defined, the viral telomeric repeats appear to play a key role [Kaufer et al., 2011; Greco et al., 2014].

Two groups have noted extra-chromosomal circular molecules consisting of viral genomic DNA with a single complete DR [Huang et al., 2014; Prusty et al., 2013b]. Providing compelling evidence supporting the hypothesis that a t-loop/t-circle excision mechanism, could permit viral reactivation. In this model, strand invasion into the T1 region of DR_R mediated by the telomeric repeats permits the excision of a circular viral genomic structure which has the potential to act as a template for rolling circle replication generating concatameric genomes with intact DRs and pac sites, as observed in infectious virions. Implying an importance to the virally encoded telomeric repeat sequences for reactivation, like that seen in MDV. Additionally, even if viral replication cannot be completed, the possibility for excision and re-integration into the cellular chromosome as a transposable element cannot be ruled out.

There was implication that this t-circle formation may be increased by telomere shortening [Prusty et al., 2013b]. As suggested earlier, TERRA transcripts could play a role in expression from the integrated viral genome through induction at shortened telomeres [Cusanelli et al., 2013]. However, certain TERRA transcripts or aberrant TERRA transcription has also been associated with telomere shortening [Pfeiffer & Lingner, 2012; Maicher et al., 2012]. Thus the integrated viral genome may actually be interfering with TERRA associated mechanisms alongside inducing them, in turn facilitating excision via the proposed t-circle excision mechanism through a feedback loop.

Prusty and colleagues also found that telomere shortening was induced by *Chlamydia trachomatis* infection [Prusty et al., 2013a; Prusty et al., 2013b]. Superinfection with HHV-7

has also been shown to reactivate latent HHV-6B [Katsafanas et al., 1996; Tanaka-Taya et al., 2000]. Additionally, the minor variant analysis performed here also suggested that superinfection with HHV-6A/-6B was associated with expression from the integrated genomes. Thus infection with exogenous viruses could also promote reactivation by a similar mechanism. Even in the absence of reactivation, co-infection of CI-HHV-6A/-6B cells with a circulating HHV-6A/-6B strain could have the potential for recombination. In this regard, the multiple loci analysis performed previously in this laboratory on these integrated genomes noted cases of intra- and inter-genic recombination [Tweedy et al., 2015a]. Recombination with superinfecting virus could potentially provide sources of divergent genes/variation for exogenous HHV-6A/-6B strains.

6.8 Functional Consequences of CI-HHV-6A/-6B

Pathological associations with CI-HHV-6A/-6B are only beginning to be assessed. To date, the most apparent consequence of germ-line integrated HHV-6A/-6B appears to be the misdiagnosis of active HHV-6 infection with subsequent unnecessary treatment [Clark & Ward, 2008]. With regard to directly associated pathologies, while there have been some implications due to the higher presence of CIHHV-6A/-6B in certain conditions, other studies have provided conflicting results [Pellet et al., 2012; Morissette & Flamand, 2010; Hubacek et al., 2009; Hill et al., 2015]. As such definitive links between many pathologies and CI-HHV-6A/-6B remain largely undefined.

However, there is speculation about a number of possible consequences of CI-HHV-6A/-6B, although these might not be present in each CI-HHV-6A/-6B individual. Firstly, integration of the virus into the telomeric regions of the host chromosome may have profound effects on the integrity of the chromosome. Indeed in somatic cells disruption of the telomere function on the integrated chromosome appears to be associated with telomeric shortening [Huang et al., 2014]. Thus hastening of the onset of cellular senescence may be a feature directly related to the integration of the viral genome [Shay & Wright, 2005]. Dysfunction of the telomeric region is also linked to a number of different disease states [O'Sullivan et al., 2002; Sahin et al., 2011], while a role in chromosomal instability could also be a factor in progression to cancer [Jefford & Irminger-Finger, 2006]. These would also likely be a factor in somatic cell integration for latency, in non-CI-HHV-6A/-6B individuals. Second there could be gene expression. Studies in addition to those on the cohort reported here have shown that expression of genes from all of the different kinetic classes can occur from CI-HHV-6A/-6B genomes, including some potentially important inflammatory modulators, U51 and U83 [Daibata et al., 1998a; Clark et al., 2006a; Strenger et al., 2014; Tweedy et al., 2015a]. Third, reactivation of the integrated viral genomes could occur, as discussed above. Thus effects may be mediated by viral gene expression and/or replication, alongside any disruption of the host chromosome. Although, whether infectious virus is produced is still under evaluation.

Since CI-HHV-6A/-6B is transmitted vertically there may also be development implications. Recent evidence has led to suggestions of the existence of a condition termed inherited herpesvirus 6 syndrome (IHS) [Pantry et al., 2013]. Individuals with CI-HHV-6A/-6B had previously been noted as possessing a decreased humoral response to an important HHV-6 antigen, gB, in comparison to a healthy population sample [Tanaka-Taya et al., 2004]. In addition, antiviral use for the successful resolution of clinical symptoms (and subsequent relapses upon cessation of treatment) in two CI-HHV-6A patients displaying cognitive dysfunction, suggested symptomatic reactivation of CI-HHV-6A or a reduced

resistance to circulating HHV-6A/-6B strains could play a role in disease [Montoya et al., 2012]. Subsequently, the demonstration that the sequence of late HHV-6 mRNAs from CI-HHV-6A patients presenting with neurological disease differed markedly from the sequence of the inherited viral genome, led to the suggestions of the existence of IHS [Pantry et al., 2013]. In which individuals with inherited CI-HHV-6A may possess a level of immune tolerance or a weakened immune response to HHV-6A antigens, resulting in an increased risk of persistent exogenous infections and associated disease states. Additionally, evidence has been presented that congenital HHV-6A/-6B, occurring primarily from vertical transmission of CI-HHV-6A/-6B, has negative effects on neurodevelopment [Hall et al., 2008; 2010; Caserta et al., 2014].

HHV-6A/-6B infection has long been associated with inflammatory pathology. Similarly, CI-HHV-6A/-6B also has links to inflammatory disease states. A recent study has found links between persistent myocardial complications and CI-HHV-6A/-6B, with symptomatic episodes associated with viral gene expression and detection of viral particles in degenerating myocytes and interstitial cells [Kuhl et al., 2015]. Both symptoms and detection of viral mRNA were abolished with antiviral treatment. This was in the same patient cohort analysed here. Furthermore, a recent large scale analysis has also found CI-HHV-6A/-6B to be a risk factor in the development of angina, with the prevalence more than three times greater in CI-HHV-6A/-6B than control groups [Gravel et al., 2015]. The implication being that expression from the integrated viral genome and/or reactivation leads to immune activation with cellular damage that promotes myocardial complications.

Finally, there may also be implications for the use of samples from donors with inherited CI-HHV-6A and/or -6B [Flamand et al., 2010]. Where highly immunosuppressive states may permit more frequent viral reactivation.

6.9 CI-HHV-6A U51

While many are regarded as dispensable for viral replication *in vitro*, the extensive presence of genes which share homology to chemokine and chemokine receptors amongst the β - and γ - herpesviruses suggests a key significance to these genes to *in vivo* replication of these viruses. It is also clear that these virally encoded modulators of the chemokine system are all highly divergent in both sequence and signalling activities. Implying specific roles for these genes in each individual viral species related to their respective tissue tropisms and replication cycles, rather than a more generalised function amongst all herpesviruses. For example, even between the two closely related HHV-6 species, significant variation can be observed in function with U83B an agonist of CCR2 [Zou et al., 1999; Luttichau et al., 2003; Clark et al., 2013], while U83A is a highly potency agonist for CCR1, CCR4, CCR5, CCR6 and CCR8 [Dewin et al., 2006; Catusse et al., 2007; Clark et al., 2013].

Although clearly distinct analogies can be noted between U51 and one of the viral GPCRs encoded by the betaherpesvirus HCMV. HCMV encodes four chemokine-like GPCRs, US27, US28, UL33 and UL78 [Chee et al., 1990; Gompels et al., 1995]. While HHV-6A/-6B U51 is a positional homologue of HCMV UL78, there are significant differences between these GPCRs. Notably, the lack of constitutive signalling and apparent ligand binding displayed by UL78. Instead, HHV-6A U51 appears to share more similarities with the most extensively studied of these HCMV GPCRs, US28, notably in its multiple ligand specificity across different chemokine ligand classes.

HCMV US28

US28 is transcribed as an early lytic gene with the ability to act as a high affinity promiscuous β -chemokine (CCL2, CCL3, CCL4 and CCL5) [Neote et al., 1993; Gao & Murphy, 1994; Kuhn et al., 1995; Vieira et al., 1998; Billstrom et al., 1998] and δ -chemokine (CX3CL1) binding GPCR [Casarosa et al., 2001]. This feature in conjunction with its ability to undergo constitutive endocytosis and recycling via multiple routes [Fraile-Ramos et al., 2001; Droese et al., 2004] has led to the suggestion that pUS28 can sequester CC chemokines from the microenvironment thereby acting as a chemokine sink for immune evasion [Bodaghi et al., 1998; Vieira et al., 1998; Randolph-Habecker et al., 2002]. However, others have challenged this notion by suggesting that physiological concentrations of chemokines are too high to be efficiently scavenged by US28 [Boomker et al., 2006a].

The signalling capabilities of pUS28 are highly complex. It is a constitutively active vGPCR [Casarosa et al., 2001; Minisin et al., 2003], with two recent crystal structures of US28 suggesting the evolution of an amino acid network in the transmembrane domain of the receptor which destabilises the receptors inactive state [Burg et al., 2015]. This constitutive activity can be differentially modulated by binding of its β - and δ -chemokine ligands [Vomaske et al., 2009a]. Leading to ligand-specific and cell type-specific activation of alternative signalling pathways [Vomaske et al., 2009b], through coupling to multiple subclasses of G proteins (G_{α_i} , $G_{\alpha_{16}}$, $G_{\alpha_{q/11}}$ and $G_{\alpha_{12}}$) [Billstrom et al., 1998; Casarosa et al., 2001; Melnychuk et al., 2004] and $G\beta\gamma$ signalling [Casarosa et al., 2001]. The signaling pathways utilised by US28, have been extensively reviewed recently [Vischer et al., 2014] so will not be fully discussed here, however, some of the major functions attributed to this signalling by pUS28 will be described below.

Constitutive signalling activity has been shown to be directly linked to modulation of the intracellular environment for activation of the HCMV lytic cycle, through transactivation of the HCMV major immediate-early promoter [Boomker et al., 2006b]. The constitutive signalling activity of pUS28 has also been implicated in an oncomodulatory role which promotes tumourigenesis. pUS28 expression upregulates the expression of vascular endothelial growth factor (VEGF) and Interleukin-6 (IL-6) and positively modulates β -catenin signalling leading to enhanced cell cycle progression and transformation of cells *in vitro* and the promotion of tumour formation *in vivo* which may be further enhanced by inflammatory chemokines (CCL2) [Maussang et al., 2006; Maussang et al., 2009; Bongers et al., 2010; Slinger et al., 2010; Langemeijer et al., 2012]. pUS28 is also an important mediator of the movement of infected cells. Its signalling capabilities can directly promote chemotaxis of primary arterial smooth muscle cells (SMCs) in the presence of constitutively produced CCL2 [Streblow et al., 1999]. It is hypothesised this virus mediated chemotaxis of infected cells to sites of inflammation provides a mechanism for the transmission of HCMV to macrophages and subsequent dissemination throughout the body. However, the migration of virally infected cells to inflammatory sites also provides a molecular basis for the link between HCMV infection and acceleration of vascular disease [Streblow et al., 1999]. In addition, pUS28 has been demonstrated to increase the motility of infected cells on a CX3CL1-modified surface, used to simulate an activated endothelial surface, thereby potentially modifying their recruitment and dispersal *in vivo* [Hjorto et al., 2013]. The expression of pUS28 in latently infected monocytes has also led to suggestions of a role HCMV mediated manipulation of the dissemination of latently infected cells

[Beisser et al., 2001]. Finally the fact that pUS28 can bind different viral proteins has led to suggestions it may serve a role in HCMV dissemination through the enhancement of cell-cell fusion [Pleskoff et al., 1998].

HHV-6A U51

The aim in this work was to assess the functional significance of variation in the amino-terminal domain of U51A noted in all CI-HHV-6A samples examined here, in addition to HHV-6A strains GS and AJ. The implication being that the ligand binding profile of this U51A variant in the integrated genomes may differ from that of strain U1102, which has been used almost exclusively in the characterisation of U51A, to date. This may be important since HHV-6A is frequently associated with inflammatory pathology, which in part likely stems from dysregulation of the chemokine system, therefore in CI-HHV-6A patients there could be distinct immunomodulatory effects, such as the case of the myocarditis patients characterised here.

Similar to US28, U51A displays high levels of constitutive activity not frequently observed amongst human GPCRs [Fitzsimons et al., 2006; Catusse et al., 2008; Smit et al., 2007]. Constitutive activity which can be modulated to multiple subclasses of G proteins through differential binding of a promiscuous ligand binding profile, encompassing ligands from multiple chemokine groups [Milne et al., 2000; Fitzsimons et al., 2006; Catusse et al., 2008]. Constitutive activity by HHV-6A pU51 has been shown to specifically lead to the transcriptional down-regulation of two cellular targets. Firstly, CCL5, which has been demonstrated in both epithelial and haematopoietic cell lines, where its roles have been hypothesised to include immune evasion or modulation of recruitment of circulating inflammatory cells for systemic spread of the virus [Milne et al., 2000; Caruso et al. 2003; Catusse et al., 2008]. The second cellular target specifically down-regulated by HHV-6A pU51 is FOG-2 in a haematopoietic cell line, FOG-2 is a transcriptional repressor which may have a role in influencing the Th1/Th2 balance by repressing Th2 development, which may also contribute to cardiac and inflammatory complications [Catusse et al., 2008]. RNAi studies have also implicated constitutive signalling of U51 to have role in aiding viral replication through a role in direct cell-to-cell fusion [Zhen et al., 2005].

The functions of ligand binding and differential signalling remain largely undefined. However, decreased PLC activation and abolishment of the inhibition of CRE-mediated gene transcription, compared to levels seen for the constitutively active receptor state are noted [Fitzsimons et al., 2006]. Making it clear that distinct changes to the intracellular environment occur in response to ligand binding, perhaps as an immune evasion strategy. Or alternatively like US28, to permit chemotactic responses to ligands for dissemination, such as that proposed for the CCL11 and CCL19 activity of U51 in dissemination to the lung and lymph nodes, respectively [Catusse et al., 2008]. Additionally, a 'chemokine sink' role as suggested for many other vGPCRs, also cannot be ruled out.

Therefore changes to the ligand binding specificity or even affinity of U51A have the potential to have significant effects on function in the viral life cycle or virally mediated pathology. While the functional significance of this amino-terminal variation could not be determined here. A panel of mutant receptor expression vectors has been constructed, which can hopefully shed light on this in the future. Additionally, the amino terminal alanine-scanning mutants may also have utility in the investigation of potential post-translational modifications. Since tyrosine sulfation and N- and O-linked glycosylation are

frequent features on this domain of a number of human and viral chemokine receptors [Szpakowska et al., 2012; Paulsen et al., 2005; Margulies & Gibson, 2007; Feng et al., 2010; Wu et al., 2015]. Finally since RNAi downregulation of U51 negatively impacts viral replication, understanding the binding interactions may have utility in the design of novel therapeutics. Recently, the receptor-ligand interaction properties of US28 have been utilised to direct toxin fusion proteins to HCMV infected cells, with promising anti-viral activity *in vitro* and *in vivo* [Spiess et al., 2015].

6.10 Future Work

Completion of the genome sequences of the CI-HHV-6A isolates examined here should be one of the first major goals. For the UL region of the genomes, this represents an easily obtainable goal through the inclusion of the under represented amplicons in further NGS runs. While the completed UL region of isolate 2284 could also be utilised as a reference template for assembly. Resolution of the DRs and the repeat regions found amongst the UL region likely requires more extensive Sanger sequencing, however, the *de novo* assembly information available for these regions should help to guide such efforts. Further genome sequences of both CI-HHV-6A/-6B and HHV-6A/-6B will be able to give a clearer understanding of the relationships of these integrated forms to both exogenous viral strains as well as amongst the integrated isolates themselves. Information that will help in confirming many of the assertions made here about ancestral relationships between integrated viral genomes as well as the observed divergence. Attempts to finalise these sequences within this laboratory are currently ongoing.

Since analysis of integration site junctions at the subtelomeric region of chromosome 17p here, suggested a shared genetic lineage. NGS sequencing of any of the 18 further isolates suspected to be integrated at this chromosomal with subsequent phylogenetic analyses would help to further confirm this suggestions. Confirmation of this would then hopefully open the avenue to attempts to date the timing of the integration event at this chromosomal location. However, differing host and viral mutation rates, plus the potential for recombination with exogenous HHV-6A strains may prove problematic, in such attempts.

It would also be of merit to determine the chromosomal location of integration for the remainder of isolates subjected to whole genomes NGS. Inverse PCR from locations within the viral DR_R, as previously utilised for CI-HHV-6A/-6B [Arbuckle et al., 2010; Huang et al., 2014], likely represents the best method to achieve characterisation of previously undefined chromosomal integration sites. This would allow better characterisation of the genetic relationships between the CI-HHV-6A/-6B viral genomes, potentially integrated at different chromosomal locations. As well as defining the sequence of further integration site junctions, which may prove valuable in providing insight into the mechanism of viral integration into the subtelomeric region in both germ-line integration events as well as potential somatic integration for the establishment of latency.

Additionally, such analyses could be extended to the whole set of CI-HHV-6A/6B isolates. Since multi loci analysis has previously been performed on this set of isolates, reported in Tweedy et al., 2015a, comparisons of the chromosomal integration site location with phylogenetic analyses at these loci could also help to determine if there are relationships between integrated genomes at different chromosomal locations. Although the possibility of the derivation of more complete genomes utilising the methodology described here

would provide a much greater level of insight.

Further characterisation of integrated genomes can be achieved by the application of NGS methodology, such as that described here, to sequencing of additional CI-HHV-6A/-6B or HHV-6A/-6B isolates, alongside more extensive deep sequencing techniques to analyse viral populations, and expression of small protein and small RNA species. This will be key in determining the mechanisms underlying the control of viral latency and subsequent switches to a reactivated state. An additional consideration will be understanding the implication of the increasing complexity to the control of HHV-6A/-6B gene expression which is just beginning to be appreciated for other herpesviruses [Arias et al., 2014; Stern-Ginossar et al., 2012; Dolken et al., 2009]. Further, much of the mechanisms underlying both viral integration and reactivation are still to be elucidated. With the viral telomeric repeats and U94 identified as likely contributing factors attempts, targeted genetic manipulation of these features utilising previously generated HHV-6A/-6B BACs [Borenstein & Frenkel, 2009; Tang et al., 2010] will be important in functional assessment of their roles in integration, latency and reactivation mechanisms.

With regard to U51, there remains a panel of U51 mutants which remain to be functionally assessed if levels of surface expression can be improved. Alternatively assessment of the amino-terminal domain in isolation remains a possibility, for example with analyses utilising surface plasmon resonance [Majka & Speck, 2007]. While this would not permit determination of receptor activation, the affinity and kinetics of the interactions between the amino-terminal domain and various ligands could be assessed. Additionally there is still likely importance in investigating the consequences of the charge changes noted in the ECL regions of the receptor, since recent reports on the crystal structures of US28 bound to its chemokine ligand CX3CL1 implicated found such residues to be involved in the 'site 2' interactions that determine receptor activation [Burg et al., 2015]. Thus understanding of the mechanism of U51A binding and activation could help to with the development of small molecule inhibitors which could be utilised as intervention strategies due to the noted ability of U51 knockdown having significant impact on viral replication [Zhen et al., 2005].

6.11 Concluding Remarks

The work described in this thesis has begun to characterise the germ-line chromosomally integrated form of HHV-6A/-6B. Methods were established to allow the NGS of CI-HHV-6A/-6B genomic DNA directly from clinical samples, with confirmation through sequencing of HHV-6A strain AJ. Analyses of CI-HHV-6A integration sites on chromosome 17 supports the interpretation of a single ancestral germ-line integration event. While genomic analyses revealed CI-HHV-6A retains the full HHV-6A gene complement and cis acting signals required for viral replication/reactivation. However, the CI-HHV-6A genomes diverged as a set of genes which are markers of speciation amongst roseoloviruses. Overall, the results define unique characters of these germ-line integrated genomes compared to circulating viral strains.

Gene expression appears possible from these integrated genomes. Additionally, there is ever increasing support to the notion that somatic integration of HHV-6A/-6B may represent a mechanism for viral latency. In which case a mechanism for viral reactivation from the chromosomally integrated forms must exist. Therefore reactivation/gene expression with possible recombination with exogenous strains suggest that the

expansion of the integrated form in the human population could serve as source of emergent infection. The potential effects of this gene expression/reactivation on human health are only beginning to be understood. Further understanding of the mechanisms underlying this integration and reactivation may help in the development of therapeutic strategies to perturb the health effects of both CI-HHV-6A/-6B and HHV-6A/-6B.

Chapter 7: Appendices

7.1 The Human Chemokine System

7.1.1 Introduction

Chemotaxis describes the directional movement of leukocytes in response to a concentration gradient of an environmental stimulus. Chemokines, **chemotactic cytokines**, are a specific class of small, low molecular weight cytokines which when released from a cell establish a concentration gradient, along which a responding cell can move. For chemokines it is engagement of a cell surface receptor of the rhodopsin-like G protein-coupled receptor (GPCR) family, known as a chemokine receptor, that initiates intracellular signalling in the responding cell. This signalling through the orchestrated polymerisation of cytoskeletal actin filaments, myosin-mediated contraction and increased adhesive interactions with the extracellular matrix and adhesion molecules, ultimately leads to movement of the cell up the established chemokine concentration gradient; that is from a region of low concentration towards a region of high concentration in the vicinity of the source of the chemokine production. This mechanism thereby allows the targeted movement, or chemotaxis, of specific cells to specific sites or tissues within the host based upon both the expression of a chemokine gene at the source and the expression of relevant chemokine receptor genes by responding cells. While this definition still stands, it is becoming increasingly clear that the chemokine system plays an array of different roles in the responding cells alongside directing motility.

The chemokine superfamily represents a large, diverse and rapidly evolving gene family. Since their emergence numerous en bloc and tandem gene duplication events and subsequent divergence have given rise to a large number of genes encoding both chemokines and chemokine receptors [Zlotnik et al. 2006; Nomiyama et al. 2010]. In humans this has led to a chemokine gene family consisting of over 40 members and a chemokine receptor family with over 20 members [Pease & Williams, 2006]. Within the chemokine system there appears to be a high level of redundancy as receptors are often highly promiscuous, binding multiple ligands and ligands also binding to multiple receptors (figure 1.6). However, in recent years a number of challenges to the notion of this high level of redundancy have arisen. It has been proposed that responding cells are often expressing multiple chemokine receptors and will be responding to a multiplicity of chemokine ligands within their environment, the signals of which in turn are integrated into the appropriate response. As such this apparent redundancy may represent a mechanism for fine-tuning the regulation of chemotactic responses [Rajagopalan & Rajarathnam 2006]. Regardless, the potential promiscuous binding and differential expression profiles create great complexity in the understanding of the chemokine system, which is further compounded by factors such as post-translational modifications of chemokines, the differential affinities of ligands at a specific receptor or between a ligands receptors, the ability of different chemokines to initiate differential signalling pathways through the same receptor, and again the components of these signalling pathways may vary in a cell type specific manner [Schall & Proudfoot, 2011].

While the chemokine system was originally recognised as a pro-inflammatory mediator for its role in the control of haematopoietic cell migration during immune challenge. The advent and utilisation of genome sequencing and bioinformatics in particular, in combination with a number of readily identifiable characteristic features of chemokines

and their receptors, as described in sections 7.1.2 and 7.1.3, led to the discovery of many new chemokines and chemokine receptors, and the subsequent identification of a number of additional roles they play in physiological processes such as leukocyte homing and development, immune surveillance, angiogenesis/angiostasis, wound healing, cell activation, maturation and differentiation. As well as roles in a huge variety of pathological processes including autoimmune diseases, cancer, transplant rejections and microbial infections [Rossi & Zlotnik, 2000; Mackay, 2001; Fernandez & Lolis, 2002].

7.1.2 Chemokines

Chemokines are the largest family of cytokines identified. They are secreted proteins, with the noted exceptions of CX3CL1 and CXCL16 which have membrane anchored forms allowing them to also act as adhesion molecules [Bazan et al., 1997; Matloubian et al., 2000]. Mature chemokines are 8-14 kDa, basic proteins which even in cases of low overall sequence identity adopt a similar tertiary folding as determined by X-ray crystallography and/or NMR [Allen et al., 2007]. This tertiary structure is maintained by the presence of disulphide bonds between four conserved cysteine residues. (figure 7.1). The spacing between the first and second of these cysteine residues is used to classify the mammalian chemokines into four groups: CXC (a.k.a. α), CC (a.k.a. β), XC (a.k.a. γ) and CX3C (a.k.a. δ); with each chemokine being assigned an 'L' as signifier for 'ligand' and an identifying arabic number. Currently there are 17 α -chemokines designated CXCL1-CXCL17; 28 β -chemokines designated CCL1-CCL28; 2 γ -chemokines designated XCL1-XCL2 and a single δ -chemokine CX3CL1, formally recognised [Zlotnik & Yoshie, 2000; Bacon et al., 2002; Zlotnik & Yoshie, 2012]. However, the existence of isoforms and alternatively spliced variants plus the possibility of post-translational modifications can be considered to greatly increase the number of functional chemokines.

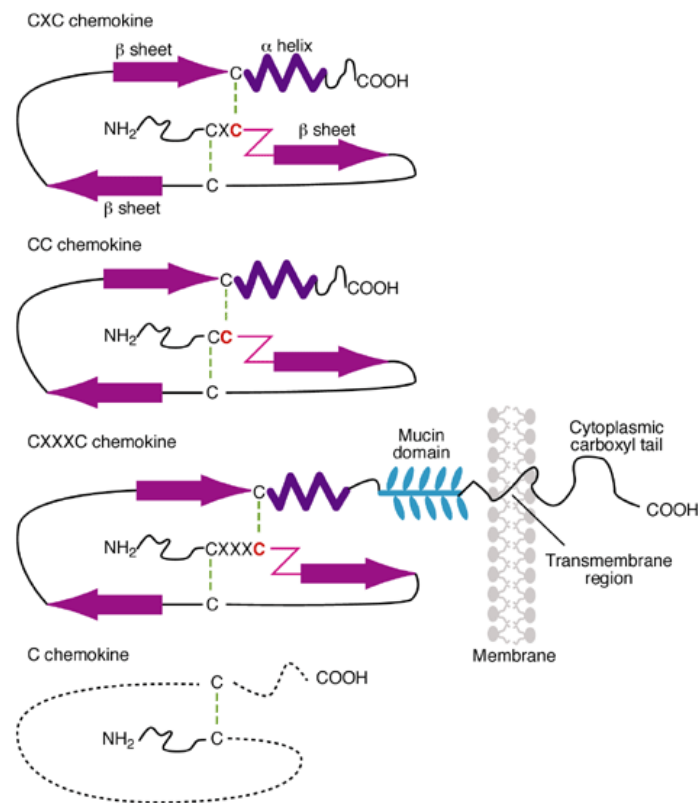


Figure 7.1 Schematic representation of the structural motifs which retain chemokine tertiary structure and define the four chemokine subfamilies. Beta sheet structures are depicted as arrows. Disulphide bonds between the conserved cysteine residues are represented by dashed lines. The CX3C chemokine is depicted with a transmembrane domain [Figure from Frederick & Clayman, 2001].

In addition to these structural criteria, chemokines may also be classified into two groups upon the basis of function: homeostatic and inflammatory chemokines (figure 1.6). The homeostatic chemokines are typically constitutively expressed in organs and tissues in the absence of inflammatory stimuli, resulting in leukocyte trafficking important during normally occurring processes such as immune cell maturation or immune surveillance. While inflammatory chemokines are predominantly only expressed by monocytic, epithelial, endothelial or fibroblastic cells in response to pro-inflammatory stimuli, to initiate the migration of leukocytes to inflammatory sites as well as activating other mediators of immune responses and wound healing. Typically, chemokines belonging to the homeostatic category are considered to bind a single receptor to exert their function(s), while the inflammatory chemokines display a more promiscuous receptor binding profile. However, as with many aspects of the chemokine system, the division of chemokines into these groups is not absolute, with a number falling into both categories [Allen et al., 2007].

While the interaction of chemokines with their respective receptors is the major interaction defining chemokine activity it is not the only interaction important for chemokine function. *In vivo*, predominantly on luminal surface of endothelial cells, certain chemokines also undergo essential lower-affinity interactions with glycosaminoglycan (GAG) moieties of proteoglycans; although the affinities and GAG specificities can vary widely, representing another mechanisms for the control of their site specific expression [Proudfoot et al., 2003; Witt & Lander, 1994; Kuschert et al., 1999]. GAGs are highly

charged, highly sulfated and heterogeneous polysaccharides which, while expressed by nearly all mammalian cells, possess expression patterns which can be varied, for example, upon a different pathological state [Mortier et al., 2012]. It is an electrostatic interaction between the negative charge on GAGs and the predominantly basic chemokine protein which is thought to be primarily responsible for binding [Hileman et al., 1998; Proudfoot, 2006]. While these GAG interactions are essential for the activity of certain chemokines *in vivo*, the exact mechanism still remains uncertain. It is believed that *in vivo* the secretion of chemokines alone, especially in presence of shear forces of blood flow, is not sufficient for the establishment and maintenance of the gradient required for leukocyte chemotaxis. As such the chemokine-GAG interaction permits a mechanism for localisation through establishment of the chemokine gradient [Handel et al., 2005]. However, the possibility that chemokine-GAG interactions may contribute to other important chemotactic processes, such as leukocyte arrest, protection from proteolysis and transcytosis of chemokines across the endothelium, still remains [Middleton et al., 2002; Salanga & Handel 2011].

In addition to the interaction with GAGs, many chemokines also possess the ability to form both homo- and hetero-oligomers both in solution at high concentrations and in the presence of GAGs [Handel et al., 2005; Proudfoot, 2006; Salanga & Handel 2011]. It should be noted that a number of chemokines exist in naturally occurring obligate monomeric forms, such as CCL1, CCL7 and CCL11 [Proudfoot et al., 2003]. While there is some evidence suggesting that these higher order oligomers may play a role *in vivo*, such as by increasing the affinity for GAGs or potentially modulating chemokine function, the true role of such structures remains unclear [Salanga & Handel 2011]. However, despite the existence of oligomeric forms, it is now generally accepted that the interaction of the chemokine with the receptor responsible for receptor activation and chemotaxis, occurs via the monomeric form. This was elucidated through the generation of obligate monomeric mutants of a number of chemokines, CCL2, CCL3, CCL4, CCL5 and CXCL8. These obligate monomeric mutants failed to induce chemotaxis in *in vivo* migration assays however, they retained wild-type receptor binding affinities and chemotactic activity *in vitro* [Paavola et al., 1998; Handel et al., 2008; Avalos et al., 1994; Laurence et al., 2000; Proudfoot et al., 2003; Rajarathnam et al., 1994]. Thus it seems while oligomeric forms may play a role in chemotactic responses *in vivo*, at least in the case of some chemokines, the actual chemokine-chemokine receptor interaction which leads to receptor activation and induction of cellular movement occurs via the monomeric chemokine form.

7.1.3 G Protein-Coupled Chemokine Receptors

Chemokine signals are transduced across the plasma membrane of responding cells through engagement of cell surface receptors of the G protein-coupled receptor superfamily (GPCRs). One of the predominant features for inclusion in the GPCR superfamily is the presence of a characteristic structure, consisting of seven predominantly hydrophobic stretches of 20-35 amino acids representing transmembrane α -helical domains, designated transmembrane domain 1-7 (TM1-TM7) [Allen et al., 2007]. These hydrophobic domains insert into the plasma membrane resulting in a protein which passes through the membrane seven times in an anti-clockwise manner to form an α -helical bundle. The location of these hydrophobic domains creates a structure with an extracellular region consisting of an N-terminal domain and three extracellular loops (ECL1-ECL3), which form a pocket for interaction with the extracellular ligand; and an

intracellular region consisting of three intracellular loops (ICL1-ICL3) and a C-terminal domain, which interact with intracellular proteins to transduce the extracellular signal from ligand binding to the interior of the cell, figure 7.2.

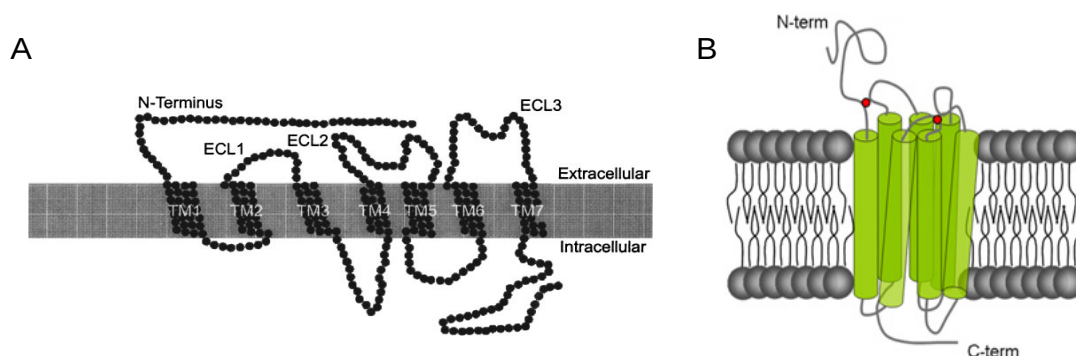


Figure 7.2 Schematic representations of chemokine receptor topology. Abbreviations: TM, transmembrane domain; ECL, extracellular domain. [Figures from Fernandez & Lolis, 2002 and Szpakowska et al., 2012].

With roughly 800 members in humans alone, GPCRs constitute the largest group of cell surface proteins involved in signal transduction [Fredriksson et al., 2003]. The sequence of these GPCRs ranges from a few hundred amino acids up to thousands in length, with most consisting of ~300-500 amino acids [Latek et al., 2012]. Generally speaking the extracellular region of GPCRs shows greater diversity in sequence length and identity, reflecting the diverse array of ligands which interact with members of this superfamily, such as hormones, peptides, amines, lipid mediators and photons [Allen et al., 2007]. While the intracellular regions show greater similarity due to their interaction with a smaller subset of signal transduction and regulatory proteins [Latek et al., 2012]. Based around sequence similarities and the presence of different characteristic motifs a number of different classification schemes have been proposed for the GPCR superfamily, however, for human GPCRs the most widely accepted is the GRAFS system which involves the division into five main families: Glutamate (G), Rhodopsin (R), Adhesion (A), Frizzled/Taste2 (F) and Secretin (S) [Fredriksson et al., 2003]. The Rhodopsin family is by far the largest of these families with characteristic features of this family including comparatively short N-terminal domains, a DRYLAIV or related motif at the TM3-ICL2 boundary, a CwxP motif in TM6 and a NPxxY motif in TM7 [Mirzadegan et al., 2003; Schioth & Fredriksson, 2005]. The Rhodopsin family can be further subdivided into four groups: α , β , γ and δ ; and it is to the γ -Rhodopsin group which the chemokine receptors belong [Fredriksson et al., 2003].

Chemokine receptors can be broadly classified into two categories, firstly, the 'classical' chemokine receptors defined as those signalling via G proteins; while more recently there has been increased recognition of chemokine receptors which apparently lack signalling capabilities or signal in a G protein-independent manner which are now referred to as atypical chemokine receptors [Bachelierie et al, 2013].

The Chemokine receptor group has been further subdivided by the class of chemokine ligand they bind: CXC (α), CC (β), XC (γ) and CX3C (δ); since although they may bind multiple chemokine ligands, human chemokine receptors are generally considered restricted to binding ligands from a single chemokine class [Bacon et al., 2002; Zlotnik et

al. 2006]. In addition to this, generally speaking functional human chemokines act as agonists at their respective receptors [Murphy et al., 2000]. In much the same manner as chemokine nomenclature, the chemokine receptors are named according to the class of chemokine they bind followed by an 'R' denoting receptor and a distinguishing arabic numeral. [Murphy et al., 2000]. To date there has been formal recognition of 18 distinct chemokine receptors: 6 CXC class receptors designated CXCR1-CXCR6; 10 CC class receptors designated CCR1-CCR10; 1 XC class receptor designated XCR1 and 1 CX3C class receptor designated CX3CR1 [Bachelier et al., 2013]. The binding profiles and commonly expressing cell types can be seen in figure 1.6 and table 1.2. As highlighted in table 1.2, while chemokine receptor expression is primarily regarded as a characteristic of leukocytes, it is not solely limited to them, indeed chemokine receptor expression has been noted on a wide variety of cell types including: epithelial cells [Dwinell et al., 1999; Jordan et al., 1999; Khurram et al., 2010], endothelial cells [Gupta et al., 1998], and a number of neuronal cell types [Bajetto et al., 1999; Flynn et al., 2003; Kremlev et al., 2004].

There is a second group of human receptors which retain chemokine binding properties and share structural homology to the G protein-coupled chemokine receptors but lack the ability to initiate 'classical' G protein signalling pathways upon ligand binding. In addition, these receptors generally also differ from their human G protein-coupled chemokine receptor counterparts in their unusual expression profiles, limited conservation of the important DRYLAIV G-protein signalling motif, and abilities to bind a broader range of chemokine ligands [Mantovani et al., 2006; Graham et al., 2012; Bachelier et al., 2013]. While previously known under a variety of names (decoy/regulatory/silent/scavenger receptors), recently the nomenclature for this group of receptors has been standardised to atypical chemokine receptors (ACKRs) along with the formal recognition of four members of this group, designated ACKR1-ACKR4, and the provisional recognition of two further members [Bachelier et al., 2013]. These receptors appear to play a vital role in the regulation of many aspects of the chemokine system through a 'fine tuning' of the chemokine levels at specific sites/tissues [Nibbs & Graham, 2013]. As such, dysregulation of these ACKRs has also been linked to important infectious disease and cancer pathologies.

7.1.4 Chemokine-Chemokine Receptor Interactions

Atomic structures of chemokine ligands in complex with their cognate receptors will be the key to fully understanding the basis of the interactions which define chemokine receptor binding and activation. However, structural determination of GPCRs has only really become viable in the past decade and still remains an arduous task. Therefore to date, mutagenesis studies and structural determination of peptide complexes have been the most readily employed methods to gain insight into chemokine-receptor interaction. Such studies on a variety of different CC, CXC and CX3C chemokines and receptors have been key to providing insight into the interactions of the chemokines with their receptors. These studies have revealed a number of generalised concepts regarding the binding and signalling epitopes present on chemokines and chemokine receptors.

The first of these is the flexible N-terminal region preceding the first conserved cysteine residue contains the determinants which are primarily responsible for the signalling/receptor activation event subsequent to receptor recognition/binding. This is demonstrated by the fact that N-terminal truncation mutants, addition mutants or modification of residues within the first 10 amino acids of numerous chemokines have

frequently been found to result in molecules which maintain high affinity binding to the ligands native receptor while diminishing or completely abolishing receptor activation and signalling [Gong & Clark-Lewis, 1995; Jarnagin et al., 1999; Laurence et al., 2000; Bondue et al., 2002; Struyf et al., 1998; Proudfoot et al., 1996; Simmons et al., 1997; Pakianathan et al., 1997; McQuibban et al., 2002; Struyf et al., 1999; Ott et al., 2004; Liston et al., 2009; Ott et al., 2006; Shinkai et al., 2002; Moser et al., 1993; Campanella et al., 2003; Crump et al., 1997; McQuibban et al., 2001; Inoue et al., 2005]. In fact *in vivo* it seems chemokines may undergo natural post-translational proteolysis resulting in N-terminal truncations, as one means of modulating or regulating chemokine activity. [Berahovich et al., 2005; Proost et al., 2006; Mortier et al., 2012; Moelants et al., 2013]. Similarly the N-loop region that follows the first two conserved cysteine residues, connecting the N-terminus to the first of the three β -sheets, is the predominant region responsible for receptor specificity and high affinity binding [Hemmerich et al., 1999; Mayer & Stone, 2001; Lowman et al., 1996; Schraufstatter et al., 1993; Hammond et al., 1996; Williams et al., 1996; Sarmiento et al., 2011; Campanella et al., 2003].

The use of extracellular region mutant receptors, receptor chimeras and extracellular region peptides has been heavily employed to investigate the ligand interacting sites on chemokine receptors. From these, the receptors N-terminal domain has generally emerged as a strong candidate for the determination of the affinity and specificity of the chemokine binding interaction [Pease et al., 1998; Monteclaro & Charo, 1997; Ye et al., 2000; Blanpain et al., 1999; Duma et al., 2007; Love et al., 2012; LaRosa et al., 1992; Hebert et al., 1993; Clubb et al., 1994; Wu et al., 1996; Skelton et al., 1999; Rajagopalan & Rajarathnam, 2004; Prado et al., 2007; Park et al., 2011; Colvin et al., 2006; Gozansky et al., 2005; Veldkamp et al., 2006; Veldkamp et al., 2008; Mizoue et al., 1999; Kokkoli et al., 2005; Chen et al., 2006]. While regions in the receptors three ECLs and the ECL-TM boundaries have generally been implicated in the interactions which result in receptor activation and consequent signal transduction [Pease et al., 1998; Zoffmann et al., 2002; Han et al., 1999; Blanpain et al., 2003; Schnur et al., 2013; Ai et al., 2004; Wu et al., 1996; Katancik et al., 2000; Brelot et al., 2000; Chen et al., 2006].

As such this led to the proposition of a “two-site” model for the interaction of chemokines with their receptors, figure 7.3. In such a model the core globular domain of the chemokine initially interacts with sites in the N-terminal domain of the chemokine receptor (site I) an interaction which determines the affinity and specificity of the binding. This initial interaction positions the chemokine in such a manner allowing for a second set of interactions to occur between the receptors ECLs and the chemokines N-terminus (site II) and it is this interaction which was proposed to result in receptor activation.

However, in many cases the ligand binding affinity, specificity and activation cannot be readily separated out into these two distinct sites [Rajagopalan & Rajarathnam, 2006; Chevigne et al., 2011]. In addition N-termini-N-loop derived peptides frequently display reduced activity when compared to the full length wild type counterparts, suggesting other regions of the chemokine must also be participating in the interactions that maintain wild-type activity [Fernandez & Lolis, 2002; Chevigne et al., 2011]. As such it seems that while still valid, this simplistic “two-site” model cannot fully explain the chemokine-receptor interactions. Regardless, from the observations it is clear that the N-terminus of chemokine receptors represents an important interacting partner with the chemokine ligand, playing a role in shaping the affinity, specificity and/or receptor activation potential of the chemokine-receptor interaction.

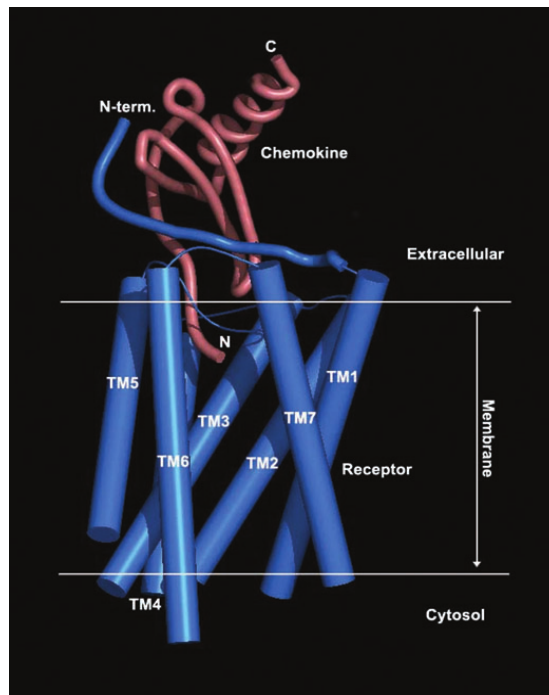


Figure 7.3 Hypothetical model of the chemokine-receptor interaction [Figure from Allen et al., 2007].

7.1.5 Chemokine Receptor Structure and Activation

While high resolution structures of soluble proteins, such as chemokines, have been comparatively easy to obtain, those of membrane proteins, such as the chemokine receptor GPCRs, has posed a much greater challenge. Purification of the proteins from the membrane in their native state and crystallisation, especially of flexible GPCRs, has posed a great deal of problems. However, recent advances in a number of methodologies required to obtain high resolution structures of membrane proteins such as membrane extraction, protein stabilisation, and obtaining structures without crystals, have allowed a number of GPCR structures to be obtained [Baker, 2010]. These include the high resolution structures of three chemokine receptors, CXCR4, CXCR1 and CCR5 [Wu et al., 2010; Park et al., 2012; Tan et al., 2013]. These structures have helped to provide huge insight into the mechanism of action of the members of the Rhodopsin-like family.

GPCRs possess 7 transmembrane (TM) hydrophobic α -helical domains, corresponding in size to the thickness of the plasma membrane. These insert into the membrane resulting in a protein passing through the membrane 7 times in an anti-clockwise manner to form an α -helical bundle with an extracellular ligand binding pocket and an intracellular signal transduction interface. Binding of a ligand to the extracellular surface induces and stabilises small conformational changes in the extracellular binding pocket. These conformational changes are propagated through the transmembrane portion of the receptor due to rotation and rearrangement of the TM helices of the receptor bundle relative to one another [Katrich et al., 2013]. These movements represent receptor activation and this modification of the tertiary structure results in larger conformational changes on the intracellular side of the receptor exposing previously masked binding sites on the intracellular side of the receptor. Interaction with signal transduction proteins at these intracellular binding sites, allows for relay of the external signal of ligand binding

across the membrane to proteins to initiate signalling cascades and bring about the relevant responses [Katrich et al., 2013]. GPCRs, therefore represent fluid structures which can adopt a number of different conformational states. The binding of an agonist shifts the conformation towards an active conformation. However, for some GPCRs, predominantly those, like chemokine receptors, which possess the ability to interact and be activated by a number of different ligands, there is often not just a single activation conformation. There is increasing recognition of so called 'functional selectivity' or 'biased agonism' in a number of GPCRs whereby the different agonists of a single GPCR can stabilise distinct active receptor conformations permitting more favourable interactions with different downstream effectors, thus allowing alternative downstream effects to be mediated through a single GPCR [Rajagopal et al., 2013]. In addition, there is also increasing evidence that the availability of downstream effectors in the cell can also play a role in determining ligand binding [Sprang, 2011]. As such, the notion of GPCRs being simple on/off switches is now largely dismissed.

The N-termini of the chemokine receptors have a number of common characteristics. In comparison to other GPCRs, chemokine receptors possess relatively short N-terminal regions (generally predicted N-terminal sequences are 25-60 amino acids in length) [Szpakowska et al., 2012]. The majority of chemokine receptors also contain conserved cysteine and proline residues in the N-terminal domain, the cysteine likely participates in a disulphide bond with another conserved cysteine found in ECL3 linking TM1 and TM7, and interaction which is aided by the presence of the proline and its conformational effects on protein structure [Rana & Baranski, 2010; Wu et al., 2010; Park et al., 2012]. This conserved cysteine has also been the basis for a division of the N-terminus into two functional domains. Firstly the 5' region from the starting methionine to the conserved cysteine and the 3' region from the conserved cysteine to the start of TM1. The 5' region is believed to be much more flexible than the 3', consistent with this being the region for which structural data has yet to be resolved. This 5' region also possesses an overall negative charge and is much more variable than the 3' region, both in sequence identity and length [Szpakowska et al., 2012]. Features which seem to be consistent with the notion of ligand binding and the selectivity of this ligand binding. Finally this 5' region also generally contains residues which has been shown or are predicted to be post-translationally modified. Such as be the presence of sulfated tyrosines [Farzan et al., 1999, 2002; Preobrazhensky et al., 2000; Fong et al., 2002; Colvin et al., 2006] and glycosylation sites [Ludwig et al., 2000; Preobrazhensky et al., 2000; Bannert et al., 2001], again features which have at least in part been linked to protein-protein interactions.

7.1.6 G proteins

The classically recognised means of GPCR signal transduction involves the activated receptor engaging and acting as a guanine nucleotide exchange factor for heterotrimeric G proteins. These proteins, as the name suggests, are formed from the association of three subunits designated $G\alpha$, $G\beta$ and $G\gamma$, of which there are a number of different types. To date, 16 $G\alpha$, 5 $G\beta$ and 12 $G\gamma$ genes have been identified in the human genome, many of which can be subject to alternative splicing and post-translational modification providing further diversity [Downes & Gautam, 1999]. Typically, G proteins are classified into four groups, $G\alpha_{i/o}$, $G\alpha_s$, $G\alpha_{q/11}$, $G\alpha_{12/13}$, based upon sequence homology of the $G\alpha$ subunit and abilities to regulate distinct classes of downstream regulators [Simon et al., 1991]. The majority of chemokine receptors have been found to mediate their signals through $G\alpha_{i/o}$ group proteins with regard to chemotactic responses, although signalling via alternative G

proteins has been suggested for some human receptors [Neptune & Bourne, 1997; New & Wong, 2003; Shi et al., 2007].

While the mechanics of the G protein-receptor interactions and the molecular basis for G protein selectivity remain hotly debated topics, there is a relatively well established 'classical' view of G protein signalling. In their inactive state G proteins exist as GDP associated trimer anchored to the intracellular face of the plasma membrane, placing them in close proximity to the GPCR. Receptor activation allows for interaction with the heterotrimeric G protein, the receptor acting as a guanine exchange factor induces conformational changes in the $G\alpha$ subunit causing an exchange of its bound GDP for a GTP which is present in much higher concentrations in the cytosol [Offermanns, 2003]. This binding of GTP represents initiation of the G protein activation cycle and it is believed to cause the $G\alpha$ subunit to dissociate from the $G\beta\gamma$ dimer, which remain as an undissociable complex at physiological conditions. The separated $G\alpha$ -GTP and $G\beta\gamma$ subunits are then free to modulate a variety of different cytosolic effectors, in addition to promoting increases in the intracellular concentration of Ca^{2+} and the opening or closing of ion channels [Wettschureck & Offermanns, 2005]. The $G\alpha$ subunit possesses a GTPase domain whose inherent activity leads to the hydrolysis of the bound GTP to GDP. This hydrolysis leads to an increase in the affinity of the $G\beta\gamma$ for the $G\alpha$ -GDP subunit, in turn leading to reassociation of the two subunits to reform the heterotrimer. The heterotrimeric G protein is then available for initiation of a new activation cycle [Offermanns, 2003]. However, as outlined below in section 7.1.7 a number of cytosolic accessory proteins are also intimately involved in control of the G protein activation, signalling and termination.

A broad array of downstream effectors have been found to be stimulated by these activated G protein subunits including PI3K, phospholipase $C\beta_2$ and β_3 , phospholipase A_3 , and phospholipase D; mitogen-activated protein kinases (MAPK); ion channels and tyrosine kinases [Bachelier et al., 2013]. With importance of these effectors varying in receptor and cell type dependent manners. In addition a multitude of cellular regulatory mechanisms are in place to ensure tight control of receptor signalling following activation.

In addition, G protein mediated stimulation of downstream effectors does not represent the only signalling consequence of chemokine receptor activation. More recently, there has also been an increased appreciation of G protein independent GPCR signalling, primarily mediated by β -arrestins. Initially these arrestin proteins were thought to act as signal terminators, discussed in section 7.1.7. However, through scaffolding of various signalling molecules both the stimulatory and inhibitory capacity of this β -arrestin mediated signalling is becoming apparent, adding a further level of complexity to chemokine receptor signalling [Musnier et al., 2010].

7.1.7 Regulation of Chemokine Receptor Signalling

Signalling from a chemokine receptor represents a transient event. While the GTPase activity of the G protein $G\alpha$ subunit is one of the major determinants of signal termination, the extent and duration of signalling is tightly regulated by a number of other mechanisms. These processes prevent both excessive stimulation of cells as well as insensitivity to ligands.

Firstly, the GTPase activity of the $G\alpha$ subunit can be modulated by members of the regulators of G-protein signalling (RGS) protein family. These proteins act to increase or

decrease the rate of GTPase activity of the G protein $G\alpha$ subunit, thus accelerating or inhibiting the termination of the G-protein signalling [Bowman et al., 1998; Ross & Wilkie, 2000]. Secondly, following ligand induced activation receptors are generally considered to be rapidly desensitised and internalised [Sorkin & von Zastrow, 2002]. However, this paradigm may not hold true as there is now increasing recognition that G protein dependent and independent GPCR signalling can occur from endosomal compartments [Irannejad et al., 2013]. Regardless, this internalisation downregulates the level of receptor surface expression, thereby attenuating further ligand induced receptor activation, in a manner that has been shown to be independent of G protein signalling [van Koppen & Jakobs, 2004]. Activated receptors are rapidly phosphorylated by GPCR kinases (GRKs) on intracellular domains, predominantly found in the C-terminal tail and ICL3 [Ritter & Hall, 2009]. Phosphorylation of the receptors permits their association with arrestin proteins, which play a variety of roles in the modulation of GPCR activity. The binding of arrestin proteins sterically inhibits the association of the receptor with G proteins [Vroon et al., 2006]. Arrestins themselves can also initiate signalling pathways distinct from those of the G proteins [Violin & Lefkowitz, 2007]. But their most well known role is to act as adapters, coupling the receptor to the machinery of the clathrin-mediated endocytosis pathway thus permitting endocytosis of the activated receptor to an endosomal compartment [Wolfe & Trejo 2007]. Following endocytosis the receptor can follow divergent pathways, the endosome may be targeted to lysosomes for receptor degradation or alternatively resensitisation of the receptor through ligand removal, dephosphorylation and recycling back to the cell surface [Kelly et al., 2008]. Thirdly, heterologous desensitisation may occur, whereby the downstream signalling responses to ligand stimulated activation at one receptor acts to desensitise other activated or unactivated receptors on the responding cell [Richardson et al., 1995; Kelly et al., 2008].

7.1.8 Chemokine Receptor Dimerisation

In addition to their association with downstream signalling and regulatory components in lipid rafts. It is becoming increasingly clear that chemokine receptors are also capable of associating with other chemokine receptors in cell membranes. Predominantly through the methods of co-immunoprecipitation, fluorescence resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) it has now been demonstrated that a number of CC (CCR2 and CCR5) [Percherancier et al., 2005; Hernanz-Falcon et al., 2004; Springael et al., 2006] and CXC (CXCR1, CXCR2, CXCR4) [Wilson et al., 2005; Hamatake et al., 2009; Levoye et al., 2009] chemokine receptors possess the ability to form homo- and hetero- dimers/oligomers. To date, the full significance of such dimerisation with regards to functional consequences remains uncertain, however, a number of theories have been proposed.

The detection of dimerised receptors in the ER/Golgi [Issafras et al., 2002; Singer et al., 2001] has led to suggestions that receptor dimerisation may play a role in the regulation of receptor trafficking to the cell surface. Support is added to this idea, through the finding that the co-expression of cytoplasmically retained mutant chemokine receptors with wild-type receptors leads to significant downregulation of the the wild-type receptor expression via retention in the ER [Wilson et al., 2005; Chelli & Alizon, 2001].

A large amount of evidence also appears to suggest heterodimerisation may represent a means of allosteric modulation, whereby the ligands of one receptor in the heterodimer, reduce or prevent binding of the ligands of the other receptor in the heterodimer. This has

been shown most extensively with CCR2 and its dimerisation with CCR5 and CXCR4, in which CCR5 and CXCR4 ligands acting at their respective receptors in the heterodimer with CCR2 actually inhibit the action of CCR2 ligands at CCR2 [El-Asmar et al., 2005; Springael et al., 2006; Sohy et al., 2007]. But similar results have also been demonstrated for some of the ACKRs, ACKR1 and its inhibition of CCR5 activation [Chakera et al., 2008] and ACKR3 and CXCR4 [Levoye et al., 2009]. Chemokine receptors have also been found to form heterodimers with other non chemokine receptor GPCRs, both CXCR2 and CXCR4 have been found to heterodimerise with opioid receptors again with the potential for allosteric modulation [Parenty et al., 2008; Pello et al., 2008]. Finally, while not explicitly demonstrated amongst the chemokine receptors, studies on other GPCRs have also implicated dimerisation in the modulation of both ligand binding affinities and the activation of downstream signalling pathways [Jordan & Devi, 1999; Terrillon & Bouvier, 2004]. While the true functional consequences of dimerisation have yet to be determined, it is clear that dimerisation is likely to play another important regulatory role in chemotactic responses, as well as adding yet further complexities to be overcome in the understanding of the chemokine system.

7.2 HHV-6A strain AJ Gene List

>gi|734986991|gb|KP257584.1| Human herpesvirus 6A isolate AJ, complete genome

>DR1

```
ATGCCGCTAACGGTGCCATATCGTCTTCCGCTCTCCAACCTATTGGTGGCTACTCTTGGGT
CGACATTTCCCTTCGTTCATGTTTCATTTCCCTACGTGCGTCTGCACAAGGGTCTACGCATTCCT
TTACCTTGGCCCCGAGCAGGAATGCCTACATTTACATCCTAAGCCTTACAAGTGTCTCCTG
CGTTACCCCTGTATAACAACACAACCCGCATCTTCTTCAGGGCTGGCCTACGAAGTCTTCT
CTATGGTTCGACCCATAACCCCTACCATCCTTCGGCCGACAGCAAGTTGCTACCGCTGGGC
CTGATCACGCTGTCCGCCGTGTTCCACGCGCGTGTGCGAGCCGACACACCCGGAGCGGCTTC
GACGTGACAGATCTGTCTCTCAGCTGGCTGACGGGATCGTCCCCTTGGCTCGTAATCCTA
CAAGGGCAGGGGGGGTCTCTGTTCTGCCACGACGTGCTGCAAGGCCGACTCTATGTCCTG
TCGCACCTCCGTGTGCTCTTTCTAAAGACGGGCCCTCCGCCACTGCGAGGCCATCTATCGC
GCACCGCTGTGGCGGACCGACCCCTGCCGAGCCTGTGGACGTGCCGGACCCCGACAAG
GCTTCTTACCGACATTACTGGCCAGAAGCGCCCGACGCGGCCTGGCCGCTTCTACGCC
CTGTGGAGACTGCATCTGGGATCGCGCTCGGAGCTCTCCACCCCGTGTGGAGTGGGAG
AGAACAGAGCTAGTCCTGACGGACTGGAGACGCGGGCGGCCGTGTACGCACCTCCCGTCC
GACTCGCGGTCCCAGCGCGCTACGTGAGTGGCGCCACGAACACGCGGGACGCGGGCACA
GAGGAGGCGGGGGGGGAAAAGGAGGCAGAGGGAGGCGGGCCAGCGTCGCGGAACAC
TTCGCGAGGTGCAGACCGCTGCTGGACGAACTGTGCGGAGAGGGCGGCTGGCTTCCCTTC
GCGTTTCTCTCGACATCTGCGCACGTCTGGCTGATCCTAACGGAAGGAGGCCCGTCTG
GCGGTGCGACCTGAACGACACCTCCGTGTGGCGCATCGCGGACGACCTGGAGCTGCTGCGT
CGCTGGGGAACCTGCTCCCTCGTCTCAGGCCTCCGGTTCCTCTCCATCTCCCGAGCGGG
AGCGGCGAGGCGGCCGAGAGCCGGGGTACGACCAAGAAGAGAGAAGAGGGAGAGCGTTCG
ACGGCGAGCGGACCGCCGCGACGTGACGCGCGGACCGACCCGTCCGACACGGGTGACG
CGTAAGGGGCGTGTGGCAACCGGAGGCGTCCATCTCCCCGCACGCCCCGAATCTGAGGAA
CAGACAGACGGCCACCACGGGCGCCAGGAGAGCAGCCACGACGACACGCGCGGGGAGC
GAACGAGGACACCGGATGACGGCGCGCACCGCCACGCGAATGAGAAAACAGAGCCCCAG
CAGCGCGGAGAGCACGAGGAAGGGAACAGACCAACTCCGGGCGCCACGAGCACGGACAG
GAGAGCCACGTGCAAGAAGAGACGAGGAGGGAACGAGACGAGGGCGGTAGCGGAAGGAGC
TGCGGGGAGCGACGACGACGTACGGCGGACAGAGCCGACATGATTCTGCCCCGTCGATC
CCCCGTGTCCTCCCGGGCCGATCCCCGCCTGTGGGTCCCGCCCGCTCATCTGTTATTC
CCTTCCCCCTCTGCCACCGATGACGCCCCGTGACGACGAGCCGTCCGTCCGCCCTCGGCGC
CCGCCGAGTCCCGCGGAGGAACCCCCACGTGTCGTCCCCGTCCGCCCCGCCCCGTGCTCC
GACACCCCGCTGTCCGCTGTCTCGCGCCCTCCGCTCCGCCCCGTCCCCCGCCGTCCTACTCTCCCTCCCCG
TGTCTCCGCCCTCTCCCGTCTCTCCCTCGTCTCCAGATCTCCGTTTCATCCCCCATC
AGATCTCCGGGACTCCGAGCGAAACCGCGAGTGTCTCCGGGCATCCCGCGACGTTCCCG
CCGGCGCCCTCGTCCGCGCCGGCCCGTTTCGAAACAGTCCCGTCCGTTCCCTCATCGGAA
TCTCCGTCCGCGCCGTGCGTCGGAAGTCGCAGCCGCCCCGCGCACACACGGCGTGA
```

>DR6

```
ATGTCGACGCGACACACAGAGAGGACGGGCGGACGCATCCCGGTCCGACGCGACAGGGCG
CGACTGGCCCAGGCGGAGCCCGTGTCTCGTTCGAGTGGCTTCTCCTGGCCCCGCGGACAG
CCGTCCAAACTCTACGGCTATGCGAGTGGCATCGCGGAGAACTGATCCACCTACCGTGG
CCGCCGTACTGGTGTCTAGAATCCACCACGATCCGTACAGAGACGCCCCGAAGTACCACC
GTGTGGGGCCACCGTGGGGTTGGCCGCCGACGCACGTGAGACCCAGATGCGTTCAAGAC
TGCGCTCTAGACTCGTGTCTGTACGTATGTTGCGGATACGGAGAGAACTTCAACCCGTG
GGTTTCGTAAAGTCGTATGTAACCAACTCCCAGCTCGACACGCTCCGCGTGTCTCCTGGTG
GGCAAAGACGGAGCCGTGTACGTCCACCACATGAGGGCGGCAAGACTCTGCCGACTGGCG
TCGAGCACAAACGGAGTTTACAAGACGAGGGCTGCAACGCGACGCCGTGACGTATGAAGAA
GACCTAGAGCTGCCGGACCAGCGTATGTGCGGAACGAACGCCCGACATCTGTTTCGACGTG
ATCGCCGCGGCCCGACGAACACAACCTGCTGACCGTCCGGCGCCTGTGTTCAGACGCAC
GCCGGAGTGTGCTGCAACTTACTAGAGACCGTGGGAGATCCGTGGACGGCGGTTCCGGGC
GCGCGCATGACTCTGACCGTCCCGCAGGTCCAGTACCGATTATGGCCCCGAAGCCCGGAGA
GACCTCCGCCGGCACCTGTACGCGGGACATCCCCCTGGGACCGTGGCTCGTGTGCGGCGTT
```

CTCTCGAGAGAGAGAGACGCAGAGGCCGTCGCCTCCCATATGCAGCGGCGGCGGCGACT
GTAGGAAACGTGGCCACGCCGGGGCCACGCGAGGTGGAGATCGTTTGGGTGGTCTTGACT
TTGGCGGGACCGCTGTTAGCGTTCGGCCCCGATACCGGCAAGATCTGTCTGTCTGGCCAAC
TCGTTCTCCACTCTGTGGAGGATGGGACCGCGGGCCATGAGAGGACACTGGACGTATTTCG
GCTCCGGGTAGACATCTCCCCGGGGATGCGTGGCCACTGTGCGAACACGTGAGACCGCAG
GTGGGAAAGCTTCCAGGAAGAGAGCGTACCTGGATTAG

>U2

ATGTTTTGTGGCTCTCCGTTTCTCGGCATATCTTCTTGGAGTCTGGCGTCTGCGGCTCTG
TGTCCTCCTTCGTGCTCTTCTCTGCGGGACGTGACTTGCGGTGCATGCGGAGGTCCCG
GAGGTTAAATGGACGGCGTTCGTTTCGCACTCTCGTCACCCGGCCGCTGTGCGGCGGACGAC
GTGTGGGACTTCGTCTCTACTTTTGCACATTGCAGGTTGGCGCTCCCGTGGCCCCGTGGGG
GCCGAACTGAGGTTCGCCACCTCCGACATGCTCGGGATCACGCAGGCCGAGGTGGCCAAG
CTGTCTCGCCGATACGGGTGCTGTCCCGGTATGGATCTCACGGTCATCGGGGTCACTATC
TTCGCAAGTTTCCGCGCTGGTGTGGTTCGGAGAGTGCAGGGGAGATCTACGCGTTTAAAC
GGTGTGTTTCGACGACGCGCTGTATCGTCTCGCCGAGGACGCGTTCGGTCTCTGGAAACAC
GGCCTGCGTGCCTTCGAGCCCGTCTACGGTTCCAAGTGTCTCACCGAAACCGGCGCCTCT
TTCTTCGGCGGCATGAGTGGAGTTGACGACGCCCTGGCTTTTCGCCGTGTCTTTTGATAAG
GCTCTCGTGCCTTGCCCTTGCCCGCGCGGGGCTTTCTTTGAGTTCGCCGTACCCGGGAGG
CCGGAAGAGGTGGCGGCTCATACCCGGCGGGCGGCTCGCGGTGGTGTATCGGTAGATTC
TTCGGACGGGGCGTCACTTACCTCTGCTTAGACGACAGCGGGTGTCTATGGATCAGGAC
GGCAAGGTGTACGCGGCCAGTCTCGACGGGGGAGCCGTGGTGCCTCTCTCCGATAGTTTC
AGGGCGTTTCTCGCCATGGGTGTGAGGAAACTCTTCAAGAACCATCGATTCCCTCCGGGA
CATCTGTGGACCATGCAGCTGCCCGTACTTGCCTTACGCGCCTGTGATAGATCTGCCC
GCCGTCTATCAACTTCTCCGCACATGGTAGAACGAGGAATGCCCGTGGCGTCTTGCGGA
GCTTCTACCGTAGTTTCGTGGGGATTACGAGGAGACGGAGACGGAGGAGTGCACATCTGGT
GATTCGGTCTCTGTTAA

>U3

ATGAAACGACAAAGAACCGGCGTGTGCGATTTACTGGAGGCCTTCATGAGCCAGCGCAA
GATTCACGATGTGTGGGAATGCTGAGCTGCAGTCTCGGGAGAGGACCAAGACACGTCAA
CGTGTCTATAACGAGTCAAGAGAGATCTCGTAGATCTCTTGCCTGTCTATACGTGCCTG
AATAGCGTTCGAATGTTTACGTTTCGTAATTTTCGGCAGGCGAATTCCTTTGGCTTGGCCT
GAAGGATACCAGCTGGTTCATCAATAACTGTAGAGACGAGAATGAATATGCCGATGAAAAG
TTGGACGAACGCGCGTCTGGTGTGTTGCCCGGAACGCCTCGTCTACTGGGATACGTG
AAGAAAAGGGGGGGATCGGATCGGAGCCGTGTTACTGGTTTTTCGAAAGGTGACATAGTC
GTTCTGCTCGGCGGCGCCGGTCTGTTTTATGCGCACACGTGGTTATTGCCGCAGCAACTG
TGCCCGTTCGGTGCACCATTTGACGATTTTCTACGTAAAGGTCTGAAGAGGTTCTATTAC
GCACATCAGCTGGTGTCTATCGTGCAGTTTGTGGTGGAGGACCCGAGGTGGACGGCGTC
CGCAGTTGTGGGACGTGCTGTGTTTTTCGAGACCGGCACATAGGTCACTTTTCGCCCTG
AGGTGGCCAAAGAACGAGACTCTGCTCTTTCATCGGCGGGCGGATTCTTTTCGCCCTCGTC
ATGTGCGACGAGGCGGCGCAGGCTCGCGGAGATGTGTTTTCTTCGGCAGCTTCGGATAC
AAGTACTTTGGGGACGAAGGTTCGATTTCTCTCTACGTGACCGACGACGGCCAGATTTTC
GGATTCAACGACAACGACGAAGACGGACGGCCGCGCTTCGTAGCGGAAGATTTTCAGCAG
TTTAGACGGATAGGAGTCACTCAGTATTATAAGAGCTACGTTTTCCGTAGGGAGCAGCCC
GAATGGGCGCTTCTTCCCTACCTGTCTATCTATCGCGAATGTTTTATCAGAAAACCTGGAGG
CAGGCCGACGCTGATGTGCTGTTTTTAATAGCTAGGAACGACGAGCGCAAGAAATTTCCG
GACGCTGGTACTAGGAGTGCCGTACACGACGTGGTTTCGATGA

>U4

ATGGAGTTATTGGACCACGACATTTATAAGGGACCCGTCCGAGAACGGGTCACGTATAACC
ATTCCCAACCATCCCTATCTCTCGCTTACCCTTCATCATTCCAGGGAACCTTGATGTTGAT
CTGAAAGACATAACGGAGGAGATGATTATAGATAGCGGCACGTTGACCGCAGAGGATCTG
TTTATGACCCGCGGCTTACGATTTTGTGACGATAGTGTCTCTGGGCGGCTCTGGCGGAG
AAAGTTTCTGTTGAATTCAGAGGCGTGGAAAGACGACCATGGATCTCTCGGCTTTTATG
TCTCTCTTAGAACGGTTAACTTTTTGAGGACGAAACGTCAGTTTTTTATCGACTCCTGACC
AAGTGCACCGCGTGGCGCACTTCTCGATTCTGGAGTACATCGTCTGATGGGGAGAAGAGG
GATACTATATCGGCGCATTTTAAACCGTCTTTTGGAGTTGTTGGACTCCCTCTTCTGACG

TTCCATCATGTTGAGGAACAGGAGCGAGTCCAATACTCTGTTGACCCTGTTTCGATATTTTG
CCCAATCCCAAGGAGATCTATGGCGACGCGGCGTCTCCGACGAGTGTGCTCTTGAAACAC
TTTTTGTCCGATCACATCTTTGTGCGGTACGCTTCTTCTTATCGCTTTGTGTCTTCTGTG
CTCTGTAGTTGCGATCAGTGCCGGGCGTCTCTCTTTTCGGATCTATCTGGGTAAGAAAGAT
TCCGGGGCCGTGCACATATCCGATATTTTCAAAATTCGATCCGGCTGATCTCATTTTTAGGC
CAGCTGCACCTCGACGCGCGTGAGGCGGTGCGTCTCCGGGGCGGAGATCGACCGAGATCTC
GGCTCTTCTCTCATCTTAACCTCTATGAAAAGGCGGCATTTGCCATTCTTCATGATAAG
CAGTTGGGGAGAGGGCATTGTAACGTAACATCTTGAAGGTGACTGCAACATCGTTCTG
TGTCTGTTTCTACTGAGACGCGTCAAACAGCGCATCGTGAGTGATCTCACTGCCATTGCT
CGTTTCTTCGTAAAAGCGATCTCCGAAATGGAGCGTTGCATCGATACCGTTTTCTGGCCTC
CGCGGTGTCAGGATCAAACGTTAGTTGTCAGTTCTCAGTTTCGAGATTAAGATCCCATG
CGCGTCCCTAGGTTGTGTTTTTTCTTTCTTGAGATCGTGGCGATTCTCGTATCGGGTTAT
GACAAGGAACGGCATCGCGTTGTCGCTCTCTTAGAGCATCTGTGCCTTAAAAAGATCTGT
CTCGGCGACTTGTGTGCATCCATGCGACTCGCTCGCGAGATAAGGCACGACGTCCTGGAG
GGTTCCCTCAATACTCTGGAGTTATCTTACATGTGCAATCCGGTTAGGTTTTTTTCGTTTT
CACGATTGCAACCTTACGGGACCCATATGTGGAGTGGGGTGTATCCGAACGTCGTTCCCT
TACGCTGTGCGCGGCGGCGGCTCGTTTGACATCAGGCTGCAGGAGCACCGGAGACGTCAG
AAGCGAATAGTCTGACAGCTAGGCTTATAAGACGCGTGCAGCAGCGAGGGTTAGACATT
CGTGAAGCGCGCCGCTTCCTAAGCCGGCTTGTGTGACCTTTATCTGA

>U7

ATGGACCTCGAGGCGAAGGAGATTTCAAGGTGATATCCGAACGATAACGCTCTTGCCGAA
TTGTGCCGTTTTATGCGAATCTGCCGATCTCTCTCAGATCGAGAATTTTGTGCGCAGGTAT
AAGAACTGGTGTCTGTCTATTATCTGGCCTCGAACTGCTGGCTGAGATTCACTCCGCGT
AAAGATGTAGCTGGATATACTGAAAAGATGTTTGAAGATATGGATGATCACTATCAGGGT
TTTGTGGAAAACTGTGTCTACTGGGGACTATTCAGATTCTTACAGAGATGTTCCAATT
CTGATTGGTGCATCTAAAAGAATTTTCTTGACGATTTAGAGACGGATACTCTCCACTTT
GTGTGTGATAATTTGAGCAATTTGTGAGATACGGCGTCTGGGTACTAATATCATCACA
TGCGCGGAGCCTGTGTACCGTCAATGGCTTGTATGAAGGCCCCAGATTTCGAATCGCTTGAA
AATTAATGAATAATGATGTGTTAAGGAGTCCCTATACTCTAAATGTCCATGTGAAGTTG
AATAGGAAAGGCGTTTTGGGCATCAAGGCTATGAGGAAACATTACATTGCCATGTTGCGG
GAGTTTGTAGTGTGAGCTGGCCCGTTGTGCCAGTTTGGACGATGTGGAACGTTTCGTGTCTC
AACATCGGCCGAGATCTTAGATTGGATATGCCGGTTTTTAAGAGTTTACTCTAGGGACT
CGTGACAGTGTCTGGACAGCGACATGTGCTTGGCAAATCTAAAAGAACAGGAAGACTTA
GTAGAAAAAGTAGTAGTGTAGGTTACCTGAATGACCATGATTACGAGTCTAAATGTACA
AGACCTATTCTTTGTATTGGGAAATCTGGAAAAATATATTATTACGATTGGATAGACAAC
GTGCTGGTTAAGCTGGGCGATTGTCTACTGACTTTTTTTGCGCGTGGGGTTTGCCAGACTT
TTTGGCGACTGCGGTTATGAGAAGATTGGAAAAATCTCCATGCGCTTCGGACGGATGTGCG
ACGTTGGGCATGTGCGAAACTTATCAATCGTGCATGAGTCTCAAGATCGTACCGGTCTGC
AACGACGAGTTCGCCTTGTGCATCATGTGTGCCTACTTTGGATTTTGTGATGTGGACGTTCTG
TCTGCAGCTTACGGAGACGGGTTGGAAATCAGTTCTCCGGGCTTGGCGGTTGTATCGCC
TGGCCTCCCATGTATGCATTAACCCCTGGGTGAGTTTTATCACTTTTACACGCATAGGTGG
GTCAGTGCCTTATGACTGGAGCAGTCTGCTGGATGCTGATGAATTTATGTCCGCTGTTGGG
TATGCTCACCCGATATATCGTGATCCAAATCCTGACTACGATCCATACGTGATGCATACT
TCAACCGGAAAGACCATTTGCCGTGGACACGGTGACGGAGAAAGTGTACATTATAGCAGAA
TCATTGGTTCAAGTTCTTCAATATTGGACTTAGACAATTTCCGCCTTTTGGCCGAAGCGGAG
TTAGATCCAGAGCAGGAAAAGATGTGGTTCCGGAGAGACTAAATGCGGCCGGGAGGAATTT
ATCCTTTTACAGAGGAATCTCCCTGCTATGAAGGATTACGTGGCAAACATTCGGCGGAG
AGAATCAGGGTGGATGCTTTCCAGGACTTTGACTTTTTCTTTTGTTCCTTGTGCGGATATT
TATTATCTGACTGGACCTGGCATCTTAGAGAAAATCACAGAAAAGGATTATGCTATCATA
GGGACGTGTGCCCGTTCTCAGGCTGAACCGAATTGCAGAGCTGCTATCGTTATGGGCTCG
AATGTCACATTTATATCTACGTAGAAAACAGAATCAGCAAAGTCTCTAAAAGTTTGAGG
ACTTTCATACGTAGAGGATTTGATGAGTTGCTCTATAAAGAGAAGTATTCTGTTGGATTGG
AGTGACGACACCTTATTTTATATAAGTGACACTGAGACGGAGAATCTCAATAGGATGTTG
AATGGGGAGTTGCCCGTTCTCCGTAGTAAGCCTCGGCATATGTGCGTGCAGAAAGGACAGG
CTTGTGAAAGACCGTAGCAAAATTTCTGTTTGGCGTGCCTGATGAAGAGGATTCACCG
ACGGTTAAATTTATCAGAAAATTTCTCACGCCGTTATTCGTGCGTGCAGTGCCTGCTACC
AACAGTTTTGTTGTTCCCGTCTCCCGCGCCAGGTTGACGAACGGTCTGCAGGGGACCGCA

GCAGCTAGATTTGGCATTAAAGGACTACATCCCTCTTCGGACTGCTTGGTGTGGAACATA
CTGCTGGATTATGAGTATGAAACTTACAAGTATCCTTCCACCTACATCAGAGCCGATCAG
ATTGCGGACATGGTGAAGGCTTAAAAATTTATGGACAATTTTAATGAGAAGTGGCAGTGC
ATCACTAAACTGGCTTTTTCTCGGTCTCTATGCCGGAGCGTCTCTGTTTAACTTCACGTCC
AAACCAACGCTGGGCTACTGGTGTCTGTTATCTAAGTGAGTATGCCTCCATGTTATTATTT
CAGTTTGAAGCAAGTTAAAAGAGTTGACCAAGGAATCTACTCGGCAACTGGGAGGTTAT
AATCTTTGTCAATTGGGGTCAAGAATTGAAGGACTGCTTGGAAAACAAATCCGATGTTTTT
TTCAGGTACGATTTCTTCGAGCGTATCGAATCCTGTCTTATCGAACACTTTTATGCTCCTC
TGTGGTTGTGTGCGAGTGTAGGCGCATGTTTATCATGTATAACAAGAGAGGTAGGAAGTTT
GATTTTGGCCACAGTGTCCGAATCCAGTGTTCCTATGATAGGTTCAATTCGCCTTCCG
GCTTTTCTGCATCTTGGTGAACCCACGATGTGAGTCTGTGAGTCTCATAGCGAAGGAT
TTAGGGTTGAGCATGATCGAAGGTCAGATAGAGCTCAGTAGGCTGCCTATTTTCGCTCCAG
ATCTCTGTCACTCCGGATAAGAAGGCATTGCTAACCTTCTTAACTAATATTGTCTTCATT
GTCTTCGTTGTGAACACCCCTTACCCTGTATATAAATGCTGAGTTGGACATATATTACGAC
CTCTTACGGAGGAAGTGGGGAAATTTGTGTGTGGCCATGGAGGAAGAGATGAAGCTGGGG
CGAATGGGTGTCTGGGAGACCTGTGTTACTTTTTCTCCGATGAAACAGATGAAAGAAATT
GTTAGGTGTCCCGGGGAAAAGAGTCAGTTTATCTTAAAGTGTGGGAAGCTCTGAGAATC
GGTTTTTCCGTTCCTGCCACAAAGATTACGACGAGACTCCGTTTCATGGAAATGTTTTTT
ATGCATCATCTTCATGTAAAACGCTTTCACGAGGACAATGATCGTGACCTGGTGTCTTGT
GATAATCTCATTCGGGGTTTTTTTTATCGTTAATACCGATGGAGAGAATTTCTTACAGCGG
CTACAACGGGTAGTGTGCCCCGTCGCCGAAGATTATTTAACTAACACCAGGTGCATAAAC
GGACTATGGCTTTTTTTTTTTCAGTGGCTTGAATATTTTTGGGAGTGGCAATCATCGCGGT
TTTCAAATAAGTCCCGAGAAAGACGTCAGAGCTATCGCATATAAATTTGGGGTCGTTGGAC
GTCTCAGGGATGATTACAAGTATTATGAATACACACCACCGGATTGTCCCGGCGAACTG
AATGGATATGGTGGGGATGAGTGA

>U10

ATGGAATAGTCACATACAAGACGGTTTTAGCGAGATCTCCGACCGTCACCTGGTCATCG
GGATTTCGGACGAGCTATAGCCTCGATTCAAAAACGGAATCAAGATAATATCAGGAAACCC
CTACGCTTCTACAGCGGTCTCCTACACTGCCTTATAAAAACAATATGAGCATTGCCTCGTC
CCGCCAAATAAATCAATACGTTTTGACAAAAGGGAAGATTGAAGTCGCTGCCCTAATTCTA
GATCTCGGACACCAGGTACTTGGTCTGTCAGATACACGTACGACAACGTATCTACAGCTGG
ACTAGCATCACACTCCCTAACTCTTTACCCCCAAAGAAGTTTACTTTTTTAGTCGCCTCT
CCAGGAGAAGAAAACATCGCATTTAATCCAACATAAACAAGGGAGGTTGGATCTCGGGG
AGTTTTTTCGTATCCCGTTGATTACAGATCAAACTTTTTCGCTAACCGGAATGTCTGCGAAC
ATATTAATGGTGCCTTTCGTGCCGTACAAATACCCACTAACTATGCTCGCTTCATATCG
TCAATCGACCTGATGATACATAACGAACAATTTCCGGAACACGAATGCGGAGATATACAG
ATTTTAAAACAGCGAACTACCTCTACCTGGGAGTCATCAAAAATCTTACGTGGAAAAAA
AGCGTACCCGGCACAGGTGACAGCGCTCCACATAGAATCCTAAAAGCTTCGTTTCATAGGA
AGTTGGCCGGACACGTCCCTGCCTGACAGAGTGGCACTGCGTTTTCTTTAATAACACTAGG
TTTACAATACACTGCCACGAATTTCTCGATTAACATTGAAAATCTCGGTCTCGTCAAAAAC
AAAGAAAAAGTTTTTGGGACACTAGCCACCGTATGCTGCGAGCAAATCCCTTCTTTACTA
ACAACGGAAAATCTACCGAGATATTTAATAGTTTCAGTTTGAAGTGGTAACGCAAATCGAA
GAGCCCGAACCGTGTATTATTTCCAGTAACCCGAAATTATATTTTACAGGAGACGTACTA
AACGCAACGATACAAC TGCAACACAACCCAAGTTATTACGAGCTACTTGTACACGCACCA
TATGACATCCATTTCTATCCAAGCCGATGCCACATAGTAATTTCTACCTATCCGATATTT
ACAAAGCCCGATAAGCAAATCTCATATCCGGCTACCAGAACGAAGGATTCTTCGAAACC
CAGGTGATGCTGTGGGCCCGGGAACACCTTGCACATTACGCTGCGTTTCAATTTCTCCA
AATCTGATCCTGCCTCAAAGCACAAAAGACGTAATCGCAAAGCTGTCTGA

>U11

ATGGATCTGCAAAGACATCCGATTCCGTTTTGCGTGGCTAGATCGAGACAAAGTTGAGTAT
CTTACAAATTTTCTCAGCAATTTGGAAAGACTGGATAATGTAGATTTGCGAGAGCATCCC
CATGTGACTAATTTCTTGTGTCTGTGAGAGAGGGAGAAGATGTAGACGATTTAAAACATTG
TATAACCTACTAGTGTATTGGCTTATGTATCACTACGTCTTATCTAAAAGGAAGCCGGAT
TATGATGCTATATGGCAAGACATCACGAACTCCAAAGTGTGCTAAACGAGTACTTAAAA
TCCAAAGGTCTGAATAAAGGAATTTTTGAAAATATGTTTACGAAACAAAGAAAAGTTTGA
TCGCAATTCAGTGATTTTAAATCGCGCTTTACTGTGTTTAGGAACTCTATTAAGTGGGGT

AGCAATGTTGCGATCGATACTCCTTATGTAAATCTTACTGCAGAAGACAGCAGCGAGATA
GAAAATAATTTGCAAGATGCTGAAAACAACATGCTGTGGTATAACCGTCTATAACATAAAT
GACCCCTGGGACGAAAACGGTTACTTAATAACGAGTATTAATAAATTAATTTATCTCGGT
AAGTTATTTTTAGCGTTAACTCAGTCCCTGGTCAAAGCTAGAAAAGGTTGCTATGAGTCAA
ATTGTAATCACGCAAAATCATCTCTCGGGTCATTTGAGGAGGCACGACAATTTTAATATT
GTATATTCTCATAGGGTTTTGCAGACTCCTCTGACTGGTCAAAGAGTAGAGAGTTTTCTG
AAAATAATCACCAGTGATTATGATATTATCAAAAAGTAGTCTGGAATCACACAGCGCATCA
AAAGCATTTTCGATGCTGAGATTGGGCCTTATTCTTTAATGGATTTTCGTCCCTTTACAC
GGCGATATACCTTCAAATTTGACTTTACCTAGTATGTCTATAGATACAAAAGAAATCATCT
GTAGATCCGGCTCGTCTGAAAAAAGTAAATCCAGAAGTTTGGATACTTTCTTAAGAATG
CAGAGACAACCGAAATTTCTAGAGTTGGATAGCGTTGATAATGCCGGGGGAAAAATTTTA
ATAAAGGAAGCAACACTCGGGGGTGAAGAAGTTAAAGCAACAACGCCTGCTTCGTCTGTC
TCTTTAATGTCCGGAGTTGAGTCGCCGTCGTCTTTCACTTCTACCAACCTGGATCTGCCG
TTGTCGTCTTTCACTTCTACTAATCTGGATCTGCGAGATAAGTCGCACGGTATTTATAAAA
ATTGGCCCTTCGGGGATTTTAGATTTTAAATGTTAAATTTCCACCTAATGCGCAATTGAAT
ACGAACGGCGTGGATTTACTACAGGATAAAAACCTTCGATCGGGAGTCCAGTAGCGGTATT
ACCGATGTGGTAAATGGTTTAGCTAATCTCAATCTGCATCAGAATAAATCGGAAGTTTTG
CCACCGTGGAGCAGAAACACAGCGGCAATGCGGATTTTTTAGATCCGGTGCATCGCTTT
GTTCTGAGCAGACAGGGACACCTTCGTGTTGAATAATTCGACGTGGCGGGATCAGAA
GCGAAGCATACGACTTACAGCACGGAGACCGGCGTGTACCCCGTAATGTTTTTCTCATT
AAAGATTTGAGAGGCAAAGACGGTTTTTAGGAAACAGAAGCAGTCAGATATTCGAAAAGC
TTAACTAAGGAAAGAAATGATAAAGCTATAATGCACTCACGCGAGGTGACCGGAGATTCT
GGCGGTGCGACTGAAACTGTAGCTGCTCGGAATTCGCCGGCGTTGAGAAAAATTAAGCAA
GCGAATGGTTTTTTTTCCGGGTAAATAAGAAAAATGATCGTGACGTATTAAGAGGGGGG
AAAGGAAATAGCAAGGACTTCTTGCATTTTGGCGGCAATGCAAAAAAAGAAATGTGCG
GGAAAGTTAATGTCGATAAAGAAATGACGCGAAACGGGCAAGAGCCATCACGTAGTTTA
ATGGGAGATGCTAGAAATGCCGGAGATGAACAATATATTCAAGCGGGTCTCGGGCAGCGA
GTTAACAATCTTCTAAGTCAATTTACAAATCTGATTAGTTTAGGCGAGAAGGGCATCGAA
GACATTTTGCAGAATCAGCGCGGGACCGATTTAAAGTTGGCTACAGAAAAAAGTTCGGGA
CGCGAATCGGAGGAAGCTAACGTAGAAAAAATTTCTGAAGCTAGTAATCCTCGAGATATG
TTTAAAGATTTTAGGTTGCAAAACGATCTCGATTCCGTTTCAGTCCCGTTTAGGCTACCG
GATGCTGATTTGTCGCGAGTTAGATTCCGCGTCATTTAAGGACGCGTTAGACTTGAAG
CTTCCGGGTAACGGAGAACAAGAAATAGATCTCGCTCTTGAAAAAGTGAAGGCAGGCGAG
ACGGAACCTCAGATTTCAAAGTCCGTCAGGATGGAAGTTTTATTCTGCGCAATTAATG
AAGTTGAGACACCTGAAGAAAAAGATGATATAATTGAACAGATGGTTCTGAGGATACGT
CAAGACGGGAAACTGATGAAAACACCGTCTCTGGGCCGGGAGTCGCTGAGTCTTTGGAT
ATAGAAGCCAAAGGCGAGTCAGCGATCGCGTCTGTA

>U12

ATGGACACTGTCATTGAGCTGTCCAAACTTCTGTTCAAAGGCAATGCTTCCTGTACCTCG
ACACCCACGCTAAAGACAGCACGAATCATGGAGAGCGCAGTTACCGGAATCACGCTCACC
ACATCTATCCCATGATCATTATAGTGGTTACCACCATGATTCTCTACCACAGAGTCGCA
AAACATAACGCTACTTCATTTTACGTCATCACACTCTTTGCTAGCGATTTTGTGCTCATG
TGGTGTGCTTTTTTTATGACAGTGAACAGAAAGCAACTGTTCTCATTCAACCGTTTTTTC
TGTCAGTTAGTTTACTTCATCTATCATGCAGTCTGTTTCATATAGCATAAGCATGCTAGCT
ATAATCGCAACTATCAGGTACAAAACCTTACACCGGCGTAAAAAGACCGAAAGTAAACA
TCTAGCACGGGCGAAACATAGGAATTCCTACTAGCTTCTCAATGTGCGCGATTCCC
ACCGTCTCTTTTGTGAAGACTAACGGAATGAAAAAACGGGAAATGCGTCTGTTTATATA
TCCTCAAAGAAAGCTTACGAACTGTTTTCTAGCGGTGAAAATTGTCTTCAGCTTCATTTGG
GGAGTCCTTCCAACCATGGTGTTCAGCTTCTTCTACGTTATATTTTGCAAAGCTTTGCAC
GACGTCACCGAGAAGAAATACAAAAAGACTCTATTTTTTATTAGAATCCTTCTCCTGTCA
TTTTTACTCATAACAGATCCCCATATAGCTATCCTTATATGTGAACTCGCATTTCTCTAC
ATACCGCAGAACACCTGCTTTTGGTTAGCCCGTGTAGAGATACTGCAGTTAATCATTCTGA
TTAATGCCACAGGTACATTGTTTTCTTAACCGTTAGTCTACGCATTTCACTGGAGGCGAA
CTCAGAAACCGATTTACCGCCTGTTTTCAATCATTTTTTCCAAAACATTATGTAGCACA
CAGAAACGCAAGGAGTCAGACGTATCCGAACATGATCAGAACTCAAATCGAAAGCGTCA
GTGAAAAAACCACCCCTTTAA

>U13

ATGGCGCACGCTAAAAACGGGTACGACGAAAACTTTTAACTTCGACAGACGACCCCATATA
TTGTCCAACACTTTTACCATGCGCCCCGACAAGTAAGATTGCAGACGCTGAAATCATTTC
AGAGAACATGATTATATCGCAAGCAAAACACAGACGGATTCAAAAAAATATCGCCCCGTG
TCTGTGATTTTAGACAAAACCTGCCCTCTTTGAATTTTACGGTATAGGAGACAACAACGAA
AAAGCTATCGTCTACCCATATCGATCCAGATTTCTTATTATGTGACTCGGAACATAATTGC
ACCCTATCACCTTTCTTATAA

>U14

ATGGTCACACACGCGATGGAAGAACCAAAGACATTCAACATTCCAACCTTTCGTCTTAGAC
GAAAATTGTAACCTTCATACCCGAAGTTCTCTCACGCGCCAACGCTAAATATATCAAAGAT
GTACTCATCCAAGACTCCTACAATACGGTTTGTCTCGCAAACAGCTTTATTCCCATGACG
ATTCAAACCACAGAGCAAATTATGGTTATCGTCAAAAGTTTAAAGTTTTACGCTGCACGA
AATCTGTTTCGAAAAAGTACTCCGCCTCGGCGTACATATAAACAGATTTTATGCCGGAAAA
AAACAAGTCAAACATATGATTACAATGATGAAAAGTCTTTTCGACACCGAAGAAGCCATG
CGACGTCTCGACAGAGCTCTTATGGGGCTGTTTGTGATGCTCGTGGTAGTTTCATATATG
CCTCTCATAGCCTTGTCTTGCACGAAAACGGCCTTCCAGATTCCAAATTCGAAAAGCC
GTCAAATTAATAAACACAACCTTAAACTCCTTTTCATAATCGACCGGACGCAGATATCGAA
CAGTATGCACAAAAATTACGAGCATAACAATTACCTCTATAAAATACCCATTATACTTTG
AAAGAAGCTGTTGACGTTTATTCCGAAAATCTTAAAGATCTCTCTATTGGAGTTAATAAA
AAGCCACATTACTGTTTACTTCTTTCAGACGACGGGTATTTATCTCATATTTTAAACGAC
ATGCTATTCTTAACTTCTACTTGGAAATATGATTTATAACTCCAAAGTACAAATCAGGCGT
TTGACCCTTGGATCAAATACGAAATCAACTCCATATTAGAAGATTGCGTATTGGTGGGA
TTCCAACCTCCAGATCTAAAGGACACAATACTAGATCTAGCGGCACTCACATCAAATATG
AATCTCGTCGACCCAGATAAAGACCTTTTTTCTCACTACAAGCTCATACTGGAAAACTT
TTTGAATCTCTATTTTCGCGACTAAGGCAAACATTTGCATTTTACCAACCTTCATAAAG
AGTCACCTAATAGAGTTCGAGGATGTCTTAAAAAGAAGCAATGATGATGAAGACCTCAAT
TATCTGCTTCTGAAATCTAACGATTCTGATGATGAATACGACGAAGACAAACTACCAATA
CAAGTCGATCCAGGCAGAGTAGACAACGTGCTTACAGATTCTAACTTTTTTAAACGTAACC
GCAGATAATGCTTTCTCGTCTATAGCGATCATGCCAATTTCTTATGATAAGATCATAGAT
GTGGAAGAAAATGTAATTCAGTCCTAGAAGTTGAGATGCAGAGCCTCTCAGCAGTTGTC
TACGGAGCAGTTGCAAACAAATACGGCTTAAAGCCTACCACAAGTCATCAAAGGTTAAAT
CAAAATGAAGGACGGGCTCTTCTCGTGCATCCTCCAGCCACAGTACTTCTACCATCCCA
TACTCCCCCGCAAAGCGGTGCTCCACTCCAACCTTCTATCCTAAGGCAAAGAGCACCA
ATACGTTCCAACAGCAGATCATCATCGGTCTTTTTCTCACAAGACGACGATAATCGTAGT
CATTACTCTGACGAAACAAATATCAGTGATTATTCATATCCCATGGCAGATCTAGATTTA
GAAGACGAAGAACCATGGAGGACCATCCACACTCCCCTCAATCAGCATCATCTAACAAC
TCCATGTCTCGTCGAAGCCGAGCTCTACAAAACGAACAAAGAAGACGTACGCCACAATG
GCCCCACCATCTCCGGCAAGAGGACAAAATAATGCCCGCTCAGACGCAGAACCAGACGC
TCAACGGAACGAACGACGAAAGGTTATAA

>U15

ATGGATGTGTGGAAGCGTCAACGGCTTCAAGAATGCCGTGAATTGTGTCTTTGCCTGTA
TTAATGTCACGTGTCGAATATGTTGTCAAAAATCGAAATCGTATACGTTAAATATCTATTT
AAAATGGACTTTTATACATATGTATAGATTTCTTTTACCGGCTCTTACGTTGAGTATGACG
GTTACAAAATCCGTAGTTATTGAAATGTTATTTATTTTAAAAAGATGGGAAGATATTGAT
CAATTTTTTAGATTGAACATCCGGAAGTCAACGACTGTATGGTCGTAGCTCAGTTCACC
CATATTCCTATAAAAAGGAAGTTAATTGTGTTGTTGTATATGTTAACAGTAGACAAGAA
AACAGCTTTTCCCTCAATATGATATATGCGTTTTTATAGAGAAGAGTCACCTTAGACTTGGT
GATGATGAAGAAAAGAATGCTATTCGCTTTTTTTCTTACGTTGATGATTTACACCTAACG
CGGGATATTTTATTGGATATGATCTACAAGCTCAAGAATACGGAATCAATCAAACAATG
GAATTTTTATTATCGTACAATGAATTGGCTAGATAA

>U17

ATGGCAGACGAACGAACCGGCGATTCTGTTAAAACCTCGTTATGATATTGCACTGATGAAT
TTAAACGATATAAAAATTGCAGTTTTTTCGTGATTCTCTATCGACATACGTAGAACAGAAG
ACCGGCTTAACTATACAGTTTAAATTGGCCGAAAAGTCGATGCTTGGTGATTTTCGACACTG
TGTAATAATCCATTTTCCACAAAGTCTGCCGCAGAGTTGCAAGAAATGTGTTTCGTTATTA

CTTTGCTGCCCGGAAAGATTACAAC TACTTGGATATGTTTTCCGTCTGGGGAGAGGAGACT
CGTGATGTGTGCC TAACTAAAACGTTGGTGT TTTGCGGGAGAAGACGAAAAATTTTACGGT
CTCGATTTTGTGAATGAGACACTGTATCTGTTGGCGGAAACGACGGAACGATTTTGCAGTT
CTCGGATTGCGTTCGCTATGATCCGGTATATCGAGAAAAGGACATTTCGCTTTTTTAACGAAA
ATTGATGATGTTCTTCAATCGTTGATTGATGCGCAGGATAATTTATGGAAATTCGCTACT
ATCGTTTTATAAAAA TAGTGGGAGGCATTACATTATGAAGTCTTGCTTGGAAAATAACGAT
GGAGTATTTGTTTTATTTCTTACTAGGAAAAGAAGATTTTCTAGTAGGATGGAATGGAAT
TTCTTTTTCGGATGTC TACGAGAGTATGCCGTATCACGGGCAGGTTATCGGAAGTATTGGA
AAAACATTACTGTATCC TAAAACGATGTTTGTGATGATGGATTTATCGGGGGCAATTTAT
GGTATTGATACGATCGGAACGGGAATTGGATCGTGTGTTAAAATAGCAGACGATTTTGAA
TCTTTTTTAAGACAAGGGATTGTTAGGGGATATCGTTCGCTATAAGTTTTTCCATAGGAAC
ATTAGCACCGTCCAAGAAAATTTGCCGTTGTGTCCCATACTTCTCTAGGGCCCACATTC
TCGAATAATTACAATTATGATTTATCATCTGAGGATGACGAGTAA

>U18

ATGAAGAAGTTAACTATGGAGTCCTTGGTAGTTTATATCTTCGTAATGGGAGTCTGCTTT
ACATCGAATTTACATGCGAGCAAAAAATAGTCTTAATTCAGGAACATAAACTTAGATCT
ATATGCATCTCTACTTGTCTATGTGAATGGAGTGTTAGCTGGGAATTC TAGCTGTGTTTTCT
GTCAAAAC TTCATATCTGATCAACCTTGGCATGTTAACAAATGGCTTTAAAGACATGAGA
GTTGGAAATATTACATCTATCTCTGAAAAACGGCTTTTCTGCGTGTGATTATTAATTAT
TATTTTAGAGGAGTGATGCTTAGAGCTTTGATCGCTAAAAGGCTTCCAAATGCCGCTCAT
TTATCGAGTACTGTAAATGCTGGCTAGATGACCATAGCGCAGGAGGTGTGATGACTCTA
TTTTATGGAACGGAACGGATTGTTCTCAATTCAGCACTGAAATCAATGCGAGTCGATGG
ATAAGTGACGGGAAGGATGCCAATGGCACTTTGAACATTTTAAATGAGAGAGTGTCTTTG
GATAGCTATTTTCTGAGCAGGATCTGTCTCAGTTGTCTGCTGAAATTTACAAGAAAAAG
GTTGCGCATCCGAAAAATTTTTCATTAATCAAAAATGATACGAAGCCGAAAAAATTC TTA
AGAAATACATGGAGGTC TGCATGGAGTAATTTGGTATAAAATATAAGGAGATCGAAGCGTTT
TTGGATTTTTCAAGCGATTACGAGAATTTATCTGAAATTACGTATTCAATGTCCGGCTGCT
GGGTTGTTCTTTTTGGCCGGGGCGCATTTACGATGCTGTTATTGCTATGCTGTTTGTCA
ATGATTACGCGGAAACATATTTGTAAGATTTTAGGATATTTAA

>U19

ATGGCATTACCGGTGATGAAC TGGCGCGCATGTTGCAGTTTAAAGATAAGATGATTTCT
TCGGCGGGTTTCGCTCTGAAATTTGAGAAAGTTGTACAGGAGGCTATGGCTTCCGGAATA
GTTCTTCAACATATCACTTGTATTAAGTTCCGATTTGTGACAATTCGATATCTTGTCT
GACCGTCAGTTGAGATGTTTATTAATAAACGGGTTGTATCCTTTTGGAGGTAGGATGAGT
ATGTTTGGTGTAAACGGAAGAATGGGAAGGTGCGTCTGCGGCTCCAGAGAGACAGGTTGTT
TTTTTGTATCGTCAACTGGTCAAGTGC TTTGGATATGAAGATGGTGTGGTATTTTATTTG
TCACCTACGTTTTT CAGATTTTTGGACAAC TGCCATGGAGTTCTCGTGT CAGAACGCGATC
TTGAGTAATTTTATCGCACAGAAGTCCC GCGATGAATACAGTGATCAATTTCAAAAATAT
TTTACGAGAATGAGACATACCCCAATTTTTTTGACAGGTGATTGCCTAGACGGTTTCAG
AAGGTAGAGAGCGGTGATTGTGTGGAAGAAGATGCTCGTGCTTCCATGAGACCAATTCAG
TGTGACTCTTTTTGGTGT TAAAGGATACATTTTGTGCGGCCGACTGAAGAATTGCTGCAGCCA
TCTGCGAATAAAGATGTTGGGGCAAGGTGTGTATGGCTTTATCGTGTCAAGAAGACAAC
TCCGCTCGACATTGCACAATCTATGGACTGACCAAACTCGTGGTATAAAAAATTATGTTT
TCGAGGCATACACAGACAGATCGGTCTGAGGTGATGTGTAACGCTGCCACCCAAACCGGA
GACGTGGTGGATAAATCTT CAGAAACTCTGTTACTCGGTGAAAATTTGGTCCATCAGTCA
ATTTTGGAAACGGAAGTCAAAACGACCGCAAAAATACTTTTTGATGTCTCTGATCCTCGA
ATCGACTCGGTGTACGATACGACAGTGTTCGGTGTCTATGGCAACCGATGATGTAGGTAGT
GAAAATGTT CAGGGCGGTGCGTCTTTGGCGCAAGAAAAACCTTTAAAAGGTTATTGTATC
ATTGCTACTCCATCAGAATGCAAACCGAACATTCATTGGCTTAAAGAGCCCAGAAAATGAT
GTT CAGGAATCGGCAGCTGTGCTGAGGTAG

>U20

ATGATAACTGTCTTTTGTGCGATGCTTATTTCAATGCGTTTTCTTCGTTGCCGGCTAAACTT
TACATCAAGACTACTCTGGCTGAAGGAATCGGTAAACTTCAGACCGTTATTGGCATTGAT
AATGATATAGTTTTTGTCTTACGAGAGGCTCTATGAAGATTTGACATTGCTGAACCATACG
GTTGTTGCAGAGACGTTGTTT GACCTTACGGGGAAC TTGGAAGAAGGAAAAAATCTTACC

GTAGACCGTTTTCTTAGGACATGTGGTGATACGTGAATTTTCATAGACTGCACGCTGGGCTT
CAATATGTCTCTCTTCAAAATTTTTCTGTAAGCGAATTGGTTTTGCTTCGTTAACAATAAT
ACGCAGTTGTCTGGCAGCTACGTTTTTTTTAGCGGGTAATACAACCTATGTACAGATAGAT
CTGTTTAAATGAAAATCGTGGTTTTGTTACAGATCTTATAAATTTGAGCAGTTTTTTTACAA
AATCGGTCTTTGCACGTATTGTCGTTTTATGCAAGACGTTTTTGTGTAGAAGACATCTTA
AATTTCTATGGTAAGGTTGTTTTCGGAGACTCTAGACATCGTCTCCGCAAGTTTTTCAGC
AAGAGAGATACAGGTTTGCCTCGTTTTGTACAGCTCGGCGATATCGGCCATTGGAACGAAT
ATTCAATGGAGTCTTCAAAATCAGACAGTCTCAGATGATCATATGACTGACGATTTTCATT
AGAACCAGAAATGCCCGTCAACTTCTATACTCTTATGAAAGAGCATTGAGTAGAGCTTTG
AGCATGACACAACGTAATTTTTTCATGTGAAATTACGCATAAATTATTAGTTACCCCAGCT
CTCTTAAACAAGAGAAGATGCTTATAGTTTTAAAGTTTTGTGAATCCCGTGAAGCAATCC
GAAGACATGTTTCCACGTCATAATTTCCAGCACCGCACAGAAAACAATTTAATAAATTTG
CAGCTGCTGTGGATTTTTGCTGTGATACCAATTGCTGCTGGGTGTATGTTTGTATATATG
TTAACGCGGTATATTCTGTTTTTTGTTTCTGGGGGATGTTTATTGAATCCAAATAGAGTG
CTAAAAGGCGTAGGAGAAATGATGAAGTGCCATGGTGATTATGGAAGTGAATACTGC
AATTACGAAGCTGATGATTACATGGAACCCATAGCGTTCAAAGGTCCGTGACAATTCG
ATAGCGTTGTCTGCGGAAATAATTCGTTTTGATATAGAACGACAGTCTAAAATCTCACGA
AATTTTTAG

>U21

ATGATGATCTGCTTTGTTTTCTTATGCGTCTTGACGTTAGCCCGGGGTGAAATCTATCCT
CCCATTTGTCCGGCCCCGCGTAGGGAACGATGAAGCGGTTAGAGCTGGGGAGTTGCTG
TTAGAGATATCTGCATACAGAAATCAAAGAAGCGGTAAAGTGGAGTTGTGGGGCAGTGCG
GCGGTCAACAACCAAGTGTTTTTATGGAGGAATGGAACAGTCAAATTGATTATGATTTT
GGGAAGTTTTTGGTTTTTCGATGTTTTCAGGTTTTTAATAACGTACATAAACTATTGTTT
AATACCGTGAGTAGTGCAACTATGCATCTTGCAAGAAAGCGTGTTCAGAAATGTGGTCAT
GGAAAAATGACATTTATTAGCATTCAAGTCCAATGCTCTGTTAATAAAAAATCGATACGA
CTGAGTAGAATGAGTGAAGCAAGTCTCAAGAAACAAGTACTACGTGTTGCTTATTTTTCTT
GATCGGAACAATAACTCTTGGATTGCAGATAAGAATTTTTCAGGGTGAAGACAGAACGATG
TTACGTCCTTTGGACTGAACCTCTCGACATATCAACAGTACTTGATATCTTCGTGCAACAAC
GACGTTAAAATCCTATCTGAGTTGTACGGTGAATTTAGACAAATAGCTCTGCCTTACGAC
GAGAAGTTGAACCTAAATTTTTATGCCGTTATCAGATCTTCTTCAGAAAGATTGTTTCAGA
GCTGATGATTTGAAATGTTCTGTTCTCTAGGTGGCTGGGCGCAGAAGGGGAGTTTGCAGTT
TGTGAATATTCAGGTTGGGGTGTGACAGGCTGGGAAAGATCGAAATCTTTGCGGAGAAG
CCACTGACGTTTGATATGGCATGGAACCGGTAAAGATGAGATCGTCCGGTGCATATACG
TCTTTATTTTCGCGACGATGTGACGTGGGGACTTATTCATTGGATAAATGGGTAGGGGAT
AAATATTTCTGCATGTGCACGAACAAGGAGAGTGGCGATAATGTTATCGTTACGCTTCCG
GAGAAGAATGTTGAAAAGTCTATTCAGATTTATGACGAAGGTTCCGGCTATGTTATCATTT
GCAGAGATGACATCTATCATATTAACCTTATGTTTCATGGGTGCGGTGGCGGTTTCTGTG
GGTATCCTGGGTATTTCTGTTTTGTAGGATTGAAGGAAATAATACTTTTATCTTTGTT
AGTGTGACTACATGTGGCTTTTTTGTGATACGTTATTGACTACGGCGGTTAATTGTTTC
TTCAAGGCCGTACCTTTCTTCGACGTGAATTAATAAATTTAA

>U22

ATGGTGCCCTCAGGGTTGTAGTTTAGTATGGGTGAGTGCATTATACGTGTCCGGTGATAGCG
TCTTTGCATATTTAACAATGAAAATTCAGTTTTTATCGCAACACACAGTGAACCGGAG
TTAAGGCATTGGTTGATTTTTGTAAAAATGACTCAGAGAAACGGAAGTCTTGGTGGCGG
ATGGCTAGTGTGCCATTAATGCTTATTTGCAAGAGATATAGCCTTTTTGTTTAAACCT
CGATGCGTTATCGAGACCGCCATGGGGTCTAAGATATTATGTCGATACAATAAAAAATATT
GGTGTGTTGTTCTGTTGACAATGATACTAAGTGAATGTTTCCTTTCCGAGCGGTGTGCAG
TTACAATTACTAAATCAATCGGTGATGGAGTCAATTAGAACTAAAACGTATGTGGTGGAT
TACGCTAGAAAAACTACAGAGAGGGGTGATTGCTTCATTTCTGTAGCATTCTGCCGAAAA
GAGAGACGACGCTTTTTTATCCCGCTGTGAGCGCTTTGTTTATTACTGTATCTCGGTTTAC
CTTTTTGCTGTTGTGGTTCTTTGTTCTTGTGTTTGGATCCGTTGTTTAAATATG
TGGCGTAA

>U23

ATGTTGTTCTTGTCTTTTTTACTCGTGTGTCTTTGTGAAGAGGTGCGAATGCTCAACCTA

ATGACTACAGAAGTTTCGGCAACTGAATTTGCATCAATTGCGTCTAAACACATGGAAACG
GATGTATCGACATCCAGCGACTATTTAACTGGGAAATCCGAAACCACGTTTTTCGGCGAAT
TCTGAAACGTGGGGTAAAAATGTTACCGAAATATCTATAGACTCTGAGACATATTTAAAT
CAGTCTTTTATGGTGACCTCTACGTTAGCCGTGGAAACCACAGATCGATCCATTGGCAAT
AATGTAAATGTTACAAGTTCGTTTTCCGACGGTCAAGGGTGGTGAGACGCAGAACATTGAG
ACGTCCTTTCACAGTAATATCGGCGTCGACATTTTCCGATGTGTCTGAAAAGACGTCTCAG
GGTTTGCCGACGAAATCCACACCCGAAAAAGACGGTTCAGGCTCTTTGGGAAACTGATACT
GTTCAAGTCCTTGAGTTTACAGATACACATGAGGGGGATGAGGAGTATTTTAAAGATTTT
CTAAGTTCATTAGTAATTTGGATCGGCGGGATTAGTTTTATCGGGGCGTTCGTAATTCTG
ATAGTCATTTTGTGTAATTTGGTATAAGAAAGACAAGCAGAGGTCTTTGTTTTGGGATGAA
GAGAAAAACCGGATGTACAAATGAGGAGGGATGCAAAGACTTGCAGATGA

>U24

ATGGATCCCCCTCGGACGCCGCCGCGTCTATTTCAGAGGTGCTGATGATGGACGTCATG
TGTGGACAGGTTTCTCCGCACGTTAGCAATGATACATCGTTTTGTTGAATGCATACCGCCT
CCGCAGTCTCGGCCGGCGTGAATTTGTGGAACAATAGACGAAAGACGTTTTCTTTTTTA
GTTTTGACGGCTTGGCTATCGCAATGATTTTTGTTTTATTGTCTTTGTTCTTTACGTTTTT
CATGTGAATCGTCAAAGGCGATGA

>U24A

ATGTTGCTTATAAGCAAACTCTATTCCCATTCTAGAGGTGCCGATGAATTCTGGGAA
GGAACATACATTTCTCGTAGTTGCATTCGTGCTCGCATTTCTCGTTTACTCCGATTTTCTT
AGTAATTTAAGTCCGTTTTGGAGAAATTTTATCTTCGCCTTGTATTTCAACGTGA

>U25

ATGGGTTTTATCTTCAAGCAAAGACATCGATTTGCTGGTAAATTTTATCGAAAATCGCCAG
GGCGCCACGTTAGCTTTACCCCTGGCCCGAGGATTATCGATTGACTCTGCATAGTGTAGAA
AACATCCCAGAAATTTACGTTGAAGATGTTCAAAATTTGGGCGAAGACCTATATGTGTTGC
GGAGGCGAATTTGGTTGTGGTTGGGGTTATACATAAGGCAAAAACGAGAACCTGCAGGGGA
CCGATTGTCTTTCAGGGGGCGCGGGGGCACATCTATGTGTATAATGGTTTTTTTTGACAGA
AGTCTTTATCACGTGGCTTCTTCTTTTTCACGATTTGTTTTCTAATGGTTTTCGGTTTTTT
TATCCGATTTATGAGACGTGTGATTATGCACCTGGACAGTACTGTGGCGCTGGATATGATT
GCTCACAGTAAATCGTTTTCTGAGCTCTTACATTACCGAAATGAAAGAAAGAATGCTTTT
TTTACCTTGAAAACATATCCGTACAAGACCTTTGTTTCGTTTTGCAATCTGAGCATGACT
GAATTTTCTAGTAGACATTTGATTCAGTGGCGAAGGAACTGCAGACGAGTCTACTAGAT
GTAGTGTATTATTGTACAGCATAATTTTTTTCGGTGACTGGAGGGAGCTCGTAGTAGTGT
GACGGCCACGGTATGCTTTTTTTCGCTCGACCGTGAGGAGAGCCTTCTCTTCATAGCTCGC
AATATGTCCGATTTTTTTAAAAATTTGGATGTCTGCGTTATAATGAGAATCGTCGGTTGCAT
ACGCAATGGTTTACGCAAGATACGGACTATATTAAGCAGGTGGATGAGATGTTTTCTAGA
GACGTCTTTTGTCTCTCCGTGAGCACTGTGAGAGGTGCGCTCGAGAGCGAGGATTATTA
AAGATATGTACGTCTCTAGTCAGGGGGTTAATTGTATTGAAAGAGGGTAG

>U26

ATGCGTCGTCTGACAGATAGTTTTATTCTCGGTCTTGCTAAGGGTGCCGTTATTCCGGGG
TTGTATACGTTTAGAATGACAGAGGGGAGGTCCCCCTCGGACAGATCGGCGTTTTGATA
ACCGTTGCGATTTCTTTTTTGTGACTTTTTAAAGATTTGATCCACGCTTTTATAAACCC
ATAGGTGATTTCAAGATTTGTGTTTCTGTCTTTAATGGCGCCGAAACTGCCATCATTATTG
TCTGCAGTGGTCATGATCTGCTTGATATTTTCCGAGATGAGGTTGAGAATGATTTTAAAGT
CGTTGTGTCATGATTATGCCGTCTTATTCCCAGCTGTGTTTACGGGAATCATGGTGTCT
TTGTTTTTTAAGAGTCAAATGTTTCGATGATTATTCTGTCTTGATAACGGCTGCGTCTCTC
CTGCCGATTACTGTGAGATATGGATGGATGATACGATCGTCTGGATTCTTCTTGGTTTG
CAAAAATATCGTCCGATTTTGAAGTCCACGTCGTTTTCGTGAAGTTGATTTGAAATGCCTT
GTAAAGTTTACGGTCAATTTTTGTTATTATTTACGATACTCTGGATCGGAAAAATGTTT
TTGAGTATGCCAAAATCTAACCATCTTTTTTCTGACTGTGGTCAACAACGTGTTTTTT
AACTAAACGTATTTCAAAGCCGCTGCGTGCGCGGTGGTGGCGATCTTATCGGGACTTATG
ATGAACGTCTGCCATATATCGAATTATCTTTGAAGCCTTTGTAGGTCTCGGGTTTAGTTG
ATTATGTTGACTTTGAGCAGCGATCTGAAAGACAGATCTTTCTATGCGGGTGATTTATTG
AACGTTTTTTTTTGTATTAGTCGTGTGTTGCATGTATTTGGTGTGTGA

>U27

ATGTGTTGGTCATTTCAATTTGTTTTTTAAAGCACATAAGGCGCGAGTTGGTGCTAGGACT
TCGTTCTTGACAGAGATGGAGCGCGGTAGTCGAGATCATCATCGTGATCACCGTGATCAT
CGGGAACATAGAGAAACGAGAGAACCGCCGACCCTAGCTTTTCATATGAAAAGCTGGAAA
ACTATTAATAAATCGCTTAAAGCGTTTGCAAACTGTTAAAAGAGAATACTACAGTGACT
TTCACACCGCAGCCGTCGATAATAATTCAGTCTGCTAAAAATCATCTCGTGCAAAAAGCTG
ACTATTCAGGCGGAATGTTTGTGTTTTGTCAGATACGGATCGTTTTTTAACGAAAACCATT
AATAATCATATCCACTGTTTAAAAGTTTTATGAACATCATCTCGAATCCTGAGGTTACG
AAGATGTACATTCAGCATGATAGTGATCTGTATACGAGGGTTTTGGTAACGGCTTCCGAT
ACATGTACACAGGCGTCGGTTCCCTGTGTGCACGGACAAGAAGTGGTGCGAGACAGCGGG
AGATCGCCGTTGAGGATTGACCTTGATCATTTCGACTGTTTCCGATGTGTTGAAATGGCTT
TCGCCGTAACATAAGACTAAACGCTCTGGTAAATCTGACGCTTTAATGGCGCACATTATA
GTACAGGTTAATCCTCCGACTATAAAATTCGTGACAGAGATGAATGAACTGGAGTTTTCG
AACAGCAATAAGGTCATATTTTACGATGTTAAGAGCATGCGATTTAATCTATCTGCCAAA
AATTTACAGCAGGCTTTAAGTATGTGTGCTGTAATCAAGACGTCGTGTAGTTTACGTACG
GTAGCCGCTAAGGACTGTAATTTGATATTAACCTTCCAAAAGCACGTTGTTGACGGTGGAA
GCTTTTTAACTCAGGAACAGCTCAAAGAGGAATCTCGCTTTGAACGAATGGGTAAACAA
GATGACGAAAAGGGGATAGGAGTCATAAGAACGAAGACGGAAGTGCCTTGGCATCAAAA
CAGGAAATGCAATATAAAAATAACCAACTACATGGTCCCCGCTAAAAACGGAAGTGCGGGC
TCTAGCTTATTTAACGAGAAGGAAGATAGCGAAAGTGACGATTCCATGCATTTTGACTAC
AGCTCTAATCCCAATCCCAAGAGGCAGAGATGCGTCTGTA

>U28

ATGAAGAGAAAAGAGCGTCGTATTAATAAAGATTATGGCTACAACAGAAAATGCGTTTGT
CATTACGAAGCTTCGCAAAAACGTTTTTGTATTCTCAATATTCATGCGCTTCTGTTTTG
TATGAACGGGTTTCGTGACATCGCTAAGATAATCGATCGTCTCGACTCAGGTTTAGATGCT
TGGTGTTCGCGGGATGCCATAATTTCTGTTTTGAGAGCTACACATTGCGTTCCACGAGTG
GATCGAATGCTAGGAAGATGGTATCTGAAGACGAGTGTTTTTTACGATTTTTTGTCCGGAC
GATTTAATTTTATCATGCCCGAATGTGATTATGCCAAATGTATTGAACTTTGTAAAGAAG
TATCGAGATTTTATTCGCAGCGTACTCTATAAAGTAAGCGTGAGTTGGAAAAACCAGTAT
ATGCCCGGAGTTCGGCCGCATCTCGCTTTTTTAGAAGAAATTTCCAATTTTAAACGGT
GTCGAAGAGAGTATTCATGCATCTATCTTCGTATGTGTGCTACTTTGACAGAGATCGTG
TTAAGAATTGGTTATCTGCGGGAGATATATCAGGAGAATCCTTACGTGATGTTTCAAGAA
CTGGCGTTTTTCCCTTTTTACTCAAAAATGGGTTTTTACCCTTTTCTGTATGACAAATTTG
GGCTCGTGGAAAAGGCTAATAGCACGGTTTTTACGCTGGCGATTTACAATACATGTTTA
TATTCGCTTGTGGATTTTACAACAGTAAACGGCGAACATCTTTTTTCCGGCTTTGTTAAAC
GGTTCATAATATAAGCATGAATTTGACTCGTTATCAACAGGAAGCCAAGAATATTTTTGAG
ATATTGCTGTCTCAAATTCGGGTGGTGGAAACGTGATACCGATAAGACGGTTCAGTTGACA
GTCTATGTGGAAGTCTGGCATGTGTCTGCACATAATGTGGCTGGACTTGTATGAAGCTTTG
CCTCAGACTTCTCGGGTCACATTTTGTGTTGATCATTCTGGGATTTTCATGGACAGGTAC
GAGCTAAAAAGGGCGCAGTGGTCTCTTTTTTCAGAAAAATATAGCTTTTGAATTGGGAAAG
TGTGACGAGATAACTTTCTCGACAAAATATTTGGAGTTTGGAGCGACAACAGATCATGCG
AAAATTACTATGTCTAGTTTTATAGAAAAAATTTGCCTTTGTTTAAAAGGGGGCAGGATG
GGACTTATTTTCCGAAAAACGTATATCAATATTCATGATTCCTCATGTGCCTTTGTAT
TGCGGGGGTGATTTTTTGGATGTTTTACCCGTCAGAGACGGTATTAACACTTGTGTTGAGA
ATGTTGCTTAATGTGGTACATTTCTTGGGTGACGAAGTAAGCGATGAGCTAACTGAGGAG
ATTGATTTCTGTCGTTTGCAGTGTAAATTTTTTATGTTTAAATGAATTAAGCGGGTGGTT
CGAAAAATGGTGTGGTTCGCTAATGCTGTGATTGATTATGCTGTAGAAAATAAAGATTTT
CTGTGTGAGGGAATCGAAGATGGACGAAGTCTAGGTATCTGTGTACGGGTCTTCATTCT
GTTTTCATGACTGTTCGGTTATCCTATGCACATCCAGATGCTCGTCGATTATATCGGATG
ATATGTGAACACATTTATTTATACCTTGTGTTTCAACGAGTGTGGATTGTTGCATGAAAGGA
GCAGAGCCATGTAATTTATTTGACAGGTCAAATACGCACTGGGGATGTTATACTTTGAT
CATTTTGATAATGTGCAATGCACGCTCCCTGAGGAATTGTGGACAACGCTGAGGAAAGAC
GTGCTGATGTACGGAGTTCGCAATATTCATTTACCGCCGGTACTGCCATGCAGAAAGAA
TTCGATATCATCAACTCTTCTGAATCGTTCTGGCCGATGGAAGACAATAAAATCTTAAGG
CGTCAAATATCAAAGTCTGTAATTTGGAAAAGACGGTTTTAATGATGTCACTTCAGTTTAT
TCGTCGGAATTTAAATCGTTATATATCCCCGTGTATAACAATCTGTTGCTCAACAGATTT

AACAAGCATCAGCAATATTTGAAGACTGTGGGATATAGGGTATTAACCGTGGACACGAAC
CTATTTACGGACAAGGAGTTGGATGATTTAGCCGTTTTCAAAGATGGTTTTAGTTATCAT
TTGAATGACTTGATTGAGATGTATAAGAGTGGATTGCCTTTTTTAGATCAGGGGCAGGCA
AATGTGTTTTATTTCAACGATACTGTGAGTTTTGAGACTGTTGCTACCGTTGCTGTATAAA
GCCGGATTTAAAGTCGCTATGTATAAGGTTTTGTGTAATTTCTGAGATGTATAAACACCTT
GATTTATCAAATCCGTTACCGTTGATCGGGAATGTTCCGGATGGTGTGGTTATGCACGTA
AGAACATTTTGTAG

>U29

ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAG
AGAAAATATATATCTATCGATGAGGGAACCTTAAATAATGTTGTAGAAAAAGAAAGGAAA
TTTCTGAAACAGTTTTTGTCCGGCAGGCAGAATTTACGTATAGCGGCAAGGGTTTTACT
CCTTGTGAATTACTGGCACCGGAGCTGGAATAATTTGGGAATGTTGATGTATCGTTTTGAG
ACTGATGTGGACAACCCATAAATTTCTGTTTTGTGGGTCTGTTTTCTTTGTTCTAACGCA
TTCAATGTGTCTACTTGTGTGAGAACCGCTTAACTGCCATGTATACAAACTCAATGGTT
GATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTT
GGTGTGACGAGCTTAGTTTTCTTCCGGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGA
AATGTCTATGACGTTAACAAGAAAACATTTACGGGCTCACGGTATTAAGAGATTATA
CTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTT
TACAAGGAGTTTTAGCAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTCAG
AACGAAGATCTGATAGTTGACATCTTGCACAGTATTTGAAAAGGAGCGGTTTTCTGTTT
TTGAACTATCTTATTAATAGCAATACCACGCTGAGTTACTTTGGAAGCGTTCAGCGTATT
GGTCGTTGCGCGACGGAGGACATTAATCGGGTTTTCTGCAGTATCGTGGAAATACCCCTG
TCAGTGATAAAGCTAGAGAACATATTCGTTGACCTTTCTGAGAAAAAGTGTGTTGTATAG

>U30

ATGACGCATGACAACAGACAACCTGCTGCCGCAAGAACTAAGCTCGTCACACCAAACAAA
GACACTTTGTCTTCCAGGTACCTACACGTATTGATCATCGATAACACGTTATCAACCATT
GAGTTTGTATACATGGCAGTTAAAGCGGTTCTCACACAAGTAGACACATTGAATGCGTTA
GAACAAAGGAAAAACAGACCCACAAACAGAATTTTAGGGTTGTCCACATCAGTCTCAAAA
CGATACATCAACATTTCCAAATTTTCCAGCTCCGGTGCCAGTAATTCACAAGGAGTGAAA
ACCCTTGCCGTATACGTAATTTCTGCCTGCCGGACAAAACTGTTTCAGAAATTTCTTT
TCTTTTTCTACAACATTTAAGGTTCCCTCATCGATAGATATATATTTTTCTTCTTCTC
ACAGTTCCACAGTATCCACAATGGCAGCTCGTGCACCTCGATTTAGAGTTCATCCTTAGT
TCACTCAAAAAAGCAGCTCTCCAGAATCATTACTAGCCGCTACGGCAAAAAATCGAACTT
TTATCCCTGGCAGCCGATAGCGTAACACACAACAGGGTAATAGCTTTTTATTAATCGACTA
ACCCGAAAAATATCACTTTGACCTCATCAGACAGTATGCGGTATTCTACTTTTTAAAC
TCTACGACCTTAACATCAGAAAATAAATTTGCTCTCCGCAGAACTTTTATACGAAGAACTC
AGCCAAATTTCCCTCTTCGAGTTCTGTCGGGCACAGAACTTGAAAACCTTAAATAATGCAGAG
GTCCTTTACGTGTTGATCGAATACTCAATTTCCATTAATAATGCTAAAAAATGAGTTATCG
TCCCAATCGGCAAGCTTCGAGCACAACCCGCCCAATGACCCAGGTACAGATAAAAACG
ATAAAATACCTCAAGATTCAGTCGATCATAAAAAGATAAACAGCTTTACACACGAAAAT
TCACTTCTTGCCAACCTGCCGCGCCGTGCTAGAGCTGATAGACGATCTATATCGGAACTT
TACTCATGGTTTTCTGCAGTTCTCACTTTTCAAGACATTCAAGTTTCCAGGCGACACATTT
CTAGACAGGCTGTTGAAAATGGACTACTGTTTTACATATTACCCTTCTTCAAATCGCCAT
CTAATAGATCTGTTGAAAAGGCGTTGGACAATCAAACGTTTACGAACCTCGATAAAATTT
TTTGACACCTCAGGAAACAGCCAGAATACTTTTATCAAAAAACGTTTAGCTTGAAAATA
TTTTCAAAAAATCTCACAGCCCAGGACAATGGATTATATATCTATCCACTACTTAAACC
GACCTGTCCATCTTCGACTTCTCAGGAACCGAAAATATTCTTCTTCCATCGAGGCCTCATC
TATCACATACTTCATCAAAAAACAATCCCTCAAGAACGGGAGAATGATCTCAATAAAATC
AATCAATTTTTTCGCTACTGTCTATCCAGCAGGTCATCGAAACGAAAAGCTCATGCTTACCC
GCATCGTTATCCCAACTGCTAGATAACAATTTTTCACTTCAACAGAATCGGTTTAAACATG
GAAACGTGCAGAACCACATCGAGATCCTCTCAAACCACATGGCAACGCCAGACACCCAA
CCCATAATCAATACATTCACCATTAACCTAACCATATTGTTTTTACCAGCGCACGTGTTT
TTTATCTGTATGAAAATTTTAGCCCAACATTCATATTCTATAATAGGAAAAAATCATA
CTAGAACAACAGAGAGCCATCTTGATCATCGAACGAAATGATTATTCCACACTCTGGAAG
CAAATATCTGATCACATCGACTGTCTCTTCAACGTCTCTCTCCTCGAATCCTTTTTTAAA
GAATATACCAAAGGAGGAAACGAAGATCAAAAAACAGTTTCTATACAAAAACCTATTTCGAA

AAATGGGGTAACGTATTTTTCCCTTTCACGTA CTCCGATCCACCTCAAACA ACTCTACG
GCATACCATATTACAACCC TTGAGCTAAGGGATATCTGCAAAGAGGTCTATCAGAGCGAT
TCCCCGGACGCATATGAATCTTTACTTCCCTACAGCACTCATCCATCTTTCAA AACATTA
TTCGTTAAAACTCTACGTTATACCTATGGTTTTACACATTACCAATCTTACTTTGATAAA
TTACAATCTGATTGCAGACTCTTAACCTAATCCACGCCTGCAA ACTCTTACTCCCCTCG
CAACATCTACTACTACACTACATGGCGTGGCTATATGCCTTCTCCATAAAACGTGGACCAT
ATAGATCTAGGCAC TTTACCGTAATAAAGTCCGTAATCTTTAAAAATCGCTGACCACATA
AACGTAATGACGCATACGATCTACTCCCCGGAGACAAACCTTCTCGTAAGCATTCTACTA
AACGCGTATACCAACTACCTCCAAAAATACGTA AACCTTGGATCAAACAAACCATAAACG
GCAAAC TTTCTCCCTGATACAGACGTACATAACTTTTACGAAACAATGCGCCAGCATCTTG
GCGACCAAATGCAACATAAACTTAGACAATCTATTTATCTACATAACTATCGGAACAGAC
AAAA TTGTAACCACTTCGTTTTGCTCCTTCATCGCTACATGCAGAAATTTAGTACGTCAA
CACGAAGAATTGAAAAATCACTACAAACGATCCAAATTTCCGAAACCACACTCACGGGG
ATGCTACGAAACATAATCACGAGCGTCTCCTCCTCCAAAGAGCTGCTAACCAACGAAGCA
CTACAAAAATTCATCGACACTGTCCAACGCATCTCCAAACACGTA AACGAAACATACCAG
TCTATTTCCATGAACCTCGAAAAATGTA AACCTCAAACGACATCTAATCGAATCCCTG
AAGAAGATCATCTACATAGTTGATGTACTCAGCTCCAATGCAATCTTAAACACGTGCTTA
GCTTCCAGATGTCTGGAAGCCGCCAACCTCGCGGTTTTGAAACAATTTCTTTACAATACTC
GAAATTA AAAAAGGACGCAGTTGCCGTTTTCAAGCCTTTTATAACACA ACTATTTGAAAGC
ATGAAACCCACCAGGTTTACATAAGAAATTTGATGTCTACCCAAAAACTAACCCACGAC
CACATTCATTTCTCGATACATTTGACGATAGGTACAACCTGGTCAGACACGTGCAACGA
CAATTAACCTGGTACGCTGCACATGCTGAGGCAGCGCAGCAAGATCTTATTACTCCTCTT
AAATTTTAA

>U31

ATGAAAATCATTACCAGTAGCACAAACCAA AATGACTCTAAATACGGACCGAGAGCCGGA
AAACAATGCATGTGCAACAGCTTTTTCATTTCTCCATACCGTTTACTTAAACGGAATTAAC
AATTCTCTCAATGCGGGTACTATAGACGCCATAATGGAAGAAGGCTATCACTTAGACACG
GCCGGCACCTTAGCCCTAATGTTAAATAACTCAGATTCGCAAGACTACAGACTCCCCACC
GAGATACCCAAGAGGATCCATTTCTAGATACGGTGTACACAGCATGAACTCTCACGGCCA
TTTAACGGAACCC TAGACACACAAAAAATTTGACAATGAGGTATATCTCGGACTCATAGAT
TTCATAC TGTATGGCAAAGCAAGAATGGTCTACTTTTGCAGTCATCACGATCGGCGTG
CTATCGCGAGCACTTTTTTTTTCTTAACAATACTCTCTACTTATTTCGACTCTCACCC TACT
GAACGAGAAGCCACGGCAGCCATCTACATCTGTCAAAACATTGAAGAAGTTTATGAACTG
CTCACCCACCACGGCACCGAAGGCTTCTACTACGACGCCAGCCTCATTTTTTTTCATAGAA
ACCTCCAATTTGTGCTCTCTAGTACGACGCCGAGCTTCTCATCTAAAAACATATAAAA
GACCCCGACATAGCGATTGCGCTAGATAAAATTTTCTTCCACGGAATTCATGAGATAAAA
AAAACAGATGATATCGGATCGCAACAAGACCTCGTTGCGGATAAAAACGACAGATCTAGAA
CACGCGCTCATAAAAAGAAAGAAAAACTCTCATAGTCTAGA ACTAGA ACTAAATGACAAA
AAGAAAAAAGATACCGCATCCTTGACATATTACGCAACAGAAGTTGACCTCATTCCGAGT
TTTTATGAACTACGATCACAATTTCAATCTTTATTTTACGATCTCAAATCTTTTCTATA
ATGAAATCTAATTTCAACTGGACCATATACTTACAAGATTCTCCTATTAATCCTAATCAG
CCTTTCGCAACACCC TTTCTCTGGAACAGAGTTTTTCCACTTATTGTGTCAAATTTATCGAC
GTGTTTGTAGGCGTCGGATCCACAAATGACGATTCTAGCAAACAACAACAACAACTATT
TTCATAAAT TATTTATTGCC TTTCAAGGACTTTTTCCGAAGTATTCAACGAGGCGTTGGCA
GCTTGTC AAGAAAACAATCTGGACATCATCTTCATTTACAATAATTACCTGTGCAAAACC
ACGAC TTTTCGAACACTCGAGAGAATCTTACTAAGTAAATTTCTGGCAATAGTTGATAAT
GATCACAAAAACAT TATGAATGGGTCAAATCGTGGACCACGCAAATGTTCCAAGGAATG
CCAAAAA ACTAGATGACATAGAAAAT TATCTAAAAGCCTACGTGGACCATAACCC TGTG
AAACACTTTTCATGAATTCATTTGTCTGAATAAGGCAGAAAAATACAAAGTAGCTGTTCTG
CTAAACGAAAAACGAAAAGAGATTCAAGAGGCTATTGAACGAGAAAAGAACAGTTTCGCA
CAACTCTCTAATTTTATTGACAAACTTGGCGAAACTCCCGCACTTCTATAGAATCAGAA
AACGTGCACAAAGTTACACGAGCGACATCACTGAAGGCATCGTTCCGCGTTTTCATAACC
GAATCGATTGAACTACCAAACATCTCCACGCTAAACAACACCCAACAAATATCCCTAGAT
AAACAAC TCAATGAAAACTCACC AACACCATTACACGCTCACAAACAAATTCACCAAA
ATCGTACAAGATAACTATAACAATATCGCTGCCGGTTTTCATGCCGTTACCGA ACTGAAT
TGCTATTTGCC TACCTAGTAAAC TTTATTTTAAACATTGAAGTTTTAAAACACAGTGGC
CTAAACATTAACACTGTGTTGCTTCAAGAGGTGGA AAAACTGTATGACAACACGCAATTT

CTACGCTTCGGAAC TTCTCACTTCAATATAAAACAACCTTTCAAAC TTTACCCTATCTATC
AGAAAAATGTTTCGTAGATTTCTATAACAGTCAAGAACCTTCGGACAGAGCTTCCGAAATC
CTAGCCGCAATCGAATCCATCCTAGCGGATCCCAGCAAAAACAAGACAGTCGTAAATATA
GAAATGATCAAATCGCAACTTGAAGAAGTGGGGAAAAATGGAAATCTCCACCACCGAAAAAC
AACAGACAGCTGCAATCACAAAACAATCCTCGGGGACCAAGAGTTAACCCCATTTTAC
GATTTCTGCACCACCTCTCTGCTTATAACCTACCTAATACTACAACCGTAAAAAATTTA
CACCTCCATTTTATCTTAGAACAACGACCAGATATCGCTATGACTCTACACGATAAAAATA
CAATCCATACTTGACATATATGTAGACGACATGCTCAATGACATAACGGTCCCTGAACAG
ACGTTTTCAACCGTTCTATTTCTTG TAGATCTCTTCCCAAACAGCACCGAAAAAACAGCA
CTGTTTGAATCCGTTCTCACGCTACGGCAATTAGCAAAGAAATGCGCTAACTTAAAAACT
CTAGATGAATTCGACGATCTAGCCCAATTCATCACGACGAATAGCGAACAACCTTCAAAAC
ATGATGAGACAGCATTTTGGAAAAAATAACCAACGCTGATGGGACATATCAAATTTCTC
TACTCTCAAAAAATAATCACCACCTGAAGAGAAAAACTGGATACAGAGAGCAAAAAACGGTC
GTCATCACATCTCCCGAAGAAGTACAGCCTTCTTAGCCACTGCACCCACTAAACACGCC
TTACAAACTTGCAAACCTGACCTCGATAAAGCACTACAACGACATATGGAAGAGCAAATG
AAACAAACTGCCGAGAAGCAGCAAAAAACACATCCTCACAAATCAGAAGCACCTTCGAAAAA
AGACTGAACGACATCTTACTTATTTCTAAAGGATGGTCAATTTTTCGTCTCTGGAAACAATG
CATTTGAATCTTTTAGAAACCTTTCTGAAACAACCTGCAAGATAACAATGTCATCATTAC
TTTACCCATGCAC TACTGCCAGTGTAAAAGACATAGAAACGACCATCAGCAAGATCATC
TCCGACGTTATTGAAAAGATATTAATCAAACTCCTCTAAACCCCGAACAGATGTCCAAG
GAAGAACAATAATACCCCTTTACTCAGTTTTCTTATCCAAGTTCAAAAAACAACATTC
TGCACAGAAGATATAAAAAACAGAAATCGACCAATGCAAAAAGAGTATCACGTTTCTCAAG
AAAATTGCTACCTCCGCCAATAAACATACTAGACTCAGCCACTCCATCTACGGACAAGAA
TTAAACCTCTACGAAGAAGCATTACAGAGCTGAGAAAAGAGACGAACAAAATGAAAGAA
CAGCTCTCCAAAGAATACGCATTAGCTGAAAAAAAATTTCTTCTCTCGTCACAAGATGCA
AAAAC TGACAAAATTTATCTAGTTCTAAATACGCACACGCTGAAAGAAATAAAAAACACT
CAATTCAGAGAAACAGCATTCGCTAAAGCTCTCACGGTGGAAAGTCAACAATAAAGAAAAT
CAACTCCAGGAATTGTTAAACCCTTTAACGCACACCTAAAGGCGAAAAATGGATCAGAAC
CATATAACAAAAC TTTTCAATTCGATACGAAATGGACAGCCTTCGTTTTCGGATTCAGGCTC
TACATCCCAGACTTCATTAACATCAAACCTGCAAGACTTTATTTCTGATCCCTTTAAAGTT
ATAAGTCAGCTAATGAACAAAGCAACCAACGAGATGCCTTACATTCAAGCAGAGATAACT
CTCAAATGGTTAACAAAATTAATACACGACATCAACAAATTTCTGCCTGTCAGCTATAAGC
GAATTCGGTAAAGAAGCAATCCCATTC AATTACGCAGCCCTACGAGATTTGGAGTACCAG
ATCAACACCAAATACGTAGAAATGAAAACAAAGTCATTTGCAACGAAACGGTCGAAAAAC
ACGAAAAACATTCAAAAC TGACTAAGCTCTTGAAGCAGTTAGATCCTAAACGCGTTGCC
GGTGGACAGGAACAGTATCAAAC TCTCATGAACAAGATTCTCAGTTCCGAAACTTCTATG
CAACAGACCTACGAAAAAGAACAAC TCAAAAAAGAATACTTTGAGATCGTCAATAATGTC
GCCAGCTTCAAGTTAGCATTC AACTTTTCTCAGCAACTACAAAATGTGGAACGGCTAATA
GAGAAGTTCAAATCGCTGCCAAAAGCCAGCCATTTGAAAAATTTCCACAGGAAAATGAT
CTACTTTCTGACTCACTCAACACAGAAAATTAATTAACGGTCTGCGCGCACTCCTCAAC
TTCATAACGGCAGCGCAAAAC TACATCCAGAACACACTTCTGAAACAGTGGGCGGTCTTT
CAGCAACAAAATTTTATCCCGATTGATTACTCAGTAGCGAATGTCAAACCTATTTCCGAT
TTATATGCCCGCCTCAGGATTGAACGGGAACGTCAAGTGTTCTACCAAGTAAACTCCGTC
TTTGGAACTCACCTAATCGTTGACGATACAGGAGTTCCGCTGCAATTCATAACATCTTT
AACAAACGCATAGTAAAGTTTTTCTCCCTGAACTACAAGCAGATT CATGTACCAGAAGAC
ACCCCCCGCTTGGTATCTAGCCAATATAAACTCCTGTCCGTATGTAAATCTTTCATAATG
ATACTACAGCAGTTCTGGGAGAATATAATCACCCCTGGACCTAGGCCCTTACCTCAGAGAC
GGTACACAAAATTTTAAAAGAGAAC TCATACCGATCGTGAACCTTGAATTTATTTATATAC
TGCATAACACAAGCCTGGACAGCCTCTGAAGACAGTACTGTTTCCACAGCGTTTCAATTA
CCCATCAAGCAATTCACACTTTTGTACTATGCTCACATCCTGAATACCTCTACGGCTGC
CTCAGTCATTCACAGATCTGGTTATAAATTCATTAGCCAAAAGCATAGACAAAAACAGT
CTCTATAACACATTCGTGCTTAGCCACAACCTCCCAGAAAGCCATGCATTTAATGAGA
AATATATGTATCGACACCCAGCTATGGCAACCCGCCAAACTTATGAAAGATACCTTCCAA
CAAACATTTTTTACACAAC TGTGCCCAAAAACGAAAAGTTTTTTTATCTATCTCACTGCC
TTTTTAATTTTACCTTACAAATTTATGAACTATATCTGGATTCAATACAAACCGGCCATT
TTTACCCAAAGATCTTATCAAACCTAATTAAGGACTTATGCTCCGAATACGTACACCAA
AATAAGATCACTACGTCTCGGTAAC TCCCACGAACAGACACAATAAAAATCCGGAGAA
AGGATCACCTCAA AAATCACCGTACACAAAGCACAAAACACACCGCACTCACTCGCCTG

CAGGCTCAGGAATATGTCTTCGACTACATTCTTTACTCTTTCTTGACAGGGTATGAAATG
ACATTTCGCTATGTACATCGATACCATCGAAAAACATACCTACTTTGCATGAGACACCTG
GAAAACGTACTCCACGATAAAGACTTCCAATCTGTCTCAGAGCACGCACATTTCGACATC
GACTACATCCTGAAACAGTCATGGACTAAAAACATAGTCGAACATTCACTCTTTTCAGTA
CAGCTCGATAAAATCGTCTCCTACCTCAACCACACAAAACAGAGCGACACCCAATATTCT
TTGATATTTAATTTATGACAACGAAGTTGTCAACGTTTATCTACCGCCGATGTCCAAC
AATCCAAAAAAGTCGCCTTTTACATCAAAAATCCTTTCCACTTTCCCGTGCAAGAATAC
GAAGCGACAAATTTAATTTCCCTTCCATCTCTACCCAAAAACAACAGATATTTTAAACCAA
CTTCCGCCAACAATACTGAGTCTACTCGCCCGGGTAAACAAAACCTCTGAAACTTTAACC
ACAAAAACCTCTCCGAACCAAAGTTTAAAAAGCCTGCCGTTACTGGACTTGTGCCCAAG
AGCCAATCGATCATCCTATCCACTGATACCAACGTACCGGAAACCAGTCCAGACGTAAAA
GCTAACACGGCCAGCGCAGCAATAAAAGATGTCACTCTAGCTCGAGAAAAAATCAACGAG
TTCTCCGAATCTATAAATAACAATATTTCAAATTTAAATCTATGTATCTGTAA

>U32

ATGACGACCATACGTAGTGATGATTTGTGCAATCAGATCACTCAGATTTTCAGGTTCTTCA
AAAAAGAAGAAGAAAAGAAAAGCAGCAGATGCTTACCGGTGTTTTGGGGCTGCAGCCT
AATATGGCCAGTCATCCCCTTCTTGGGGTTTTTTTTGCCAAAATATGCTAAGCAGAACGGC
GGTAACGTGGATAAAACGGCGTTCCGACTCGATTTGATAAGAATGCTGGCGCTGCATAGA
CTGAACACAAAAACAGGGAGTGATTGA

>U33

ATGTTCTACTTAAAAAACTCCTACGCCAGCTTCTGGCTCCGTTGTGTAAACATGGGCCG
TACACGCATTTGCAATTTGTTGTGATGGGTGAGGCCTGTGTGCCAGGAAAGTGTGTTCTT
ACAATGTTTTTGACAAATAAAAAATCTTGAATAAGGAAGTAACGGAAAAATTTTATAAT
GAGTTTTTCGCAATTTGGTTGCGTTGCAGGCCGGAGACGCGATTAATCACTAAGCGACTT
TTTAAACAAAATGGTTATGACAAAGGGTCTTTTTGTTTTGCTCGCCTATCTGTATTTCTGTG
TATCGGCAGTGTAAGTTCTCGAATTTGTTGTCTCTTTATAAACTTAAAAACAATCAAATGG
GTGGACGTCGAGACTAGATTCAGGGTTTTATCCCCGTACAAGCTTAATAAGTTGTTGGAG
ATGCCCTCGTTTTCCGAAATCAATGAGTTGCACATGTTCTGTTCGAGCAACAGTTGTTG
CTTCCGATTCCGACGCACGTGAACTTACCGTGTATGAGGCTATTCTGTTTACGTGACTAT
GAGCAGACCGAGACGGTGATGTTGCGATATAGGCAGAGAGAGCACGTTTTGTCAATTTCCC
TCTATGTTGCGAAAGTATGCTTTGAAATCTCCGGCGGGGAATTTTATGTTTACTATGGCA
AAGGCTTTGGTGGAAAACCTCTGTTTTAGTGCAGACAGATATCTTATTCCAGTGGAGCAT
AATAATCTAGTACCTATGGTGCCGAGCAAACCTGAACGGGGTGATTTTCCAAAGATTCTA
ACCTTTGCGTTGGCGACGTCTTAAAGGATGGCTTAGCTACCAGTGTTATTTTCGTTACCC
GTTATGTGCTATTTGTAAGACTAAATGTTCTCGCTTTTATACTCGAAGAGAGTTACATTTGC
GTTATTTGCGCAAAATGTGGTCACTGCTTAAACTCCGAAAAGAAAAGCTGTGTTCCGCCG
CAGGGCTTTTTCGCTGAGCTCCATGTTCTATTTTAGAGACAAGCAGGAGAAAAATTTGATA
TACAGCATGCACACGGATGTAATGTATTGTTCCGCTGTGTGGGAGTCAGCAGTTGGTGTTC
GAGCGGATATATGAGATGTCGAGCATTGTGTATTGGGGATGAAGGTGAAAACCGTTTCA
TGGAAGGCGGTGATCGGAACAAATTTGCTTGCCTTGCCTATTTTAAATGACAATGTGAAGTTC
GATGTTATTTGTTCCCTTGTTCGTGTAGATCCTGTTATTCTACTGTGCATCTCTACAACGTG
ACCGTCAAGAAGTTGTTACGGCTTGTTCACACGGTTCTGATTTTTCAGTGTGAGCGCTGT
CAGCATTCGTTTAGAGAGACTTGCCTTAGATCTGGAAGACTGTGTTAATATCTGTCAAGGT
TGCCAGATTTCCAAAACGTGAGGTGTATTTGA

>U34

ATGGCTAACGTACTCAAAGAAAAGATGTACGATGAACTACTGAGTGCGACCTGTTCGTATC
TTAAAGTTGGGGTCTCATGACTATAGGATAACTGAGCGGAATCTATTGTGCAAGAACCCG
AAGTTTCCTTTGTGTGATATCATCTCAAGTTAGACTACGCTTATAATCTAGAATATCTG
ATAAGTCTTTGGGAACATGTAACATAAGCAGGAACCCAGATTTGTGTTCAAGAACACAGGC
GGAGCTGTGTCGATGCTTGCCTATTTGCATGCCCCGGTTAAAGTAGAAGGTTCATCATGCT
GTGAGAGAATGCAACATCCTGCGTGTAAATGAATGTTTACTGTCCGCATGAGCGACATA
GTGGCTATGAAGCCTTCAACTTTTGCAGTGTTCACGAAATGTATAATCAGGCGTAATCGA
GACGACACTTACGTTGTTGAGTTTGTGGCGTTCGGACCTGAAAACGAATCTGAATACATT
TCGCTTTTGAAGCAATTTTCTTAAAGAAGTGTCTATGGGAAACAGCATCTCGAATCA
AATAGTTTTTGTCAAGTTTTGAGAAGAAGGTCTTACATGTTTTGAAAAGGGGCGATTTC

GAAAGTAGCGGGAAAGTTGTAATAAAGCTTCTGCGGTTGTGACATCACAGGAGAGTATA
AAGCAGTTCTATGAAAAAGAGAAGTCATTACTGTCCGGTGTGAAGTTTTGGAGACTAAGT
GAAAGACATTGTGCGTTTTGCGCTTGTGCGGCATCTGTTTTCTTTTAGCCTTGTATTTTTGT
TATGTTCTACTTAAAAAACTCCTACGCCAGCTTCTGGCTCCGTTGTGTAA

>U35

ATGATGGATTCTCACCTGAAAAGACGAATCTTGAACTTTTGTATGAGCGAGTCTGTGAG
CAGGGAAGGGAGTTTGAGGTCGTGTTTTATCCAATGTTGCCACGATTATATGAAATGATG
CTTCCTTCTTTGGAGGCGGACTGAATTTTTCTTAGTGTTGGATATAGACACGTTGCTTTT
GCTAGATACGTTACGGGGATGTGCGATTGCGTGCATAGAGAGGTCATGGCTCAAAAAATG
GTACTTCTTACTTCGATATTGTCAAAGCTGTTAAATGTAATGGAATATTAGAACATCAA
GAATATTTAAATACGGAATAG

>U36

ATGTACAAAGGATGGGACTCAGAAGCGTTGCCTATACAATCTGAGATACTAAACGAGATT
CTGTTATATTGCTACCTCACTCCTCAACCACCGATCCCTTCCGAGACGACAACAGCCACC
AGCCCCACAAACATAGAAAACGAAATATCCAATTTAGACAACCTTGAAAACCTAGAGGAA
TTAAAGCAACTATCCGTGCGATTAAATATAGACAGACGATGCAACATCTGCTCCATCGTA
AATCTGTGCC TCAAACAAAACAAATCATGGATCTACGACTACAGTCTACTGTGCTACAAA
TGCAACTACGCGCAAAAAC TCCACTGTCTCTTCTGATCGTTTTCCACAGAATTTTTAATG
TTAATACGGGAACGATTTCCAAACATCAATTTGACGGTCTTTTTCTAAACAACATTGTC
TCCATCTTTGACTTTCAGTTTCAATTTTTTATTACAGATGCTTCGCAAATACAGTCAAC
GATCACATCCAAAGCGAAAACATAACCCTGAATCACATGGCAATAATCCGATCGACCCTG
CTAAAAGAAGACACTATACCGCACATCAAAATAAAAAAATTTCTAACCAAAAAAATGAAC
CCGAAAAAACACAAAGCTCCGAAC TGAATAAAAAACTCACAGTGCCCATGAAGACTCGC
TTCACGACACTCTTATTCTATATGTGGTCTGGCAGAACGTCTTCGACCGAGTTCCCTTC
ACGGACCTAACTATCCGCAAACATCGCTTCATCAAAAACCTGTATTCCAATAAAACCGAC
ATCGAACTCACAGCTGGTCCCATTTCTTTGGCACAGATTCCATTTTCTATACCAAAAAAC
AAAACAACCAGCGTCTGTCTCCTATGTGAACTGATGACAGCTTCCAAACAAGATTACATT
TTTTTAAAATATCTACATCAGAGCATCATGGACTACTGTCAAAAACAACCTAAAGATGATA
GATAGAGTTCAGTTCGTTCATCGCCGATATCTTCGAGAAGACAAAAATACACACGCACGTC
AAAAATCTTTCCGACTACAGCAAGGCGATATTCGATAACGAATTTCTATTCTCAGATGAT
AACTTCACTCTCGATACCCACGTTTACCTGATACTCCGCCAGACTGGAACAGTGGGAGTC
TATAAGCACTTCTTTTGTGATCCCTGTGTCTAGCAAAC TGTAAAACCATCAACCCGGAG
GTTCTCTTCAACACTACAGACGCCGGAGAAATCCAAGACTTGAAAGTCACCATATGCTAC
CGTAACGAATATTTAAGCATCGTCGAAAAACATGTTTGGCTAGCGATACATCTTTTCAA
GCTTTTTCAGATTATTTAAACCAAAACATAAAAAATAAACTCAAATTACAGAGTTTCTCAA
GACTTCACTAACCTATTAGCACTTACCACCTTCGATATCATCGATCCCATTTTACCCTG
AACTACTACGTTTGA

>U37

ATGACTGTACACAAGAGCCGTGTTTCGACGATCGCGAAGCCTGTCCGTGACACATAGAATA
CAAAAACGACCTGATCACAGAGAAAAACAAAGCTTTACTTACAAC TGAACCTTCACGAT
CTACACACCGTCTTCAACATGTTTCCAGAATACGAACAAAAATTTCTAGCGATCATAAAA
CTGCCAATTACCGGTAAAGAACCATCGACGTCCATTTCAGTCTCAGCAATCACCAACAA
CACACCTGCC TAGAATTTCCCCCTACGCTAACGAACAAATCTCCAAAAGCGCCTGCCTA
CACTGCGAATCCGTCTCCGTTCCAACCTCCTCAGATGCAATGGTCGCCCACTTGAACCAA
GTCAACAACGTATGCAAAAACAGACTTTACTTCTACGGTTTTTCGAAAGGACATGGAGCTC
ATACGAATGTCCGCCAAACAACCCACTATCTTTCAAATCTTTTATATTGTCCACAACACG
ATCAACAACATATTTCCCTATCATGTTTCGAAAGAAAAACAAAGCTCGGAATGCACATAGTT
TTCCAGTCTCGCACTCTGCACATTCCTTGTGAATGCATTAACAATAAGTAGCGGTATCC
TCGGGCTACAACGTTTATCTAGATATACTACAAGAAAGTGAATCCTAACTGTCTCTGC
GAAACACTAGATACCAATAC TAATATCCACATCGACATAGGGATGCTTCAAAAAAACTC
GAAGAAATGGACATCCCAAACGAAATAAGCGACAGACTCGAAAAATACAAAGGACACCTA
ATAGGTTTCCATTAA

>U38

ATGGATTCCGGTGTCAATTTTTTAATCCATATTTGGAAGCGAATCGCTTAAAGAAAAAAGC

AGATCGAGTTACATTCGTATACTTCCCTCGCGGTATAATGCATGATGGTGC GGCGGGCTTA
ATAAAGGATGTTTGTGACTCTGAACCGCGTATGTTTTATCGAGATCGACAGTATTTACTG
AGCAAAGAAATGACCTGGCCGAGTTTGGACATAGCTCGGTCCAAGGATTATGATCATATG
AGGATGAAGTTTCACATATATGATGCTGTAGAAACGTTAATGTTTACGGATTTCGATCGAG
AATCTTCCTTTTTCAGTATAGACATTTTTGTGATTCTTTCGGGGACAGTGATTAGAATGTTT
GGGAGAACTGAGGACGGTGAGAAGATCTGCGTGAACGTGTTTGGGCAGGAGCAATATTTTC
TACTGCGAATGCGTCGACGGAAAAAGCCTGAAGGCTACTATAAAACAATTTGATGTTAACC
GGCGAGGTTAAAATGTCGTGTTCTTTTTGTAATTGAGCCGGCTGATAAGTTGTCGTTGTAT
GGGTACAATGCCAACACTGTCGTTAATCTGTTTAAAGTGAGTTTTGGAAATTTTTTATGTA
TCTCAACGTATTGGAAAGATTCTGCAGAATGAGGGATTTCGTAGTTTTATGAAATCGACGTA
GATGTTTTGACTCGTTTCTTCGTGATAATGGTTTTTTGAGTTTTCGGATGGTATAATGTA
AAAAATATATTCCTCAAGATATGGGAAAAGGGAGTAATCTTGAGGTGGAAATTAATTGT
CATGTCTCTGATTTAGTTTTCTTTGGAAGACGTTAATTGGCCCTTATATGGATGCTGGTCC
TTCGACATAGAGTGTGGGCCAAAATGGGAATTTCCCGGATGCCGAAAATTTAGGCGAT
ATAGTTATTCAGATTTCTGTAATTAGTTTTCGATACGGAAGGCGACCGTGATGAGCGACAT
CTGTTTACTCTGGGAACATGTGAAAAAATTGACGGCGTGCATATATATGAATTTGCGTCA
GAGTTTGAATTAATTTTGGGTTTTTTCATATTTTTAAGGATTGAATCTCCGGAGTTTATT
ACCGTTATAATATTAATAATTTTGATTTAAAATATTTGTGTATAAGGATGGATAAGATT
TACCATTATGATATTGGTTGTTTTCTAAACTGAAGAATGGAAAGATTGGAATCTCTGTC
CCTCACGAACAGTACAAGAAGGGTTCCCTTCAGGCGCAAACCAAGGTGTTTACTTCCGGA
GTGTTGTATCTGGATATGTATCCCGTCTATTCTAGTAAGATAACGGCGCAGAATTACAAA
CTGGATACTATTGCTAAGATCTGTCTCCAGCAAGAAAAGGAGCAGTTATCGTACAAGGAA
ATACCAAAGAAATTTATTAGTGACCCAGTGGCAGGGCTGTTGTGCGTAAGTATTGTCTG
CAGGACTCTGTCTTAGTTGTGCGTCTCTTTAAACAGATTAATTATCATTTTGAGGTTGCC
GAGGTCGCCAGATTGGCACACGTCACGGCTAGATGTGTGGTGTTCGAGGGTCAGCAGAAG
AAGATATTTCCCTGCATTCCTTACGGAAGCAAACGCCGTAATATGATTCCTCCGAGTATG
GTGCTTTCGCACAACAGACAAGGGATAGGTTACAAAGGGGCTACCGTTTTGGAGCCTAAG
ACGGGTTATTATGCTGTGCCACCGTGGTGTGTTGATTTTCAAAGTTTGTATCCGAGCATT
ATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATAGCTGGA
TTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGAATGAGACTCATCGGTTTGTGAAG
CCTTGATCCGTGAGTCTGTGCTTGGAAAGTCTACTAAAGGACTGGCTGGCCAAGAGACGA
GAAGTGAAGGCGGAGATGCAGAACTGTTTCGGATCCGATGATGAAACTTCTTCTGGATAAAA
AAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACGGGAGCGGCGCAC
GGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAAGAGAGATGCTT
TGTTCCACGGTGGATTATGTTAATTTCCAAGATGCAGTCCGAGCAATTTTTTTGCGAGGAA
TTCGGTTTAAACGTCATCAGATTTTACTGGTGATTTGGAAGTGGAGGTAATTTATGGTGAT
ACGGATAGCATCTTTATGTCTGTGAGAAATATGGTTAATCAGTCTCTGCGAAGGATTGCG
CCGATGATCGCCAAACATATCACAGATCGTCTGTTCAAGTCGCCTATCAAGCTCGAGTTT
GAAAAGATTTTATGTCCGTTAATTTTGATTTGTAAGAAAGATAACATTGGTAGACAGGAT
GATTCGCTTTTTAATTTTTAAGGGGGTAGATCTGGTGAGAAAGACTTCTTTCGATTTTGTG
AAGGTTGTGGTGAAGATATCGTGGACTTGTGTTTTTTGATGAAGAGGTTTCAGACTGCT
GCTGTGGAGTTTTCTCACATGACACAAACACAGTTGCGTGAGCAAGGAGTGCCTGTGGGT
GTTCATAAAATTTTGCCTGCTCTGTGCGAAGCGCGGGAGGAGCTTTTTTCAGAATCGGGCA
GACGTGAGACATTTAATGTTGTCTCTGTGCTTTCCAAAGAAATGGCTGCATATAAGCAA
CCGAATTTGGCTCACCTTAGCGTCATTAGAAGGTTGGCGCAGAGAAAGGAAGAAATCCG
AATGTAGGTGACCGAATTATGTACGTGTTAATAGCACCATCTATTGGCAATAAACAGACG
CATAACTATGAATTAGCAGAAGATCCCAACTATGTGATAGAACACAAGATTCCCTATACAT
GCGGAGAAGTATTTGATCAGATTATCAAGGCTGTGACTAATGCGATCTCACCCATTTTT
CCGAAAACCGATATAAAAAAAGAGAAGTTACTGTTGTATTTACTTCTATGAAAGTGTAT
TTGGATGAAACATTTTCTGCTATTGCAGAGGTAATGTGA

>U39

ATGAGCAAGATGGCAGTATTATTCCTGGCTGTCTTTTTGATGAATAGTGTTTTAAATGATA
TATTGTGATCCGGATCATTATATCAGAGCGGGCTATAATCACAAATATCCTTTTCGGATT
TGTTGATTTGCCAAAGGCACGGATTTAATGCGGTTTCGACAGAGATATTTTCGTGCTCGCCG
TATAAGTCTAATGCAAAGATGTCGGAGGGTTTTTTTCATCATTTACAAAACAAATATCGAG
ACCTACATTTTTCCAGTGAGAACGTATAAAAAAGAGTTGACGTTCCAAAGTAGTTACCGT
GATGTGGGTGTGGTTTTATTTCTGGATCGGACGGTGATGGGTTTGGCCATGCCGGTGTAC

GAAGCAAATTTAGTTAATTCATGCGCAGTGTTATTCGCCGTAGCGATGAAACGACCC
GATGGTACGGTGTGTTAGTGCCTTTCATGAGGATAATAATAAAAAACAATACTCTAAATTTA
TTTCCTCTGAATTTCAAATCTATAACTAATAAAAAGATTTATCACTACGAAAGAACCCTAC
TTTGCAAGGGGTCCTTTGTGGCTATATTTCTACATCAACGTCTCTCAATTGTATTGTGACG
GAGGCTACGGCTAAGGCGAAATACCCGTTTGTGTTACTTTGCTTTGACGACTGGCGAAATC
GTGGAAGGGTCTCCGTTCTTCAACGGTTCAAACGGTAAACATTTTGCAGAGCCGTTAGAA
AAATTGACGATCTTGGAAAACATACTATGATAGAAGATCTAATGAATGGTATGAATGGG
GCCACTACGTTAGTGAGGAAAATCGCTTTTCTCGAGAAAAGCGGATACTTTGTTTTCTTGG
GAAATCAAGGAAGAGAATGAATCGGTGTGTATGCTAAAGCACTGGACCACGGTGACTCAC
GGCTTCGAGCGGAGACGAATGAGACTTATCACTTTATCTCTAAGGAGTTGACAGCCGCT
TTCGTGCCCCCAAGGAGTCCTTAAATCTTACCGATCCGAAACAAACGTGTATTAAGGAT
GAATTTGAAAAATAATTAATGAAGTCTATATGTGATGATATAATGATACATATAGCATG
AATGGTAGTTATCAAATTTTTAAGACTACGGGAGATTTGATTTTGATTTGGCAGCCTCTT
GTGCAAAAATCTCTTATGTTTCTTGTGAGCAGGGTTTCGAAAAAATACGTAGGAGGCGAGAT
GTGGTGGATGTTAAGTCTAGACATGACATTCTCTATGTGCAATTACAGTATCTTTATGAT
ACTTTGAAAGATTATATCAATGATGCATTGGGGAATTTGGCAGAATCTTGGTGTCTCGAT
CAAAAGCGAACGATAACGATGTTGCACGAACCTTAGTAAGATTAGTCCGTGAGTATCGTG
TCTGAGGTTTACGGTCTCCGATATCTGCACAGTTGCATGGTGACGTGTTAGCTATCTCG
AAATGCATAGAGGTTAATCAGTCATCCGTTTCAGCTTCATAAGAGTATGCGGGTTGTTGAT
GCAAAGGGAGTAAGGAGTGAACGATGTGTTATAATCGGCCTTTGGTGACGTTTAGCTTT
GTGAACTCGACGCTGAGGTTGTCCCTGGGCAGCTAGGGTTAGATAATGAAATTTCTGTTG
GGTGATCATAGGACAGAGGAATGTGAAATACCTAGTACGAAGATCTTTTTATCTGGAAAT
CATGCGCACGTGTATACCGATTATACGCATACGAATTCGACGCCCATAGAAGACATTGAG
GTATTGGATGCTTTTTATTAGACTAAAAATCGATCCTCTTGAATGCGGATTTTAAAGTA
CTCGATTTATATTCGCCGACGAATGAGTAGAGCAAACGTTTTTCGATCTAGAGAATATT
CTTCGTGAATATAACTCATATAAGAGCGCACTATATACTATAGAAGCTAAAATTGCTACG
AATACACCTTCGTATGTTAATGGGATTAATTTCTTTTTTACAAGGGCTTGGGGCTATAGGC
ACTGGATTGGGCTCGGTTATAAGTGTACGGCAGGAGCGCTTGGGGATATTGTGGGTGGA
GTGGTGTCTTTTTCTAAAAATCCATTCGGGGGTGGTCTCATGTTGATTTTAGCGATAGTA
GTTGTTGTGATAATAATTTGTGGTTTTTCGTTAGACAAAGACATGTGCTCAGTAAGCCTATT
GACATGATGTTTTCTTATGCCACCAATCCGGTGACTACTGTGTCCAGTGTACGGGGACC
ACTGTGCTCAAGACACCTAGTGTAAAGATGTTGACGGGGGCACATCTGTTGCGGTTTCG
GAAAAGGAGGAGGATGAGTGCAGTGCAGTGGGCAAGTAAGTGCAGTGAATATTACAAA
GAAGATGCTTTAAAAATGCTCAAGGCCATAAAGTCTTAGACGAGTCTACAGAAGAAAA
CCTTCGTCTTCGAGTCTCATGCCCTCAAACCTAGTTTGATAGACAGGATCAGGTATAGA
GGTTATAAGAGTGTAAATGTAGAAGAAGCGTGA

>U40

ATGAATTCGTTACAATCTCTGTGTGTGTTGTGCGCTAGGCTTAATGAATGTGCTTTGGAC
CTTGAATGTTTTAAAGTTTTGTGATCCTGTAATAGTATTGTCTGATATGGCAAATTTCAAA
AAGAATGGGATTTGTGATTTCTGCATTTATATCAAACTTTTTTTTGAAGGTATCAAAGAACAA
AATCTTCTTTGTGCGTCAGCTTTGACGGTTTATATGCAGGTAAGTGTAAAAGCTATGTAT
GAACAGGTGCTTCTTTTGGATGCAGCGTTGGAAAGCTTTATGGTAGACCAGGATAGAAAA
AAATATTTTGAAGGTTGTTGTGTTTGTGAGACGTTGTGCTGAGCACTTGTCAATAAACATT
TCATTGAATAACGGTGTGAGTTTATCGTGCAATTATCTACGCTAAATGATATAGAACAA
TTAATCAGTAAGATCAACTCCGTATATGCTTTACTTTTTGCCCAAGAAGGGTTGCAGATT
TGCGGAAAAATTTATAGACTTGTAACTATCATGTGTGGTGCATGTATGGTTGCGAAACCG
GAGAGTTATCTCGAAACCAAGACTTGCATGAAGTGTACGAAGAATTGACTTTGACTCCA
AATCAAGGAAAATCACCTTCAAGAAGATTACATGGGAAGTTTTGTAATCATTTGACCGAA
CAAAAAGCTTTTTTTTTAATATTGAAAAAACATTGAGACGATAGAGAAAAGACTTAGGGGAG
GCTATTTTGAACATATGGTACTGTTTCAGTCTGTTGCGACTGAAATCAAAAAGATTTTTAAG
CAGCAGCGTTTCAGCCGAGTCTGTTGCATGTTAGTGATGCTGAGAAAACCTCTGAAAAAGTAT
AATATTTTTTAGCAAGGTACCTGACGTGATATATTCTTTAAGTGAAGTTTACGTATTGGTCT
AAAATATCGGAAACGATAGTGAGAAATGTGCGTATTACGCTTCAGCAGTTAAATAGTTGT
CATACGTTGTATAAGCAACTACAGAATGATGTTAGTCTTTATCTTTACGGTGAAGTGTCT
GAAGACTTCTTGGCTTTGAGTGAATAATTTGTTAACGCATGACGAGCGCTTGTACGTTGGC
TCCATTTATGTGTCTCCTTCGAGGCTGATAGATCTTGTCACTGGCCTGAGCATTAAAAAT
CTCGAGGAAAGTCCAATATTTAAAAGATTGGCTGAGGAAGATGAGGTTCAACATAAAAT

AAGAGTCTCTTGCATGACATTCGGGATCCACAGACGACAGAAACGCCGGGAAGGCTGAAC
ACGATCAATTGCATGCTTCAAACCTATAATCTTCAGCAGGAAGTCTTGGCCCCGAAAAAG
GCGTATTTTCAAAAGGTTTCTGAATCTGGTTATAATAGGGTGATGGCATGTATTCGAGAG
CAGGAATCGCTCATCAATAAGGTAGTGAGTGTGAATGTGTATGGGAATTTTATTTTGGAG
GCCTTATCGAAGATTATGAACGGTTTTCGTTCTCAGAAAAATGTACTTGGATGGCTCTCTC
CGTGTAGATAGCTGTACATACGATGAACATCTTTATATTA AAAACAATCTGATGCCGAAG
AAGCTCCCTTTGGAGCTCTTGGCGGATTTAAGTGAGATTATGTACACATTGTTAACGGGT
CCGTTATCGGATTTTCATAAATCTGCCTATCCACTTCCGGCGAATATCTCAATGGCGTAT
GTTGCGACCACGCGGAGATGTTACCGCATATGAAGGAAGATTTAGCTAGATGCATAGAG
GGCACTATTCATCCTAGTGTGTGGATGGTGTGCGAATACAATGAATTTTTTAACTTTTCC
GGTGTAACTGATGTAATGACATGCAGAAAAAATGTGGAATTTTATCAGAGAGTTAACA
CTTCTGTGTTGCACATATAATGATGTTTTTGGAAAGCGATTA AAAAATTTGTGCGTATAGAT
GAGGAGGGGATTTGAGTGGAAATGTCGTTTTGACATTTAATCACGAGTCTCCTTTATTA
TTTCACACTGGTGGTGGCATGACAAAGTTTTAAAGACGTGTATTTCGCTGTTATATTGTGAT
TTGCAAGCTCAGTTGTCGAGAGAGACTGTAGACGTGCCGGAAGGAGTAAGTTACTCTGTT
CGAACGCCAATCTACTAGATCTCGTTAGGGAAAAATGAGCAAGATGGCAGTATTATTCTC
GGCTGTCTTTTTTGATGAATAG

>U41

ATGGCTGATGAAAACGAAACGGTAGTTTTCTGCTCCCGTTTTCTACTGCTGCTTGGATCTAT
GTTTTCCAAAAGATAAAGAACTTTTGGATGTTTTATCTGTTTTATCGTTAATGGAGAGA
AATTCCTCTGTAGTTATTTCCCGTTGTTAATGAATTTGACTGTAGAAAACGATTTCTCG
ACCACGGTTAAAACACCTATCACTAATTTTTGGAGGCACAATTTCTAACTAAAATTACGTCT
TTTATGCCCGTGTGTTTTTTTTTTCATGAACTGAGCAGTTAGTTGGTATGGCCGAAGAT
CATGGTGATCTTATTCGATTATGTGAGCAAACCAGACAGAAATTTCACTTGCAGTCGTTT
GAGGTGCCACAGCAGCAAAGTAATCGATATCAAAGCTCTTTGTTCTGCTGTGCGCAAAA
GATGCAGATTCGTATTTGTGATGTTGCGTGTGAAACGGTTTTCAAAGAGTTGTTGTTT
GCTGGTCTGTTAATTCCTTGTGTTGAAGAACAATTTCAAGTACAAGTTGGGGAGTATTG
TGTGTGAAAATTCCTCTATATTCGCAACTTTATTTGAACTGAAGAAAACGATTTCTTTA
TCTTCTGTACAGAATTTATTCAGGAACGAGCATTTTTTCTGCTGCGTTAAGTGAGACG
CTTTTTTATTATGTTTTTACATCGTGGGGTACAACATTACGATTTTCTAATACTAAGGAA
TTAATCGATGCTGGCCTGAAACAATTTACACAGGATGGGGAGCAAACGTCAAATTAGCA
CCCCATAAACTTATCTGGGTATCTCCGGACAGAAAATTTAGCGGTCGAGAAAGATTTT
TTGATGTTGGTTGATTCCGTTGTGACAGAATTATCGTTTAGTCATGTAGCCGAATATCTG
GATTCGGTGTATGACCCTAGTCAGATAATGAATTTAATGACTGGCCAATCATAAGAAAC
TCAGAGACACACGCTGAACGCATGGCGCAGCTCACGAATTTAAAGTTACATCTGTCTAGT
CATCTTGCTGTTTTGATCTTTGCCCAAATTCATCTTATATTGCTCAAAAATGGCATT
ATACCAAACGTGAAACAAGCGTTTAATTCGGTGATGACTCAGGAGTTACTGCTGAGATCT
TTGTCTTTTTGTAAATGCGTTGTGCTCTTTGACTGAGGATGTGTATAACGATAATAGAAAA
ATTATCAAATGTGATTCGACATCAGGCAAAGATGACAAAATTTTCACTAATCACTTGGCT
TATGCCCTGTGCGACTTCTCCACAGCTGCTTTCTTATGTAGTGTGGAATTTAAATCGTATG
AGCGTTTATAACGCGGGTAATGCGCATACTGAGATTTACAATCATTTGGTAAATTTGTTG
GCGAATTTATGTGAATTCGTGATGGGAAATGTTGTCAATCGTGTATTGGAACCTGCCATG
GTGCGGGTCGGTAGTCGTTTGGCGGCAATACCAAAGAATGTTAAAAGGAACCCTTGTT
ATGAGCATGTTTTCTCGGTATTATGCCGAGGTTGATATATTAGGCTCTTTTGGCAGGAAG
CCTGTCAGCGAGCTCAAAGAGATTGGAAGAAGATCAACAAAACACGCTCTCACTGGACAGG
GGTAAATTCGTATCTCAGATCTTCGATTACTGTAAAAAATTTCTCTGATTGATCCTGTA
ACAGGCGAAGATACTTTCAATGTTGCAAGTAAAAAAGATTTTTGTTTCTATAATTCACGGA
TTAACACAGTGTATCGAGGAATGTGTTTCTAGATGCATTGTTGAAATGAGAAGGACACAA
ACGCCTAGAGAACAGATAGAAAATTTGTCTACAGTCATTTAATGTAGATACGACCCCTTAT
GCAACTGCCTTTTCTCCTTTTTTGGACGTTTTTATATAAGGTCATTTTGGACGTTTTG
CAGAACCCTCGATTGATAGTGGCTTCAAGGTCATGTTGTCGATAGACCTTGTACCGGTAAT
TCAATTTCTAAATGGCTCGTTTCAAGCAGTACCAGTCACTATACGGAACGTTTACAGCAGC
TATCTCAAAAAGGTTTTCTGAATACGAGAACAGTTAAAGTGCCTTCAAATGTTGATATG
GAGCAAATCTTAGATTGCGATCTATATAAATCGGGTAAATATGTGAAAACCTACGATTGAG
GCGAAACTTTGCCGTTTGTCTATGCAATGTCTTCGAGATTTTGAATAAAAAATAGACCG
TTCAACAAGTCTAGTAAGACAGCACACAACAATCCTTATTTTTAAAAAACGTGAAACAT
AAGAAGAATCCTTTGCTGGCTGTATCTCCTTTCTTTTTATTTAAATATCATGATAAATG

TTTCCGAATGTGAAGATTTCTGCTTAGAGCTATGGCAACGTTTTCTGCTCAACAATGTT
CCGAAAACTTTAGATATTGGTAACCCCGAAGAAGTTAAAACTTTTATTAAATTTGCATTC
AGTATTACGAATACCTATGATGAGATAGATATTATCGACATACAGCCAGAGTGTCTCTCA
ACTTTTATAGATTGTTATTTTCATAACAAATTTTTGTCTGGCTTTAGGTTTTTCATGACTAC
TTGACTAGCTTTACACGGTCTTACATCTAAGCTGGTTACGCAGAATCCTGTGTTGTTTCCC
GTGGTGTTAGACAAGCAGCCTAAATTTTTCTTCTATTTCAGGAATATTTAGTTTTATGTTAAA
AAGTTAGTTTTGGATGGCGTTCCGAATCCTGTGATTGCGTCACTGTCCAAAGAACCCAAT
TTTGGTACAATTTTCACAAGTCGATCTTTGGTTACATTTGGTTTTGACATTAGAGAAAATTT
GTGAGTTTGGCTAATAGGGAATATTTTCAGTTTGGCCAGTTAGGTTGGATAGGCGGAAGT
GGCGTGGATAGGAACTTAAACCCCTACCTCTTCTGCATTGCAGGATTTTAGATTTATGAGA
CAGAAAACGATTATTGCTACAAAGTTTTTCAGAAGTTATAGTGAAAAAGGTTCCGGCGTGAG
GCAATTTATGTTTGATACGGAAGTAGTTAAGGGCAAGGTTCTCAGCATTGTTGAGAATTTA
ACTAACGATATTGATCCAGAATTATTAATAATAGCTGAGGTGATGAGGGATAGGGAAGAC
AAGCCGACTATGGATGATATGCTTTTTTTTTGTGGATGGTAGGGAAGCCTTAGCGGCATCC
ATAATGCTTAAATTGAATCACCTTAGTTGATATGAATGTGAAAGACTTTTCAATAAGTAAT
CTGCAGTCGGTATTCGAGACGGTGTCTTCAAACGATGCTCCGGTCTATGACTTTTTCTGAA
ATTCTAGCGGAGGAGGATGATCAAGGAAACGGTGTATTAATAATGTGATGAAACGGAACCC
GAAACAGATGAGCCGATGACTAAGAAAAACCGATTATAA

>U42

ATGTATCCTCGTGGAGTAAAACGCTCACATCATGATTATCATAGACAAACCGCCTTTAGA
ACGATAAAACGTTCCACGCATAGACAAACCAGCAAGTTCATCAGCCATTTTCGCCAAAAAT
TTTCGAGGAAAACTCGCACCTCTGAAACAACCTGGATGAATCCAGACTAGACGCACTAAGT
CTGACAGAGTTAGAACAGCTGAAGACCATCATCGAAGAGAAAACAACAAGAAAAACGCGCT
CAAAACAATGCGATCACCTTTCTGCCTAATCTTCCAACAGTCCCCTTTGCCGACACAAAC
TTTTTCGCTAAAATCTCTCGGCCCTGAGACCTTACAACGGAGACGCCCGTGATCCAAAACAA
CGCATCCGAGACAGATTTCCACAGACTCACGAACGAATTTGCCTCCTGACTAATGATATA
CTGAAAACCGACCTGCTGCTTCGCTACCGACAATGTCTAGACTCGCTCACACGAGAAGAA
AATCAGCAACTGATGGGCGACCGGATATTCTCCCTGACCAACTCACCTTGTCTTGCTTTT
ACAGTTGCTACAGTAGAAGAAGCCTGCAGTTATTTTTAAATTTTACGATCTTACACAACCTG
CCGGTCAACCCACAAGACTTATTCATGTATACCATCACCGTCATGAAATTCGAATTTTTTC
AACAACTCAACATGGCAAAGTTAACCTGTGTTTTCAACGACAACGGACACGGTGACATC
GAATACCGAAAACTCAGGCAACTCTGCGGTAAACCTGTACTAGATCGAGAGATGCCAAAC
TCCGAGCTTGAAGTACAACAACAACCTCCAGACTCGTTTAGACACCCAATTAACAAGCC
ATGTCCATCGTTGTTACGTTTGTCTGAATACTACGACAAATTAAGAGCAAATCATACAA
ACAAAGAAACCGCAATTTATCCGCGATTTTCGACACGGGACGAGTAGCCGAGCGATACGAA
TGCGGTCTGATGAGCAGACTCATCGGAAAACAATTCAGTAACCATAAATGCGATGATGTC
AGCTGTCAAATAGAATCGAACGAATCATGGCTCCTTGAAAACCGAGTCTATTTTTTTTGC
ACTTATTTTCGCCAAAGACGCACCAAAGTTTTAAGCTATTTTCAAATTTTCTGAAGAATAC
CAAATCTTTCTTTTACTTGTCCAAAGGTGACACGGAACCTTCATGCAGCTATTCAACA
AATCAGACCTACCTCAAACCTCTCATCGTTTACATAAAAAACATGATACGCCTAAACTC
AAGTCCAAAGTTTGTGTGCAAAAACCGACACCTCAATCCTAACAACGACAAAAACAAC
ACCGAAATTCATAGAAGAATCAATGGAAAACCGACAACAATAACCAAATCCTAGAGAA
CTAAATTCATCAAGCAAAAACAAGAAGAAATTTGTGATTATTAACATTAATGAAAACGTC
AACTCTAAACATGAAAGTGAATCGAGTGTGAAATGGACTTAGATTTAGACTATGAAGCT
GACACCTGTGAAACTAACTTAAACGCATGTTCTTCTGACTCAGAATAA

>U43

ATGACCATCGTAGTGTGTTGCGACTGAATACGATGCCGCCAACGTAATTTTTTAGTATCCTA
TGTAGAAGTCCCAGTGAACATTTAATCTTTCCGATTATTGTTAAATATAAACCATCAAAT
AACGTTAGTTTCTGCCTTCAAACCTCAAAAATGCAAAAATTCAAAGCGAATCGACACGGTA
TTTGTCTGCCATGCCGAAAATTAATCTCAGCCACTACATACAAACTGCAAGTCCGATA
AAGGTGAAGATGTCGCCAACAGCCTCAACGACAAAGAAACCGAATCGTTATATGTCGAC
ATGATATTATCTCAAACCTGAAAAGAAAGAGAGGACGTGCAATTTAAATACATGGCCTAT
TTCCACAAATCGCTCATTATCAAATATCTCACCGTAAATTCCTATTGCCCACTTCCCCC
TTTTGGTTTTTTATCGACCTACGGGCAAACAGAAGGTATGCTTTTTATTAACAATGTTTTAC
TATTTGTTTCGAGGAACAAAAGAGTACCATCACGACTACCAAAAATTTATGTTCAATGCTTC
ACGAAAACACCGGAAACATGGTGTTCACGTATTTCTTCATGTCCGAATTTATAAACATC

ACTTTAAAATCAAAATTTAGAAAACATTTTGCAGACTTCTCAACGTACGCCAGACAAAAA
AATCTTCGCGATAAGGAAGAATTTAAACATCTAGACACTCAAATCAATCTATTTAGAAAAG
AGTTCACATCTAACCAACACATTCCTGTCCATTACATATATATCGCTTACAACACGGCC
CTAGAGACAACAAAATTCGTCAATTACTGCAATCTCACTTCTTACGACTCTAATTTACCA
ATAGGGCAACAATGTCAGAGAAATCTCCACATTTCTGGAAATTCCTGCACGAAAATTTA
CTGTGTATCATGAAGCAATACTTTAAATGCGGATTGTTACTTTAAAACATACATAGACATA
AAAAGATTAAAAAATCCAAATCTAAACGTTACAGAATACGAGTATGCTCTGGTCTCTAAA
AAGAAGACCATCCAAGCATTAACTTCCGAACAAAATAACGCGTGCCATAGCGAAATGTAAT
AAAAACGGAGAAGGTCTTTTTTCTCCAGTAAAACCAGGATTACAAGGATTATTAGAAATC
TCCGCTTCCGATAGATACGTTTCCAGATACAAGATAAGCGCATCTACAGAAGACAACACCTA
CACAAAGATTATCACCGTCCCTTTCTGTATTTCAGGGTGCAGCTTTTACATAAAAAATATT
TTTTGCTTCGGAAACTCTGAGGATTGGTACGAAAACATGGGATTTAACCGCATCTTACAG
TATCTCCCCGACGAATATATCTCTGATGAAGCTCTAACTCGCGCCATCTGGCTGCAAGAC
ACACATTTTCTCTGCGACGACGTCGAAAAACAATTTCTATACCACACGACACGAAATATTC
AACGAACGCATACCAGTCACAAATTACATAGGAGACTTGGATCTTCCGCTGCAAGATACA
GCAACTATTACGGAGGAAACATTTTTCTCGATGTGTGCGCTTATAAGACTTACCCTCATC
AACGCTTGGAAAAAAATCTTTCCAAGCATCGACACAGACACACATCCCATCTTCTTCTTC
AAAACCCAGTGCACACAACCTAACGACACACTAGACTATACTGAAGATCCAACCGAAATC
AAACAATCTCTGTGTTTGTGCGAAAAAGATCGGTCTTAGAATATCAATCCCATTACCTAAT
GGCACAGCGATCGCCGGTGGAGAACCATTAAAACAATTATCCAAGATTTTAAATCACGTT
ATGTGTTTAGATCAAGAGCTTAGTCAGATCCTCAATTCATAACATTTCCCGGGGAATGT
TTCGACATCGGCATTTACCACACAGGTCCTGATCAGAATAGGTTATATGTACAAGACA
GACATGGACAAAGGGAAAATGCTACATGGACGTTAACTCCCATATTTATCGTCCCAGCA
GGTTACCGAAATCTTGAAGACGTTTATTTCAGATGCAAATGGATTTAAACAACCTTCTA
CATCACGGTACTAAGAAGGCTCCCATCGAAGAATTAATCTATAACATAACGGATAAAGGC
TGCCCTAAAGAAAACCTAAGTTTCATGGATTTAAAATCCAGACAACCTGTGGAACAAAGTT
AACATAGCTACAGAGACTCTAATCACCAAGTACCTAAACACGCATGGTTTTAAACAACAAT
GCAACATCAGCGGATGACTCCCTTTTATCGTTTATCAGATTAATAGGATGGCCAGTCATT
AAAACGCAACTAATTACTCATACGAGACTAGGATAGCGCAACAGTTTTTACAAGTAACT
TTTATTAATAATAGATTCAAAGAACCTCCAGATCAAAAAACACAATTTGGAAGAGTTTCG
GATTTTAGCTGCCGTAACAGACAACATAGAGGAAATCGAGATAACGTCCTAGTTTACATA
CAACTCAAAGCCGACGGAATCGGTTAATTTCTCATCTTGTGGAGCACGTGCTTTGCTACT
AAATGCCAGTCCAACAGCAACAAGTACATTGTTCCATCGCCTTGGAGCAATTAAAAAAC
TAA

>U44

ATGGGAAATGTTTTCGACGATTATCTTTGTTTGGACAACCGACTCAGGCAAAGTATGAA
ATGTTGTGCTCTTCCGACGATTTAGAGACCGAAGAGCTGACGATTTTTCTGGATATGACG
TATAAAGATTTTGGAGTTTATCAGAATGATATTATAAGTCATCAAAAAGACACGGAAACT
ATGAAAACCTCTGTTAGGCTTATTACCAATGTACAAAAAACTAAGCTCAGGCATACGGTT
ATGGAGCGTTGTTTATCTAATTTGTCTAACCATGTTAAAGATGCGCTCTGTGTAGAGCTT
ATGAAAGCTGAAAAAATATTACAAACGATGGATGTCGTTTTTATGAAAACTTTAAATTGGC
GAATTTTCCATGTGCACCGAAAATCTTAATCAGTTGCTTAACAAATTCGCTACAGATCAG
TCAACACTTAGCGATGTGAGAAAATAAACAGTCTCATAGAGATTGATGGCGAAAACAGC
AAACGCTTTTTAGTCGAGTTGGATCCGATTTCTCCATGAGGAGACAGGATTATATCAAGCG
CTGCCAAATGTTGTGACGGAAGCTCCAAGTAAAAGGTTAAATCGATACGCGTTGAGTCT
GAAGGTGAGAGTGTATAG

>U45

ATGTACAGCGCAATTTAGAAAAGATCTCCGAAACGATTACCCTACAGAGACAAAACGAGC
TCGAGATACATTGAATTTTTCTGCTTTTCCGAAATGTGGACATCAATGAACTTTGGACTACA
GACATATCTGAAGACAAAACGCACGATGTGTGGCCGGCTGTAAATAAGAAATCATTTAAA
AAATTTCTAGAAAACGAACCTAACCTCTTATCAACGCCAATCCCCTTCTAGGTATTCCG
CAAAATGGAACAGTGTGCAAGACCTGCAAAAAAGAGAAGCAAAGAGAAACCGATTGTGTT
AACTATGAAAGGAAACACGACAACCCAGTAACTTTCTACCCAAGGCACCGCGCAAAAAGA
AATGCAAAATACGGACACTTGTATTTCCGGAAGAACCATCTATTTTAGTAAGCCATCACAGG
AATAGTAAAATGGACGTTTTTATGGACACCAACAAAATCACACTCGTAAATAGAGAACTC
ATATGGGTGCCGACGATCAAGTTGCAATCGTAAAACCTAGATATTTCCCTGTACATCCCA

GATGGTTTCTTCGGAGTTATTACTGGCCACAGCAACGACGTATTCTGCGAATGCGTCACT
GAAATAATAACCGACGAAACCGACATTTTCAGTTTTTCTAATGAACTTATCCGAGCACTCT
CTAATGCTTCTCCCAGGAGATGTGGAATTCTCAATCAACTTTCTCCCTTGCTACATCCCG
GAACCATGGGAGATGATCAATCTGTCTCCCCCGAATTTGCAATTTTCCATCTAAAAGCA
TCCAGAGAATTTATAGCTAAACCGAACAGCTATACGATACAATACTTCGACGCCATGTAC
GTTTGCGCCGATGAACTGAAGGCTCTCATGATTCCCTCAAAAGAAATTGCCAAACTCGGC
TTATTGATCGAAACCTACATCTGGAACAAAGACACAATTCCTCCATAAAGATCTTCAAC
TCCACACGCAAACTATCTACATCCCTACAGGCATCTGCATCGCCAGGATAATTTTTACT
TGCGGACATTTTTGCCTCAGTTAATGCCTGAACGTGCAATAAACCGGCTACAGGTACTA
GACGCCAATTCCTCTTTCTTGTTCATTATGCCGCCTCCAATAACGCTTAA

>U46

ATGAGTTGTAAGAAAAGCGAAAGGCAACGCTTATATGTCAGTCTCTGGCTCTTTTCCATT
CTTGTTTTTTCGGCCGCTACGGAAGTGGACTTTTTATTCTCCAGAGTGCCATTCGCATACC
TATGAGATCGTCTTGAATTCGTTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATTG
TTATGTTCTTTCGCGATTTTCTTGAATACTGGTGTATAAAGACTTTTGTTCAGAGACG
GTGAAGGGTTATTAG

>U47

ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAGCAAATACTATTTTAGTCAACA
TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTCACATCTCGCAACTC
GGCTTTTTTCTGGGCCTTTTTCGCGATAGTCATACACTCCGCTAATCTAATAAAAATACACG
TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAGA
GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTGACACAAATCTATCCCTC
GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC
AAGTCTGCGGTTAACTTGTCTCCATTACAGAATTTACCGCGGACAGAAACCAAATAGTG
CTATCGTCGACGGCTGCAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT
CTGTTTGCCACACTATCTACATATAACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT
AAAATTCCTGGGAGCACAGTTATTTGACATAAAAAGTACCGAAGAATCCTTGCGGATGGCA
ATGAGCAAATTTTCCAACCTCGAATCTGACACGGTCATTGACTTTCGTTTACGTCAGAAATC
TTCTTTAATTACACCAGCTTTGTTTACTTCTTGTCTATAACACAACATCATGCGTCCCT
TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTTGGGA
CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT
CTCACATACCACATATCCCTAGTACCAACATAACCAACCCAGAACCTCCCCCGTAACA
AAAACTCTACAAAACCTGCATACAGACACCATAAAAAGTTACACCGAACACACCCACCATA
ACAACGCAACCAACGAAAGCATCAAAAAAATAGTTAAACGTTTACGATTTTCTCGACCC
ATGTACACCCCAACCGACATTTCCAACCTTTACAATCCGTCTTAAACGCCACTATTTAAACC
GAACAAAACACCGAAAACCCAAAAGTCCACCAAAAACCAACAAATTTTGAAAATACCACA
ATCAGAAATTCCTAAAACCTTTGAGAGCGCTACAGTAACAACAACGCAACCCAAAAGATC
GAAAGCACCACCTTCAACAACATAGGAATCAAGGAAATTAACGGCAATACCTATTCTTCA
CCAAAAACTATATTTATCTTAAAGAGCAAATCACAGCAGAGTACAACAATAATTCACCGAC
GCCGAACACACCCTCCGATTTTAAAGTTTACCCTTGGCAAAACACGGCACGCACATAC
ATGAGTCACAACACAGAAGTACAAAACATGACCGGCAAATTCAGAGGACAACCTTGAAA
TCCTCAAACGAGCTACCTACCATTTCAGACGTTGTCTGTCACTCCAAAACAAAACCTACCG
TCGAATGTAAC TGCCAAAAC TGAAGTACACATAACTAACAACGTTTTTACCATCTAGTAAT
TCATCATACTCAATCACTAAAGTCACTAAAGAGGTAAAGCATACTAGAATGTCAGCGTCC
ACTCACGAAGAGATAAACCACACAGAAATAGCACAAATGACACCAATTCCTAACGCTCAC
ACATCGGAAAAATCAACTACACCTCAACGGTCTTACCAGCTGAAACGTTCTTAACGACA
TCTTCAAAGCCCGCCATCCTAACCTGGTCAAATTTATTATCAACAACGCCTAAGGAACCA
TTAACGAATACAAGTCTAAGGTGGACAGATCATATCACTACACAGCTAACGACTAGCAAT
AGAACTCAATCAGCCAAACTAACAAAAGCTAACATCTCGTCACAAACGACTAACATCTAC
CCCCAAACAATCACGGGACGATCTACAGAGGTTTAA

>U47A

ATGACCTCAAATGCATGCTGTGTTGGACCGCGATTACGAATGCCCTGGCTATTTTAAATG
TTATTACTAGTAATTTACCGGGCGATAATCGGATTTAATGGCGATATTTATACGAGCAGG
TTGTCGGCGTTTAGCTGTATACGGACAAATTCATCTAATACCTAA

>U48

ATGCTCCTCCGACTCTGGGTCTTTGTCTGTTGACTCCCTGTTACGGTTGGAGACCGTTG
AACATATCGAACTCGAGCCATTGTAGAAATGGAACTTTGAAAATCCAATCGTTCCGCCCC
GGCTTTATAACATTTAACTTTTATACAAAAACGACACTCGGATATATCAAGTCCCTAAA
TGCTTACTCGGCTCCGATATCACATACCACCTGTTTGTATGCCATCAACACGACAGAATCG
TTAACAAATTATGAAAAACGCGTTACACGTTTTTATGAGCCGCAATGAACGATATTTTA
AGACTTTCTCCTGTACCTTCGGTCAAACAATTTAACTTAGATCGCTCTATCCAACCGCAG
GTTGTGTATTCTTAAACATGTACCCTTACAGGGAATTTATTACGTCAGGGTTCGTAGAA
GTTTCGACAGATGCAATACGACAACGTTTTCTGTAAAGCTGCCTAATTCTCTCAAGGAAC
ATATTTCCAGTCCAAGTCAGATGCGCTAAAATTACGCGCTATGTGGGCGAAGACATCTAT
ACCCATTTCTTTACTCCGGACTTTATGATACTGTACATCCAGAATCCCGCGGGGAGATCTG
ACTATGATGTACGGAAACACCACCAGCATAAACTTTAAAGCCCCTTATAAGAAAAGTTCA
TTCATATTCAAACAGACATTGACAGACGATTTACTATTGATAGTCGAAAAAGACGTAATA
GATGTACAATACCGTTTCATATCAGATGCGACATTCGTAGACGAAACGTTGAATGACGTA
GATGAAGTAGAAGCTCTACTACTCAAATTTAATAACCTAGGAATCCAAACCTATTAAGA
GGAGACTGTAAAAAACCAACTATGCCGGCATAACCGCAGATGATGTTTTCTTTACGGTATC
GTACATTTCTCATATAGCACAAAAACACAGGACCAATGCCCGTGTAAAGAGTGTAAAG
ACACACGAAAAATCTCCTGTCCATCGACTCATTTGTCAACCGATGTGTGAACGTATCGGAA
GGTACGTTACAATACCCAAAAATGAAGGAATTTTTAAAATACGAGCCCTCGGACTATAGC
TACATAACCAAAAACAAATCCATTTCCGTATCTACGCTGCTCACGTACTTAGCGACAGCG
TACGAATCCAATGTAACGATTTCCAAGTACAAGTGGACTGACATTGCCAACACTCTACAA
AACATCTATGAAAAACACATGTTTTTTACTAATCTGACATTTTCCGATAGGGAAACTCTA
TTCATGCTAGCCGAAATAGCGAATATCATCCCTACCGATGAACGCATGCAGAGGCACATG
CAACTACTAATTGGAACCTGTGTAACCCCGTAGAAATAGTTTCATGGGCCCGCATGCTT
ACAGCTGACAGGGCACCGAATCTAGAAAATATTTATTCGCCTTGTGCCTCCCCCGTACGC
AGAGATGTAACAAATCTTTTTCTAAAAACAGTTCTCACGTACGCTTCCCTTGACCGTTAT
CGATCAGACATGATGGAGATGCTATCCGTATACAGACCGCCAAATATGGAGAGAGTAGCG
GCTATTCAGTGTCTCTCCCAAGTGAACCAGCAGCTTCTCTGACTCTGCCGAATGTGACA
TTTGTAAATTTCTCCCTCTTACGTGATTAAGGAGTGAGCCTAACAAATTACAACGACAATT
GTGGCTACGAGTATAATAATCACAGCCATACCTCTCAATTCACCTACGTTTCAACCAAC
TATAAATATGCAGGACAAGATCTGCTTGTGCTACGAAACATCTCATCTCAAACATGCGAG
TTCGTGTCAGAGCGTAGTCATGGAATATGATGATATCGACGGTCCCTTACAATACATTTAC
ATAAAAAACATAGACGAACATAAAAAACATTGACAGATCCAAACAACAATTTACTTGTCCC
AACACCAGGACGCATCTTTTTGTTAGCCAAAAACGGCTCTGTTTTTGGAGATGTCTGAA
GTCGGAATCGATATAGACCAAGTGTCTATCATATTGGTTATCATTTATATTCTGATCGCA
ATAATTGCTTTTATTCGGATTATATAGACTTATCAGATTGTGTTGA

>U49

ATGTCCTCTGACAGGTTTGCCTGATATCCGGAAAAAATTTGGACAGTTTCACCACTTGAGG
ATATATAAACAGATCTTATCTTTTTCAAGGAAATTTTGC GCGACTAAATTTATTTTTTAGGA
GACGTGTTTCCGGCTAATTTGCGATCGGCATCTGTCTCGGTGTTTTTTCGAGGTGCGTCTA
GGGCCGCGCATTCGGGATTGTATCGTCTTGCTAAAGAGTGTAGATGTCAAAGATGAGTTT
GCTTTTCATTGTTATTTTTTCGAGTTTAAAGACGACTTTGGGGAAGTCGACGATGCAGTCT
GTGCATCACAACATGCATTCACCAGGCGCAATATTTGCAAGGTTTGGAGACAGTTGCAACAG
TCTATTTTCGTTCTTGGATCAATATTTAATCGCAGACGAGGTTTCGTGGAACGTTGTTCCC
GTAATCTGTTTTTTTCAGGCAGTGGGGACTTAAGTTAGATTTTTTTAAAAAGTTTTTCGGGG
AAGACCAAGCGGTTGTCAATTTTCGTTTTATACGTGATTTGTTTCGCCAGGTCTCAGGATGGC
GCAGTGCAATCTCTTTTATCAATACCCAATTACACCAATTTTAGAAGGGCATGTTCAGAAA
CATACTGATTTGTACAGAAAAAGATGTGCAAAAAGCTCCAAAGTCAGTCCTCACTAAGACT
TCGGGAGAAAAATAGATCAAGGGCATCGCGACAAGTTGCTAAGAATGCGCCTAAAAACCGA
ATTAGACGCACGTCAAAAAAAGATGCAAAAAAGACAGTGA

>U50

ATGGCGCAGTGCAATCTCTTTTTATCAATACCCAATTACACCAATTTTAGAAGGGCATGTC
AGAAACATACTGATTTGTACAGAAAAAGATGTGCAAAAAGCTCCAAAGTCAGTCCTCACTA
AGACTTCGGGAGAAAAATAGATCAAGGGCATCGCGACAAGTTGCTAAGAATGCGCCTAAAA
ACCGAATTAGACGCACGTCAAAAAAAGATGCAAAAAAGACAGTACGTTTTTGAATTTCTCAT
TTGAAAGCGATAGAGGATGCTTTGCTGTTTACGAATGATGGGGAGGTAAACGTCGAAACG

AAAGCTGATGCGCAACTGATACCGAAATCTCCTGAGAAGCTGGAAAAATTCAATCAGGTC
GCTATAACCCCACTCGACCCCTTTATAAGATTTACAGACGATTTTAGAGGAGAAATGATC
AATACATTTTTCAATAATGCCAGATGTGGAATTTTACGTTTCGGGTCTTGGTTCTATAAG
CTCAAGAGGGTTTTTTTACAATGAGCCCGTTTTAAGAAGGGTTTTAAACTTACTAATGTG
GATTCCTGACTATTTCAAAGAGTTACTAGCAGTTACCGTCAATGCGTTGGAGCAAGCG
ACGGTGTATCCTATATTCGGAAGTGAATGTCAGACCTGGAAGCAGCCTTGTGTATCCTG
GCAGCTTTTTTATTCCACTTATGAGAATTCACAGATCGACGAACGTACAACCTAGAAGAT
GTTATAACCCCTCTTGCCTGTGATATTTAGATTACTCGGCAGTGAATTTACAGCATTGAAA
AATGTTTCTCCTTCTGGCACCTATTTTGGATTTAATGATCCTTCTTGTATGAAATTTTTT
GTTCCCATGAGAAAAGGCAAGCATTATGCAGAAAACACATTCGGGAATCATGTGCTTATA
AAGATGTTGTTAGGTAGAGGGGTAATGCAAAAAATTCCTGGTGAGAAAATTAGTCAAAAT
TTTGATGTTGAAGCTCGTTTACACGGGGCCATTAAGAATGATGTTTTGGTGTATTGGACA
TATCAGCTAATGAGGCCAAAATGGGAAATAATGTTCCAATCTTCATACACGATCAACAT
TATTTACGGTCTGGCTTGGTTGCAATGAAAGTCTTTTTTTATTGTGGCGGATCCTGAAC
TCAGAAAGTTTGTAAACAAGAGAGTTGGGAAGTTTCTCTTAACTTCGGTTTTTCCACAG
TTGGAGAATGTCGATTTCCGCGAAAATAACTTTGAAGCTGGAAATATTCAAAATTTTGA
TACTTAATGCACCATTACGTTGTGCCATGTACAACCTGCAGAATGACATATCAATTTTCG
ACTCTTTTTCTGGTCTCGTGGCCGTGTGTGTAATGAGAGTGTGAGATTGGGATGGGAA
CATAAGTGTGCGGGGGCTCCGTCCGATGCCGTTTCAAGTTCAATCTAAGGAAAATCCGTTT
GTGGAGTATATCCGCGCTCAAATGGAACAACAGGCGGATGTGGCGATACTTGAAGGAT
GACTGCATTCTCTTTTATCTTGAAGACGGTTTTAAACATCACTCTATCTTTTACATTACCG
AGACAGAGATTGTTTGAATGGCTTCTTCAATTATTTAATGTAATGATACTTATGATTTT
ATCTATTTTCTAGTTTTGGGTTTTCTGCCAATACCGCCGTTATATAA

>U51

ATGGGAAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
TTTATTTTTTTCGTCATACAGTTGTGTACAATGGTTTTGCTGACGGTTAGGTTTAATAGT
TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTTAATTGTTTT
TTACTTCCGATAAAGATGGGATTTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
GCAATCTTCTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGT
ATGGCGATCGAGCGCATCAACCCTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGAAGCTCTT
CAGATGCTGAATAATTACAAAGCAACAGCTTTCGAGAACGAGACCTCACAGTGTCTCTC
GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTTATTCCGGTCATGATTCCA
GTTATGCTGGTGTATTATTTACCTATACAGCAAAAATTTTATGCTTTTGAAGATGAACTT
TCGACCGTGACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG
AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTTGCAGGATTTTTTT
ATGGCGCTTCTGTTTTTAAAGTTTTATTATATACCGTTATGGCCTATTGCATGGCATGT
GATGATCACACTGTGCCGTTTCTGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
TGTTTTTCTGTGTACGCGGCGGAAAAGGAGTCCGATTTAGAGGTTGGAATAAAAATGTTA
AAATGA

>U52

ATGACCCAGATAGGACAGTACATAAAAATTGAACGACGCCGTCCTAATTTAATATTACAC
ATAACCACAAAACCTGCTTAGAAACGAGAATCTGACTAGCTTTAAACAAGAGGAATTACTT
TTAATTC AACACGTATGTACATCAATGCTGTACACGGAATCAAGATCTTACTACTACGC
GAAAGTTTGTACAATTCGGAATCGGAGACATAGTATTTGAATCGTAAAATCAGCAAC
AACTACTGGTTTTCGAATATTTTCTATTTCTGAAACAGCATTCTGACGCGGAACTACTCCGA
CATATGTTTAAACGAAAGTCACTCAGCATATATCTCAAAAAGCTGCACTACAGCGGAAAT
GTTAGTCACATGATTAATTTTTCTTTTTCATGGACGAGTTTGGTGTAAATCTCAAAATTCG
GAAGAATTAATCTGTGAAGGCAATATAGTTTTCTCCGTTGGAGCTATCTATAATCACCGA
CTTTTAAAAATCTGCAGATCTTTTAAACCGATTTTGGGGGGATCAAGAACGAGAACCAACA
GTGCGGCTCATCTGCAAACACCTATGGTTTTGCCTATCTTATAATGTTTGGAAAATTTGAG
ATATCCACACTGGCGTACAATCAGCAAAGAGCCGAGCACAAGCCGACTGTTTTTCAATTT
CTCCAAAACGACTTTAAAGTCTTCTGCGGTATGTCTGAAAATCCACAATTTACTCGACAGC
TCTGCCATATTTGACCTTACAGGAATCAGCGCCGAAGATCTATTCTCTTATGAATAA

>U53

ATGTCTAAAGTTTGGGTAGGTGGATTCCTTTTGTGTTTATGGTGAAGAACCGTCCGGAAGAA
TGTTTAGCTCTGCCAGGGACACGGTTCAAAAAGAACTGGGGTCCGGAATATTCCTTTA
CCGTTGAATATTAATCACAAACGAAAAGGCCACTATTGGAATGGTACGCGGTCTTTTCGAT
TTAGAGCATGGGCTTTTCTGCGTTGCGCAGATACAGTCTCAAACGTTTCATGGACATCATA
AGAAATATTGCTGGCAAGTCAAAGTTAATCACAGCAGGCTCGGTAATTGAACCTTTACCA
CCGGACCCGGAGATTGAGTGTTTAAGTTCAAGTTTTCTGGTCTATCGCTGTCTAGCAAA
GTTTTGCAGGATGAAAATTTAGACGGTAAACCGTTTTTTCATCATGTGTCTGTATGTGGA
GTCGGTCCGAGACCCGGAACTATAGCGATTTTCGGGCGAGAAATAAGCTGGATTCTAGAT
AGGTTTTTCATGCATTAGTGAGAGTGAGAAGAGACAGTTTTTGGATGGGGTCAATGTTTAT
TCTCAGGGTTTCGATGAAAATTTATTTTCGGCGGATTTATATGACTTGCTCGCGGATAGT
TTGGATACATCTTATATTAGAAAACGGTTTTCCAAAGTTGCAGTTGGATAAACAGCTATGT
GGTTTGTCTAAATGTACGTATATTAAGCTAGTGAACCACCCGTGGAGATTATTGTGCGCA
GCGACCAAAGTTGCCGGTGATCAAGTTCAGTTAACTACGGAACCTGGGTCTGAATTAGCG
GTTGAAACGTGTGACGTGCCGTGTTGTGCACGGGAATTACGACGCCGTTGAGTCTGCGACA
GCTACAACGGCTATGAGTAATCAAAATTTGCCAAATACTACTCCTTTGCTGTGCGAGTCCA
CCGTTTTCCGATTGTGTTTTTTTTACCGAAAGACGCTTTTTTTTTCTCTTTTAAATGTTACA
ACCGGACAGCAACCGAAAATAGTTCCGCCTGTTTTCTGTTTCATCCGCCCGTAACTGAACAG
TATCAAATGCTACCGTATTCTGAGTCCGGCTGCTAAGATTGCAGAACATGAGTCAATCGA
TATCACAGTCCCTGTGTCAGGCAATGTATCCCTATTGGCAATATTCCTCCGTTCCCCAGTAT
CCGGCCCGTTACACGGTTATCGTCAATCGAAAACGTTAAAAAGCGGCATTTCCAGAGT
GATTCGGAAGATGAATTAAGTTTCCCGGGAGATCCGGAATACACAAAAAAGGAGGCGC
CATAGAGTTGACAACGACGACGATAAGGAGATGGCTCGAGAAAAGAACGATTTAAGAGAA
TTGGTGGATATGATAGGAATGTTAAGACAAGAGATTAGTGCCTGTTGAAGCACGTTTCGCGCT
CAATCGCCGACAGAGACATATCGTTCCGATGGAGACTCTGCCTACGATCGAGGAGAAAAGGC
GCCGCTCCCAAAGCCATCTATTTTAAACGTTCTTTGGCGCTGAAACCGTAAATAGG
AGCCTTGCTGGTCAGAACGAATCCACGGATCTGCTGAAACTCAACAAGAAATTGTTTGT
GACGCGTTAAATAAAATGGATAGTTAA

>U53.5

ATGAGTAATCAAAATTTGCCAAATACTACTCCTTTGCTGTGCGAGTCCACCGTTTTCCGAT
TGTTTTTTTTTTACCGAAAGACGCTTTTTTTTTCTCTTTTAAATGTTACAACCGGACAGCAA
CCGAAAATAGTTCCGCCTGTTTTCTGTTTCATCCGCCCGTAACTGAACAGTATCAAATGCTA
CCGTATTCTGAGTCCGGCTGCTAAGATTGCAGAACATGAGTCAATCGATATCACAGTCCCT
TGTCAGGCAATGTATCCCTATTGGCAATATTCCTCCGTTCCCCAGTATCCGGCCCGGTTA
CACGGTTATCGTCAATCGAAAACGTTAAAAAGCGGCATTTCCAGAGTGATTCGGAAGAT
GAATTAAGTTTCCCGGGAGATCCGGAATACACAAAAAAGGAGGCGCCATAGAGTTGAC
AACGACGACGATAAGGAGATGGCTCGAGAAAAGAACGATTTAAGAGAATTGGTGGATATG
ATAGGAATGTTAAGACAAGAGATTAGTGCCTGTTGAAGCACGTTTCGCGCTCAATCGCCGAC
AGACATATCGTTCCGATGGAGACTCTGCCTACGATCGAGGAGAAAAGGCGCCGCTCCCA
AAGCCATCTATTTTAAACGTTCTTTGGCGCTGAAACCGTAAATAGGAGCCTTGCTGGT
CAGAACGAATCCACGGATCTGCTGAAACTCAACAAGAAATTGTTTGTGACGCGTTAAAT
AAAATGGATAGTTAA

>U54

ATGCAACCCGCCACTCTACAATGGTCTTCGTACGTACTACAGCTAAGAGTAACAACCTACC
GCCATATTTAAACCCGGCGAAGTGAAGATTGTTTAAATGCGGATTAGGTATCACCCCCCT
TCATCCGCCGTAGTGTGCATTTGTAGAGACGAAAGCAGTTTTTACATCATCGCCTTTTACT
TACATCGATCCAAAAGATTACGGCAACATACCCTTCGCGGTGCAGAACATCTCAGATCTA
GATCTGGATCTCAGTAGATTACCTATCGTCTTAATATATTTGCACTTCTCAGCAAAC
GTGAATATATCAAATTTACCGGTGCAACGCATCGAAGCCACGATCGCCACATTATACCC
CACGGACAGTGCAGCGAAGTTTGTATATATGGCCCTCTCACGCGTATTTAAATTTAA
GTGGCTGATATACGATGGATAGAACAACACCAGAAGAGCCAACACGATATTCATTTAAG
GCAGAAATATGGGTTAATCTCCAAAACACTCCGCTCGACCAAATTTTTTAAACTGCTAAC
ATCGAATTTATATCACAAAAACGTCTATATTTCTAGAATCTTACTCTGCGGAAATCTA
TTAATTTTGAAGGCAAAATACGAAAACAACCTATCTCCTTGACGATAATTTTTATCTGAA
CAACTTGCCATACAAATTTCACTGACCCCCAAACCGCAGACATTACATTGTGCGAGAAT
CAAGAAACACTTCTAAATGCAACACCCATTCAATAACTGTCTGCGCAACTAAAAATATC
ATCCCTAACACGGTTACTCCTGTCCATTGTTCTTTTAACTATCTTTAATTCAAATTC

AACTTTACGGGTTTATTTATACCCAACTCTTGCTCGGAATCTCTACAACCACCGGCATT
TGGGATGAGACACGGCCTCTCTACATCACCATGAAAGCATTAAAGAAGAACCCTCGCATC
AATTA CTGCAACCTATAGGCGTAGTATACTTCTTCCCAAACAAATTTTACCACCGGGG
AATAATGTGCAATTCAACTGGACAGAAACAAGCAAACCTTTATGTAAACACAGAATCTCCA
AATGGTCCCGTAAGAAACATCGTAACTTTTACAAAACCAAGCAGTATTGCGAACACCAAGT
CTAAGCACGGTTGCAAACCTAACGCCAGACATCAACATGGGAATTTTTTTGAGTTCACTC
CGCGTTGCGTTGACACCGCACATATGGTACCCCTGCACCTTTCTCTAAAACCCGGAGAA
TCTACTCGCATGGAATTCATGCCGCTGGAACCCACAAAACCTAACGATATTAGAAGGA
GACGTCGGCATTCAATTCATTCTTGTCCACAATCACTCCCATAGATCTTCCCGTAA

>U55

ATGTTTTCTCCTCAACTAAACCATGTCTTGCAACTCGTTTTTGATTGCAATCACATAGTT
TCGAATTCCTTACCAAGAAATCATTTTGAAATGTGCAATCTCAAGCACAAAACCAGGAATC
CTATTCATAACAGATGGCACTGAATCGGAAACTGAGGACGTTCTGTTTCGATGTACAGTTT
CTCGAAAAACAACGACTATCAAAAATAAAAAATTTTACCTATTTCTAGGACGGTTCCCGTA
CGGATTCATTTATTTTCCATACCCATTTATCTGCCGTCACATGCACTGGAGATTGACGAC
ATCAATGCGCTATGTCAAAAAGAAAACATCGATCCGACCGTATACGAACAATATGGCCTA
ACCAATATGCCATCTCGACAACCAACGTGATAAATCAATTTAAAGTAGTTTCAGAATGAT
AGCGTTGAAATGTCTTTGATCTTATTTAACCTAAACTGGACAGAAAGTCCTTATCAGTCT
CAAAAATTAGACCGCTATGGCAGAAAGCTGCTAATAGTTTTTTAGCGTGTTCAGCATCGAT
ACCAGAGAATGTCCCTATTGGGAAAACCTATTCAAAAAGTATCATCCACTCTGCAGAGTC
AACATGATTTCCGAACCCGGCATCTCTATCTACAAAATAGAATTTAACGATCCTTTTCTC
ACAATCTTTTTGCGTAACGTAGTGTGGCCGAAAAACAAGCGACGGTTTTCAATCACGAAA
GAATGTTATCTTCGCTTGACCTTTTCCCCCCTGATCGGAGCACCGTCACGCTAAATATG
AATATGCCGTATTTTAAAATCTGCGCAGACGGGAAGAATCTAGAGGTTTTCTTTCCAGC
GATATGTCTTAGTAAGTAATACAAAACAAGAGATCACGCTACGTGGCGAGTTCACTAAC
AAAAAATCGTAGGCATCTTCATCCCTGTCCAGACGCATGTCTTCAAGCTGTACCCATTT
TTATGGTTTGCCCAAGAACCCTTAAAATGCGGATTTTCGTGCGAAAGAACAGTTTCCGTA
ACCGAACACGACATGTTAGGTGCTGTCTTTTTTTGTGACGAACAAGTTGTTTAGAAAGCCG
TTTCGCTCCCTCTCAAACAGCGACATGAAATCTCTTTTAAACACACACGGAGAATACCGAT
AAACCATTGACGTACACTTCATGGGCAATTGTTTTCTGCATACTTGCTACCGGAACTC
ACATTACATCCCTTAGAATACGAAGACAAGCCTGCGCGTGGAACCCGCAGACTCTCCCG
TCCGTGGA AAAACACTCTATTATAAAAATACGCATATGA

>U56

ATGGA AACCGTCTACTGCACCTTCGATCACAACTGTCACTTTCCGATATCAGCACCCTA
TGCAAGCTCATGAACATCGTCATACCGATCCCAGCTCACCACCATCTAATAGGTAGCGGC
AATTTAGGTCTTTATCCCATCGTCTCCTCCAACAAAGATTACGTCCACGTCCGAAACGTA
TTACGGACCATGGTCGTAACCATCTGCAAAAGGTGGAAGGCAATCAACTGGTTTTGAGA
AAGCCGGTGACCGGCCATCAGTATGCCATAAAGAACAACAGGACCTTCCCCTGGGAAAAG
GGTGACACCTTGACATTGATTCCACCTTTATCGACGCACTCGGAAGAAAAGCTATTGAAA
CTAGGTGATTGGGAGCTCACTGTACCTTTAGTCGTTCCGACAGCCATTGCGGCCGAGATC
AACATCAGATTGCTATGCATTGGTCTTATAGCAGTTCACAGAGAATACAATGAGATGCAA
ACAATCATCGATGAGTTATGCTCTATCCAATACAGAGACGTA CTATTAAATTACCAGAC
ATAGTCAACGACAAACAATCAATGTACTCCATGAAAACAGCTTGCACTCTCGCTGTCCATG
ATCACGGCCATGGCCCCAGACATCGTGAGA ACTTACATAGATCGGCTAACGCTGGAAGAC
CATTC AATGCTTTTTGATCAAATGCCAAGA ACTATTATCTAAAAGAACAACGTTAAACACA
CAGCGCTGCGGACA ACTTACGCTACGGA AATTAAGATGAATTA AAAAAAGTCAAATCA
GTCTTAACTATGATCGACCAATTAATCTTTGACCAACGAGAAAACGATTTTTGTTGTC
TGTGACGTTTTCCGCAGATAACAGAATGGCTACATGTATTTATAAAAATTA

>U57

ATGGA AAAATTGGCAGGCGACCGAAATTTTACCTAAGATCGAAGCACCTCTAAATATTTTC
AATGACATTA AAACATACACAGCCGAACA ACTTTTTGACAATTTGCGAATTTATTTTCGGT
GACGATCCGAGCCGTTACAACATCAGTTTTTGAAGCCTTACTCGGAATCTACTGCAACAAA
ATAGAATGGATTAACTTTTTTACCACCCCGATCGCCGTTGCAGCGAACGTAATCCGCTTC
AATGATGTGAGTCGAATGACCTCGGGAAGGTTCTCTTCTTTATTCAATTACCAGAGTC
GCCACAGGAAACGACGTAACCGCTTCAA AAGAAACCACCATCATGGTAGCCAAACACTCA

GAAAAACACCCCATAAACATATCGTTTCGATTTGAGCGCTGCCTGTCTGGAACATCTGGAA
AACACATTTAAAAACACAGTCATCGATCAGATTTTAAACATCAATGCGTTGCATACAGTC
TTAAGATCTTTAAAGAATTCAGCCGATTCGCTCGAGCGAGGTCTGATTCACGCATTCATG
CAAACCTTATTGAGGAAATCCCCCCGCAATTTATCGTCTGACCATGAATGAGAACAAA
GTACATAATAAACAAGCTCTGAGCCGAGTACAGCGCAGCAACATGTTTTCAAAGCCTAAAG
AACAGATTGTTAACGTCAATATTTTTTTTTGAACAGAAATAATAACTTTCATATATCTAT
AGAATTCATAACGACATGATGGAATCGGTTACGGAAAGCATTCTAAATGATACGAACAAC
TACACTTCCAAAGAAAACGTCCCCCTAGATGGCGTTTTATTAGGACCGATCGGCTCTATC
CAAAAACCTCACCAGCATACTCTCCAGTACATCTCCACACAAGTCGTCTCCGCCCAATC
TCATATGGTCACTTTATTATGGGCAAAGAAAACGCAGTGACTGCGATTGCATACCGTGCA
ATCATGGCCGATTTTACTCAATTCACCGTGAACGCCGGGACAGAACAACAAGACACTAAC
AACAAATCAGAAATATTCGACAAAAACCGCGCGTACGCCGACCTAAAGCTGAACACGTTG
AAATGGGAGATAAATTAGTCGCATTCGACCACCTACACAAAGTTTACAAAAACACAGAC
GTCACGATCCGCTAGAACAGAGCTTACAACCTGACATTCTTTTTCCCTTTGGGTATCTAC
ATACCGAGCGAAACCGTTTTTCAGTACAATGGAAACACGTGTGAAATTAACGACACCATG
GAAAACAACCTACCCACCAGCGTTTTTTTTCCACAATAAAGACCAAATCGTGCAGCGAATT
GATTTTGCCGACATATTGCCGTGCGTTTTGCCATCCATTGTCCACGACTCGACCATCGTC
GAACGACTCATGAAAAGCGAACCATTGCCTACCGGCCACCGCTTTTTCCCAACTATGTCAA
CTAAAAATTACCCGAGAAAACCCAGCCAGGATCTTACAGACCTTATACAACCTTATACGAA
AGTCGACAAGAAGTACCCAAAAACACCAACGTCTTAAAAAACGAATTAACATTGAAGAT
TTTTACAAACCGGACAATCCAACACTGCCGACCGAAAAGACACCCCTTCTTCGATCTCAG
TATATCCAGAAAAACCGAGCCACAGAAGTACTCTGCACACCAAGAATAATGATAGGCAAC
ATACCTTTACCGTTAGCTCCAGTCTCTTTCCACGAAGCCCGTACAAATCAAATACTGGAA
CATGCAAAGACGAAC TGCCAAAAC TACGACTTACCCTCAAATTTGTCACCGAAAAGCTTG
ACGAGTGGCTCGTACCCAGAATTGGCTTACGTTATCGAGACCTTAGTGCATGGAAACAAG
CATGCTTTTATGATCCTAAAACAAGTAATTAGCCAATGTATTTCTTATTGGTTTTAACATG
AAACATATACTTCTTTTTTGCAACAGCTTCGAGATGATCATGCTAATCTCTAACCACATG
GGCGACGAAC TGATCCCGGAGCAGCTTTCACTCACTACAGAAATCTTGTGTGCGTAATT
CGCTTAGTGAAGAGAACAATCTCTATCTCCAACCTCAACGAGCAACTTTGCGGCGAACCT
CTGGTGAATTTGCGCAACGCGTTGTTGACGCGGACGTCTGTTCTGCCCGTTTCGTCCATACC
ATGCCAGAAAACGACACGAATGCAAAAATAACAGCGGATGATACACCCTGACACAGAAC
ACCGTAAGAGTTAGAAATACGAAATATCCGATGTGCAAAGAATGAATCTAATAGATTCA
AGCGTCTCTTACCAGCAATGACAGACCATCGAACGAAACCACCATCCTGAGCAAGATA
TTTTACTTCTGCGTACTCCCGGCATCAAAATAACAAGGCCTGTGGCGCTGGCGTCAAC
GTAAAGGAAC TAGTTCTAGACTTATTCTACACGGAACCGTTCATCAGTCCAGATGATTAT
TTCCAGGAGAATCCGATTACCAGCGACGTTCTAATGTCTCTGATCCGAGAAGGTATGGGC
CCTGGCTACACCGTAGCCAACACATCCTGTATCGCAAAACAGCTGTTTTAAATCGCTAATC
TACATTAATGAAAATACGAAAATACTGGAAGTGGAGGTCTCCTTAGATCCTGCGCAGCGA
CACGGCAACTCCGTTCAATTTCAATCACTACAACACATTCTATAACAACGGGCTTTGCCTG
ATCTACCAATCACCACCTAAGACGGTACTATCAACCAATCCATTTTCATCGATTCTTC
TCCGACCCAGGAATCTGTGGCACCATGAATGCTGATATCCAAGTTTTCTAAATACATTT
CCTCACTATCAAAGAAACGACGGCGTTTTTCCACTCCCGCCCCATTAGCATTAGAATTT
TATAATTGGCAACGAACACCGTTTTTCCGTGTACTCAGCCTTCTGCCCAATTCCTGTG
AGCATTATGACGCTTGCCGCCATGCACTCAAATTTGTCTCCCGTTGCCATAGCGATCCAA
AGCAAAAACAAAATCCATCCGGGCTTTGCGGCCACACTAGTCCGGACGGATAATTTTGAC
GTCGAGTGCC TATTATACAGTTCAGAGCAGCCACATCTATAATTTTAGACGATCCACG
GTCACCGCGGAAGCTAAAGATATCGCAACCCTTACAACCTTACCCAGCACCTAAGTTTT
GTAGATATGGGCTTAGGTTTTAGCTCTACCACCGCCACTGCCAATCTTAAGCGAATTA
TCAGATATGGGGAGCAAGATACAAAACCTTTTTCTCCGCCTTCCCGATACACGCGTTTACC
AACGCGGACATAAATACGTGGATTTCGACATCACGTGCGGATAGAAAAACCTAATCCCTCC
GAGAACGAAGCATAAACATCATAACATTCCGGCGGAATTAACAAAAACCCACCTCCATA
CTACTGCATGGTCAACAAGCTATCTGCGAAGTTATACTGACCCCGGTCACGACAAACATT
AATTTTTTCAAATCGCCCCACAACCAAGAGGCAGGGAATCGTGTATGATGGGAACGGAC
CCGCACAACGAAGAGGCGGCTAGAAAAGCATTGTACGACCACACCCAAACAGACAGCGAT
ACATTCGCCGCAACCACAAATCCTTGGGCATCTCTACCAGGCTCCTTAGGCGATATTCTA
TACAACACGGCACACAGAGAACAAC TATGTTACAACCCCAAGACATACAGTCCCAACGCT
CAATTTTTTACC GAATCTGACATCTTAAAAACAACAAGATGATGTACAAAGTGATAAGC
GAATAC TG CATGAAATCGAAC TCGTGT TTAACAGCGATAGCGAAATACAATACTCGTGC

TCTGAGGGCACGGATAGCTTCGTAAGCAGACCATGCCAGTTCTTACAAAACGCTCTGCCT
CTTCACTGTTTCATCCAACCAAGCTCTATTAGAGAGTCGGTCTAAAACCGGCAATACGCAG
ATCAGCGAAACCCATTATTGTAATTACGCCATAGGAGAAACCATACCTTTCCAACCTCATT
ATCGAATCATCTATATAA

>U58

ATGCAGCACACCGGTAATTGTGAAACTTTAATAGTTAATTCATGTTTTGGTTCCACTTGT
GCACGGTCGATTCCAGTTTTTATAGATTTCGTGTGATCTGACCGCTGAAGTGTCTCGCGAT
GAGGAAACTCGTTTAGCGCGATCCGTGCCTGTAGTTCTGGAAAAGATCGAGTCAATCATA
GAGAAAATTTTTCAAACGTCCGGGCCAAATATCGTTTCATGATAAAGACAGGGCTAAAATT
GCCTTGTGTCTGTGTTGTTAGGACCAGTTGCGGTGCCATGCTTCTGTGAAGAGTGGGAC
ACCAACGACTATTTGTGCGAAATCTGGCTGTAAATGCATAGGTCCAATCTTATATATTCAT
ACCAGTCGATGTCGTTGTAGCGACATTCCGATTTTTAAGTTTTCTATTATGAAAGATTAC
TATGCTTACACGTGTTTCAGAGGTTTTATTATCTCTGAAAGAGTGGAAATACACATCTACCG
AATGTATTGTGTACGTGTGAGTTGTGCGATGAGCGATAGATATGTGGCGACAGTGTATCCT
AAGCAAAATTCATTTATCTAGAATACTATCCGATTTTTTTGTGCTACCTGTGTGCGCCAT
CTTACTGTCAATTGAGATTGAGCAGTGTACAAATGACTTAATTTTCGCTTCTTGGCCCTAAA
GTAGCTCAGCGAGTCATAATTCATTTTAAATTGCTTTTTTGGTTTTTCGTCAAGCCGCAT
ATTGGTACTGTTGATTTCGTGGTTCTGGGAAAATTTTTTTATGTTAGAATTGCATAAGCTT
TGTTAACCCTAGTCAAACATAATCGGGTGACGACAGATTTTTTTAATGTAGTCTACGAG
AAAATTCAAAACATAAAACAATACGCAATCAAGACCCTGAGGATGTCGTCTAAGGCGGTT
CCTGCAATACAGAGGTTGTGTTTGGCAAAAATTTAAGCAGCAACTGTTGTATCTAAATATT
AAGGTTACGGTAAAAAAAATAAGCGAGAAATGTGTTTGAATGGCTTTGTTTACGGTAAA
ACATTGTATGTGCTTGAATCTTCTCAGTTAATCTTCCGGAATTTGCTTCTGTTGTATTAC
GATTACAGTTTGC CGGACGAATGCAAGACAAACGAAGAAAACGTTTTGACGGCTCATTAC
ATACGAGTAATTTGAGATTGTGCTTTAAGCGGTCTCGGAGTGCCTCCCGCCAGGCGTG
AGATCAGATTTTATCTTTGTGGCACAACAACCTAAACGTAAAGAGTTACCTAATGTTCCC
GGTGGTATCGATTTTGTGAAATTACCTCAGTGAGGCATGGCGCGGTAACCTTAAACGCG
TTTAAATACGAACAAAGTCATGAATTTAAAAGCAACCATTTCAAAAAGGGCTAACTTTGTA
TATCATCGCATTCCTAAGACGATGACCCACAGTTTTGTGTCATGTACAAGCATAACGTTTAAA
GAACCTGCGTTTACCGTAAGCACGTTTTGTTTCAAACGATGATTTAGATATGAGTTTCGTTG
AATATCAACATACGTGGACCGTACTGCGACTTTTTTATATGCTTTAGGCGTTTATAAGATG
CATGTTTCTATCCGAGATCTATTTTTACCGCGTTTCGTTTGAATAGCAATAATTCAGTG
GATTTACAGGGACTGGAATAATCAGGATGTTGTGAGAAATAGAAAAGAAAAGGTGTATTGG
ATCACTAACTTTCCGTGCATGATTTCTAATGCTAACAAAGTGAATGTGGGATGGTTTAAA
GCAGGAACGGGTATTATTCCTCGGGTGTCTGGGGAGGACCTTCAAAATGTTTTGCTTACG
GAATTAATAACGTTTCGAGAGATTCCTCGGGTTAGTCTTTGATATGGATTTGCATCAACTG
CTTGTTTTACTGGAACAACGAAATCTACATCAGATTCGTTTTCTCGTTAAACAGTTTCTT
ATTTTTTTACGTCTCGGTCTGTTAATGGGTTACGGGCACTCTCGGCGCAACAAGGTGCAT
GATATCATGTTACATTTAATTTCGAATGGTCTGTTTGAATTTAATAAGAACTCCGTAGCA
AATACAAAATCAAACACGGGTGTGCGTTGGTTGGGACGCGGCTCGCCAACAATGTTCCG
AAAATCATTGCTAGGCAGAAGAAAATGAAGCTAGATCACATGGGACGAAATGCTAATTCG
CTCGCCGTGTTGCGTTTTATCGTTAAAAGTGGGGAACAGAAAATAAAACTGTTTTTCATT
AAATTGTTGGAATATTTAGCGGAAACCTCAACTGCCATAAATACCGGGAATGAAGTCGCC
AGATTACTTCAGACTCTGACGGCTAAGGTGAAAACATGA

>U59

ATGAATGTACTCGTGGCCGACGAATGGTTTCGATTGCGCGATTAGGTTAGATTCGGAAACC
ATAGCTGTCCATGAGATTTTCAATCCGGAGTTAAGTAAACTGCTTAACTTGCCTCGAAA
ACAGTCTACATGTCCGACCTATGCGCTTTTTATTTCTGGTTGTGTTAATCGGAATGTCGGT
AACTTACCATATATTGGCATGTGAACGGGGACATAATCTACGCATTGACTGGTATTTTA
CATTGTGTAATAAATAAAGATAGAGTGCGGGGAGAGAATTGCCGATGGTTCGATATAGATTA
TACGAAATTCCTAAATTTATTTTTAATGAGAGGACAGTCAACACCCATGGAATTGAAGTGG
AAGCACGCCGTGGGTATCGCGACGACGAATAAGCCTTTGCTGACGCACGTTTTAACAGAT
GTGTTGGAACGCTCTCCTTTTACCTTGCCAGATACGCTTCTGTGCGGTGCAGGAGTTGTCT
ATTTTCAGAGAGAGATTGTGCTACATTTACTATGTGCTGGGATCAGATGTTGATATCGTA
GCGCGGACAGAGAGAGAGATTTTTTCAAAAATGTGCAGAACTAGCTCGCCTACAGCAAGTG
TTCTTATTCAAGGAAATATTATGGAAAACCTTTGTCTCTCGTCCAGGCTTGTCTATTTTCAG

CTGGGGGCTGATGGTTTGTGGGAAGAGATATCTGGTTCTGTACGTCCTAGGCCGGAGTTG
ATGTCAGTGCCTTCATTCAACACAGAGTAATGTTGAATAATTGTTATTGTATCGCTGTC
ATCTTCAATGCCATTTACAAACACAACTTTCCCTGCCTACCGTAGAAAGAAGCCACGAA
ACCGTTAATCGTGTAGTTCAGGAATATTATAAGTCTTATGTGAATGCTCCTCTCTGTT
CTTGTGTGTGCGACTAAGGTGCTTACTTTGTTTACAGAAGAATATAACTTTAAGTCAGCT
CTCGTATTTGTCAGTCAGTTTTTCCAGGTGGACGTCGAGGCTTCGAGAGCGGATGTGATT
CGTCTGTTTTTAGCGTGTCTAAAGGGTGATTAA

>U60/66

ATGCTACGTACCTGCGACATAACCCACATAAAAAATAACTATGAAGCGATCATATGGAAA
GGCGAAAGAGACTGCTCGACAATATCTACCAAGTACCCCAACAGTGCCATCTTCTATAAG
AAGCGTTTTATAATGCTCAGCCAGAACTCGGCTTTGCCATTTCATATAACCAACAGGTA
AACTCTATATACATTTTGTGAAAAACAGCGACATCTTAAAAACCGTAAGCCTCTTACT
ATACTACCGAGCCTTTCGCACAAGCTACAAGAAATGAAGTTTTTACCCGCTTCCGACAAA
TCCTTTGAATCACAATACACGGAATTTCTTGAATCCTTTAAGATTCTATATCGTGAACCG
CTGTTCTTGCAAATAGACGGTTTTATCAAAGATTTTCAGAAAATGGATCAAAGGAGAATTC
AACGACTTTGGAGACACGCGAAAATACAACCTGGAACCATTTTCAGAAGAACATTTCTCATA
CATGTAATCTTCTTCATCGCCGTTACCAAACCTTCAGCTTTAGCCAATCGTGTATCAAC
TATCTGACCCACGTGTTTCGACATTGAATTCGTAACGAAAGTACTCTGAATACGCTGAAA
CAAAAACTAACGTATTCCTAGTCCCAAGAAGACATGGAAAAACGTGGTTTCATAGTTCTT
ATTATAAGTTTTTTTATTGAAGAACATTGAAGGAATCAGCATCGGTTACGTGGCACATCAA
AACACGTATCTCACTTTGTGATGAAAGAAGTCGAGTTTTAAATGCCGCCGAATGTTTCCA
GAAAAGACAATCACATGCCCTGGATAATGTAATTACAATCGACCATCAAAATATAAAGAGC
ACAGCACTTTTTCGCCAGCTGCTACAATACACACAGTATACGCGGACAGAGCTTCAATCTG
CTTATTGTGCGACGAGGCCACTTTATAAAGAAAGACGCATTCAGCACGATTTTAGGTTTC
CTGCCCAAGCATCTACCAAGATCCTATTTATATCTTCGACCAATTCGGAAACCATTCA
ACATCTTTTTTAAATGAAGCTGAACAACCTCTCCGTTTGAAATGCTCTCCGTAGTTTCTTAC
GTATGCGAAGATCACGCCCATATGTTAAACGAGCGCGGTAACGCCACAGCCTGTTCTTGC
TACCGACTCCACAAACCAAGTTCATCAGCATTAACGCCGAAGTTAAAAAACCGCCAAC
CTCTTCCTAGAAGGAGCTTTTCATCCATGAAATCATGGGCGGAGCCACTTGCAATGTGATC
AACGACGTTTTTAAATTACGGAACAAGGCCAAACCGAATTCGAGTTTTTCCGATACAGCACC
ATCAATAAAAAATCTAATTCCTTTTTTAGGTAAGATCTGTACGTATACCTAGATCCTGCC
TATACAGGCAACCGCGTGTCTTGGCACAGGCATAGCAGCAATAGGCACCTATCTAGAT
CAATATATCGTCTACGGCATGGAGCACTACTTTCTAGAAAGTTTAAATGACAAGTTCCGAC
ACGGCAATAGCTGAATGTGCCGCGCATATGATATTGTCAATCCTAGATCTACATCCTTTC
TTCACGGAAGTAAAAATCATCATAGAAGGGAATTCCAATCAAGCTTCTGCCGTAAAAATT
GCATGTATCATAAAGGAAAACATCACGGCTAATAAATCTATACAGGTGACTTTTTTCCAC
ACCCAGACCAAAATCAGATAGCACAGCCATTTTACCTACTCGGTAAAGAAAAAAAACCTA
GCAGTTGAGTTTTTTATCTCGAATTTTAAATTCAGGAAACATCAAGGCGTCACAGGAATTG
ATATCGTTTACAATAAAAAATTACCTACGATCCCGTGAATACGCACTAGAACAAATCAGA
AACATTCACCAGATCTCCGTTAACAATTACATAACATACAGCGCTAAAAACAAGCCTGT
TCCGACGACTTAATTATCGCTATAAATTATGGCGATTTATGTATGCTCTGGAACAGTTCA
GCCTCTTTCGAGAGATTTAA

>U62

ATGAATTCGGCGTTAAACGATATAAAAAGACGATTTTTGAGAACTGTGAAACGAAAGACGAC
CTTTTTAAAAATAATTGATAAAAATAATCAAAAATTGCGATTTTATAGTGGAACAGGTCGAG
TCTTTGCCCTCGGAGGGTGGATTCAGCGGCCATCCTATTTGATAATCTCGCGGTGGAGATA
TTTAAACGATGTAATATATCGACAAAATGGAGTTGCCGCGAAAATACGACAGGGTAACGGG
CAGGATATTGACACATAA

>U63

ATGGAGTTGCCGCGAAAATACGACAGGGTAACGGCAGGATATTGACACATAAGAATAAC
CAGATGTGTACAACCGAATGTTCTCAGATGTATAATTTACACAATCCTATCACGTTTGAG
TTGGGACTTGAAACGTGTTTGTCTGTATGCGGTGTTTACCGTTTACCCTGTGATATG
CAAACGACTGTACCATTGTCAACACGCATGAGGGGTATGTTTGTGCAAAAACGGGTTTA
TTTTATAGCGGTTGGATGCCGTGCGTATGCCGACTGCTTCTTAGAACCAATCTGTGAGCCG
AATATTGAAACGTTAATGTCGTGGTGGTCTTGTATCTTATGTATATAGTTTTTTGATG

GAAAACAAGGAACGTTATGCTGCCATTATTGATAGCATTATTAAGATGGGAAATTTATA
AAAAACGTGGAAGACGCTGTGTTTTTATACTTTTTAACGCGGTTTTTACTAACTCAACTTTC
AATAAGATTCCCTGACGACGATAAGTCGGCTTTTTGTTTCAGTTGATTATAGGAGGACAC
GCTAAAGGAACGATTTATGACAGTAATGTAATTCGTGTGTCAGTCGTGCGAAACGAGAAGAC
AGTTTACTAAAAAGATGAGATTGGAGTATGGAACGCACTTATACTATGA

>U64

ATGGAACGCACTTATACTATGACATCCTGTATCAATATCAAGGCGGGTGTATCCGGCT
CATATTTGCCGCGCACAGATGTGTGCTTCCGATGAGAGTGGATTGTATCGAGTCTTTA
TATTTTCGGTGTGTATTTTTTAAGAATGGGATGCATTATACTGAATGGAGTAAATTAAG
TTTACTGTGATTTACAGTGAAATAAAGTTTTAAAGATGTGTTAAAGGATGCAGACTCTGAC
GAAGTTTTTACCGGTTTGGTGGTAATGACTATCCCAATTCGATAGTAGATTTTCATTTT
GATATCGATTCTGTAATTTTGAATTTGGTTTATCCGCAGTTAGTGCACCGGAAATAGTG
CTGAGACTCTACGATATTATATGCGTCAGACCTCCGTCAGACCGGCCGTCAGAAGCATCA
GCTAAAAATATTGCTAATGATTTCTATCAACTAACCTCACAGGGAAATAAACAGACACCC
GATGAGGAAAAACGTTGTCTATTTTTTTCAGCAGGGACCTTTAGAGCCACCCTCTACCGTC
AGAGGCTTAAAGGCGCCCGTAATGAAAAGCCAATACAATTTCCCGCCATGCTAACGAA
AAAATGACCGAATCTTTTTTAAGCGATAGTTGGTTCCGACAAAAAGTCAGATGCAAAAA
ATTTTGGATTTTACGCAAACGTATCAAGTCGTGGTATGTTGGTACGAGCTTTTCGTTTTCC
CGCGAGATGCAGATCGAGAATAATTTACTGTCCGCTTCCAGCTAAAGCGGGTTAACGCT
GCGGATTTTTGGGATAGAATAATCGGTATTTGCGAGATATTGGAAGCAGGGTATTGACA
CACATCGTAAAAACGCTTCAGATTCATAATAGGCAATTTAAACAGAAATTTAATTGCAAT
TTCCAGATAATTTAGCTTTGATCGTCTATTATCATTATGCAGCTCGGGAAAGATTTT
TGGATTTTAACTTAACTTTAGACAGCTGCATTATTAAGGCAATTATCTGTTTCTAGGT
TTTCAAAACGGGGAAAAATCTTTTTTAGCCCAAGATGAAGTTTGGGGGATTTAATAGAC
TGTTCTAAAGGATCGGTGATCTACGGGAAAAAGATCCAATGGATTTTGGACTCGACTAAC
AATTTATATTCGACGTGTGCGGAAAAACAGAATAAGTCGTGGGAATTATATGTTGATTGC
TGTGCTTTGTATGTATCTGAAAAGTTAGAGTTGGATTTTGTGCTACCCGGCGGATTTGCA
ATCACCGGTAAATTCGCGCTTACTGATGGCGATATCGACTTTTTCAATTGGCGATTTGGG
TTATCTTAG

>U65

ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAATTTTCTGCAGAAT
GAATGTAACGGTTCAGAATTTGTAAAAAACATTCTATCGCAATATCGTAGCGTTGCG
ACATCGTCTCCTACATTTCTCGTTGAATAATAAGCCTAAGAAATTTTGCATGCATTGCGAG
ATTGTAATACTCAAGCGAAGTGAAGAATTTATGTTTCAGCCTTGCGGTAAATGGCATACAT
TTTGGGCAGTTTTTAACCGGAAAAATGAAATTTAATAAGAAAGCAGTTCCGGAAGGGCTC
TATTACTATATATTGGAATTTGGGAAGCATAACCCCTATCGATTTGGGCTTTATTCCGAGA
TATAATTCGACTGTGTTACAAACATGCGTTGTGTTACACCGGAGGTTATTTATGAAAAT
TGCTCTATTGTGTGTCCCGAAGAGGCAAATCGCCTCACGGTAAAAGGGTCCGGGGACAAT
AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT
ATCTATACTTTTGCCTCGCTTACGATCTTTTCTAACTTGTATGACAAATCCACCTTT
CCATCTCTGGCGAAAATTTATTTTTGATATGATAGCTTGCGAATCCGAAGATTGTGTCTTT
TGTAAGATCACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTAAT
CAAGAAACCTGTTTTTGTACACACCGTGTAAAGAAAAAATGGCTAATATTAACAATCCG
GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCCAA
ATAAAAGCATCGTTATCACCTGGACATTAATTTACGCTCATGGGTACTTCGGGTGACGAC
CCTTATGCGTTAAAATGTGTTAATTTGGATAACAGTCAGGATCAGCGCAGCTCTGAGCAGG
CTGATTGTTTTGTGCTGTCTGTGTGTAAGCGTGTGGTAATGGACTAA

>U67

ATGTGGGTTATGTGCGAGGTACGTAGCATGGAGCCCGACCTTACGTTGGCGGCGGTCTAT
CAGGCGGCGGCAACCTCACAGAGCAAGATAAGGAGATTTTTGCCGAAGCGGTAAAACT
GCGTTTTTCAGTGTGTAGTTCCGGCAGCCCCGAGCGCTAGGTTGAGGATGATTGAAACGCCT
ACACAGAATTTTATGTTTTGTGACGAGCGTTATTCCTTCGGGTGTGACGTCTGGTGAAAA
AAAACAAAGTTAAATATCGATGCCGCTCTGGATAATTTGGCTTTGTGTTTTGCGAACAAA
AAATCAAAGAAGATGGCTAGAACGTATTTGCTGCAGAACGTTTTGCGGACTCAAGATCAA
CAAGTTGCCATTTCCGGGAAGTACATTTTGTATACAAAAAACACATTGAAACGTCTTTG

ATGATCGATAAGACGAAGTTAGTTAAAAAATTCTCGAGTATGCCGAGACCCCTAATCTG
TTAGGATATACCGATGTGCGTGATCTTGAATGTTTACTTTGGTTAGTGTGTTTTGTGGTCCT
AAAAGTTTTTGGCAGTCAGACAGTTGTTTCGGATACAGTAAGACGGGATATAATGCCGCG
TTTCCAAATCTATTGCCCTCCGTATCTGTACGAATGCGGCCAAAATAATGGACTGTTTTTT
GGCATTGTGCAAGCTTACGTGTTTTCTTGGTACTCAGATTTTGATTTTTCGGCGCTTGAG
ATTCAGAACGCGCTCGTCTGTCGAATCAGGTCCTGTACGACTTAAAACAGAAGTTT
GCGGAGCAAGAAGTTTTCTGTTTTACCGGTAGCGTCACAGATGTGTATCTTTTTGTGCATTA
TATAAACAAAACAACTTAGTCTAGAATACGTTTTCTGGTGACTIONTAAAGACTTCTGTTTTT
AGTCCAATTATAATAAAGGATTGTTTTATGCGCGCAGACGACAATTTCTACAACCTCAGATG
CTGCCAGGTACAAAGAGCTCAGCCATATTTCCGGTATATGACCTTCGTAAGCTGCTCGGT
GCGCTTGTCAATTCGGAAGGTAGTGTCAAGTTCGACATATAA

>U68

ATGTCCTTTGAAGGACTATCTGAGACAGTCCATTTCTAAAGATTTGGAGGTGAGACATCGA
GATTCCTTTAAAGATTAGATTAGGGGAGAGACATCCATTGAGTGTGCATCAGCATATGATC
GCCGCTAGGCAGATCATCAAATCGGACAATGCAGAACAACAGCATGTAATATCTTCTTTG
AGTGGTTTTTTGGATAAGCAGAAGAGTTTTTTAAGAGTGCAACAAAAGCTTTAAAGCAG
CTAGAGAAATTGGACGTCGACGAAATAATTGATACGGCAGCGGAAGTGAAAGCAGTCAGT
AATGATATAAAAAGAACTCTTATAGCAAGCACTGAATTAGAATAA

>U69

ATGGACAACGGTGTGGAGACACCTCAGGGTCAAAAACTCAGCCGATAAAATTTGCCACCT
GTCAGGAAAAAGTTAAGAAAACATGAGGGACTCGGAAAAGGTGTTAAACGAAAACTTTTT
GCCGAAGATAGCCCTCCCTTAAAGAAACAGATTTCCGCCTGTAGTGATATGGAAACACTT
TCTTCGCCCTGTAAAGTCTGGATGCGAGTCGCGAAGTGCTTCTCTCGATGAAAGTTTTCGGA
AAATGTA AACACGAAACTGCTTGCATTGTTCTGCGATAGAGGAATTGCTTTGTACAGG
TCGCTTCTAGACTCGCCGATAAACTGTGCAATGCCACACCATCTTCAGCTCAGACAAA
CGGAAACTGGAGCTAGAGAAAATTATAGCTTCAAAGCAGATCTTTCTAGACATGAGTGAG
AATGCTGAAC TTGCGGCC TACGGCGAAACTTTGTGTAACCTGAGAATTTTCGAAAAGATC
AGCTCGCCGTTTTTTGTTTGATGTGCAAAGCGAAGAGCGTTTCGTATTCAGTGGTCTACGTC
CCTCACAAACAAAGAACTCTGTGGACAGTTTTGTCAACCTGAGAAAACCTATGGCTCGAGTT
CTCGGAGTGGGTGCTTACGGGAAGGTGTTTGATCTAGATAAAGTGGCCATAAAGACGGCC
AATGAAGATGAGAGTGTCAATTCGGCTTTCATAGCTGGTGTTCATCCGTGCAAAATCAGGA
GCCGACTTATTATCTCACGAGTGTGTTATTAATAACCTATTGATTTCAAATTCGGTTTTGT
ATGAGTCATAAAGTGTCTTTGTACGTAATGATATTGATCTCCATAAGTTTCAAGAT
TGGGATGTCAGGAATGTAATGAATTACTACAGTGTGTTTTGTAAGTTAGCTGATGCTGTA
AGGTTTTCTAAATCTGAAATGTAGAATTAATCAATTTTCGATATCTCACCTATGAATATATTT
TTAAATCATAAAAAAGAGATCATCTTTGATGCCGTGTTGGCGGATTACAGCTTGTCCGAG
ATGCATCCCAATTATAACGGCACGTGTGCTATTGCTAAAGAGTATGACAAAAATCTTCAA
CTTGTCGAATTAGTCTGTAACAAATCTGTGACATGTTAATCCTGGATTTTCGACCACTT
GTCGCCAATGCAATGATATTGGTCAATGTGTGCGGGGCTTTTGATGGTGAAAATAATCCT
CTTAGACATTTGTAAC TTGGATCTGTGTGCTTTTGCCAGGTTGTATTATCGTGTGTGCTA
AGAATGACAGATAAACGTGGATGCCGCGAAGCTCAGTTATACTACGAGAAAAGGTTGTTT
GCC TTGGCTAACGAGGCC TGTGATTGAATCCTCTCAAATATCCATTTGCTTACAGGGAT
GCTTGTGTGTAAGTATTGGCTGAGCATGTAGTATTACTAGGGTTATTGTTTTACCGAGAC
GTAGTTGAGATATATGAAAACTATAACGATTTTCTAGATGAGAGAGGGGAATTTGGGTCA
CGAGACCTTTTTGAGGCACTTTTTTTAAATAATAGTAACTTACCAGACGTCAGCCAATC
AGAGAGGCTCTCGGCTCTTACAGTCTCGGAGTATGGAGAAAAC TTTTACATGACCTT
AGAGAGTTGTTCTTGATCAATCTACTGCGGATCTGGATAAAGATAATCATCTCTCTTT
CATATGTGA

>U70

ATGGATCTTGATCAAATATCTGAAACACTGAGTTCTGTGGCCGAAGAAGAGCCTTTAACC
ATGTTTTTACTTTGATAAAC TGTATGCAATACGGGAAAAGATCAAGCAAGTTCCATTTTCG
ATTGTTTCGCTTGTGTCAATGTTTACTGCATGCTAATAAAAATATAACGCTTCTAACAATAAT
TGCATCTGGGCCGTAAACTTATTGAGGAAATGCAGCAGTTTTTTGTGCGGCACAAGAGTG
GATGGATCGGAAGACATTTCTATGGATCTGAGTGAATTGTGCAAGCTCTATGATTACTGT
CCATTTATGTTCTGCTCTGTGTCGTGCGCCTTGTGTATCTGTGAACAAGTTATTTAAA

ATCGTAGAGCGTGAAACTCGGGGGCAGTCGGAAAATCCCCTCTGGCACGCGTTACGGAAA
TATACCGTGACGGCGACTAAGTTGTATGACATCTATACGACAAGAAGTTTTTTAGAGTAT
AAAGGACAGCAGTTTTTTGGGGAAGCGGTGATTTACGGCGCGAAACATGAGCGTGTTATT
AGACACCTCGTAGCGACCTTTTACGTTAAAAGGGAAGTTAAGGAAACGTTGGGATTGCTG
CTTGATCCTTCTTCTGGAGTTTTTGGTGCCTCTCTAGACGCGTGTTTTGGAATTTTCATTC
AATGAGGATGGATTCCGTGATGGTTAAGGAAAAAGCGCTGATTTTTGAGATTAAATTCAAA
TATAAATATTTACGGGATAAAGAAGATCACTTTGTTTCTGAACTATTA AAAAACCCACG
GAGAAATCCTTTTCGGATTTTATTTTATCTCATCCGGTGCCTGTCATAGAGTTTTCGAGAA
AGAGGGAAGATTCCGTCACTCTAGAGAATATTTAATGACGTATGATTTTCAATATCGTCCT
CAGAGAAAGTTGCGTACTTGCCCACTCCAGCTATTTTGGCACCTCATATCAAACAACCTG
CTGTGTTTGAACGAAACACAGAAGTCTACGGTAATCGTTTTTGGATTGTAAGAGCGACTTG
TGTGAGCAGAAGCTGCTGTGTTTCAGAAGGCTGTGTTTACTGTGAATGTATTTCGTAAAT
CCGAAACACAGGTATTTTTTTTTCAGAGTCTGTTACAACAATATGTAATGACTCAGTTTTAT
ATTAATGATCATAATAATCCAGAATATATCGAAAAGTACGGAAGTGCCTTCTGTTCACATT
GTGACGGCTTTCTTCAGGAGAAGAACAGAGGAAGAAAGGAGTTTGCATTTGGTGATTGAT
GAGACGGAATATATAGAAGAAGAAATACCTTTGGCCTTGATTGTGACTCCGGTAGCACCG
AATCCGGAATTTACTTGTGTGTTATAACAGACATATGCAATCTGTGGGAAAATAATATT
TGCAAGCAGACCAGTTTACAGGTATGGGCGCAAAGTGTGTAAACCAGTATCTTGCGGCA
TGTGTAAGAAAACCGAAAACACCTTGA

>U71

ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAACACC
TTGATTGATTACAAGGGAAATCCTATATCACTAGCGGATGAATTTACTGTTTTAACCGAT
ACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGA AAAAGCCCTTACTTGAAAAGTACAG
ACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAATCAAAA
AAATGA

>U72

ATGGCCTCCAGTCGTGTAGACACCATTAATTTACGGATCTGGCTGGTCTCTATAATCTGT
GCAGCCTTGTCAATTCATAAACGTCACGGTCCATCTGATAGCCATAAACTTTCCAAATTTA
GGATTTCCGTGCGCCTACTTTGAAATTAATGACCTCAAAGCGGTTAATCTCTCCGCTAAC
AATCAAATATATCAGATGACGCATCAACTTTACATAAACCCCGTGCAGATTATATGCTAC
GTACTGATAATGGCTATGTTATTCCTCCTGATTATTATTTATTACATTGTCTGTGTGCA
AAAGTGTTCCTCTAACAAAACATCTAACGTTAACCAAACAACGAGGGATATTACCTGG
ATGGGGACACGTCATCCTGCTTTCAATTCATCTTAATCATGGATACATTTCAATTATTC
GTCACCGCCTTGTCTTTTCAGACTGGTAGCCCTAGGCGCATTTGCCTATTGCATCTTCTTC
GTGTGTTTCACCACCTTTAACGTTACGCTGATAACGCAATTTCAATCAGCGGATAAGTCA
TTTTTCGCCTTTTCAGAAAATTCATCCTAACCTAAAGGGGACCGTACAATTC AAGACCGTT
GTTATCAATCTAACCGAATCATGCTAGGTTACTCTACAATGTTTCTCGGCATTACTACT
TGCTTGGGCGTCGGAAATAGTATCTACATACGTTCCATAACCGTTGCTTACTCTTCCATC
AATACCTTTCTTGTGTCATGGCTTGCATATATAGTATAGTAATAGAAGCCGCTCTGGTTGCG
TACGTTAAACCTTGTTCGGCTATTACGTCGGTATGTTTTGTGGCGCCGTAGGCCTATCT
TTTCCATTTCTCCAGTACGAAACTTTCTTTGAATCCGAGTGGAGCACAGGACTTATTATC
AATTTAGCGGTCACTCGCCATAATTTTCGATCGGTTTTATAATTTGTGCACTTGTACGCTAC
CTAGTTAAAAGAAGCGTCGGTACAAGCAACTACTTAATACTGAAAGCTCGAGCCTAATG
GATGAAAACGAGTGA

>U73

ATGGAGAACAGGTTTTCAGAACGATACTCTGTTCTCAGAGTGGTTTCGGTCAGAGTTTGTCC
GATGTAAGATTTCCGGACAATGTCACCGTTTATTCTCAAGCCGACTCGGCGGTGAGTTTT
GAAAACGTCAGACAGCCGATAAAACTGGTTTCGCGCAGCGATGGGTTTCAGGAAAAACA
GCCTTGATTCATTTTTTGAAACAGGTACCCAGGGAATCTGTCTTTTTGATTTCTTGC
CGTAAAACGTTTCGCTGCTGAAATTTTGCACCGTTTTACTCTCAATGGGTTGAAGGATTTT
GAATTATATTGCGATATTACGGAGCGGCAGATAAACAACAGAAAAGTCATTGTCCAGATA
GAAAGCTTGCATCGTCTTACAGAAAATTACGATGTTTTAATATTAGATGAAATTTATGTCG
ATTATAAAGCAGTTTTTATTCAAAACGATGACAAAACCAAAGAAGTTGATTGTAAGTTT
TTATCGCTGATCAAAAATTCAGTCAATAGCCATGGATGCGACGCTGACTCGGCAT
GTGGTGGAGTTTTTTGCCGCTTCAAGCCAGATACTCAAATCGCCCTCATCAGAAATACA

TTTGTGTCCGCGATGTTTTCCAATAGGATCGCTTACTTTTTGTGACACGTTTTTTGGTAAA
GAGTTTTCTTTTTTGCACGCCTTGGAAAGACAAGCTACGATGGGATAAGAAATTGTGTTT
TTTTGCAGCACAGTGTGGCAGCGGAATACATGCATGATTTGATACGGTCTCGGTTTTCCA
CTCAAAAAAGTCTCTTGTGTTGACGTCTAAACAAGGGAAATGCTCTTCCATAGAATCGTGG
ATACGGTATGACGTAGTGATTTATACGTCCGTCTGACTGTCCGGTCTCAGTTTTCGAACCG
GTTTATTTTCTAGCTCTTTGTTTGTTCATTTCAGTTGGCCAAAGGAGGGCCGGACATGGTT
TCAATATTTCAATCCATTGGTCTGTGAGGCGAGTTATCGATGAAGATATATACATTTTAC
ATGAATCCCGTATTAATTAGAAGCTATGATCCTTTGGCACCTATTGCCATACCTCCGTGC
AGTGATTGGAGTGTCTGCTGAGCAGTCTATCATCTCAGAGAGTTGCATCGACTTCCGCGGA
AAGTGCAGCGGAGCGCATAAATAAATTTTTGTTCCGTAAGTCTCTTTTTTCGGTAC
AGGCATTATATCGAAAAGACTACTATTACAAGTTTATCTGATAGTTTATTTCTGCTATGC
AGTCTGTTATGTGAGAAATCTATTAAGGTGGACATCGTTGGAAATGGATTCCCGATGCGC
AAAGAGGTTTTCTGAGTTTCTTGCAGATTCTAGTTGAGGAGTGTCATTTTATCGAAAAG
AAGATTACTTTGCCGGGTGACGACATGACGTTCCAAGAGATTATAAGTAGCAGAGAAACT
ATCATGAACGGGGATTTTTATGAAAACGGTGATCAACTATTACATAAGGACTATATTACC
GACATGGGAAAAGTTCAGAGCGACTTTCTTATCTCCGGGCGTTGACATTTTCATCGCATCG
GACATTGTATATGATCTAAAAAATGAAAGCAAACGCTATGTTTTTGTGAACGTCTGGTTG
CAAAAATGTGTATCTGCCGGTGTGGAAAGCACGGAAATCGAAAGAGTTTTTTGTGAGCGG
ATTAAGTGCATATATGCTGCCGAAGAGTTTTCTGTGCGACGAGTACTTTGTTCTAGGTGAC
ATTTCTGGGGTTTACGAGTGGGGAATGTTAATAGATCTGGCCTTCTGGCGGAAATGATC
CGAAAAGATTTGAAACTGAAATCGTGCACGGATACTACTACTGACATTTCTGAAGATGAT
CTCCTTCTGTGTGCGGCCAGGAGGTCAAGTGACATCTTACAAATAATGCAATTAGTGT
ACGGTGCATGTGCAATTTTTTCAAAAATACAGTCTGCAGACTCTGCAGCTTTTTAACAAG
TTGAGAGGCATGCGGATTGTGACGGGTGTGTTCTCTATAGAAAATTCAGCATTTCCATC
CTCAGGCTTTTCTTTAAATGTGCATTTAATATGACTCTGTCCGCCAGCAGACCTCGGTAC
ATTCCCGGGAAGGCGTATCGCAATCTAACAAAAAATGATATGGAGAACATGCTAGACAAT
TGGGAGATTTGCGGCACCAATCTTAAGACTTGCAAAGAACTACGTAAAGCCCTCACCGAG
GCCTCGAGAGCAAGGCGGAAACAACAATTTATAAACTGCAAGGCTCGGATATAAGCCTC
TCCGTTAGTGAAGTTGGAGTTTTTCGGGCAACATGCATCTCCGGGGGTGTGCGTGTATCT
TAG

>U74

ATGCATCTCCGGGGTGTGCGTGTATCTTAGCCTTTATTGTGTGTACAATGACTGGGAG
AATAAGATCTACAGAGTACCGATTTTTTCAAGTGTCTGTTTTTGAAGCGGAGACTCGTTG
TTAAAAACGTTTTTGTATTAGGGGACAGAGCCTAGACCAAGAATCGTTAAATGAGATAGAA
GTTACAAGAAAAGAAACAATGCTTTGGGACCTTCAGGAACAAAGTAATATGATGGACAAA
AAAATAGCAGCAATCAGCAACCTCATTATGAATAACGGAGAGCTTGTAGGAAACTTTCA
AAATTTTTCGTGCCCTTAACCGTTGTTCTTGGAGATGACGGGTAGAGATTTTGGAAAGCA
TACGTCTGTGGCGAGGAATCATGTTGCCTCTGGACACGGTACCTGTTATCTTACGGTGT
ATCGGAGACTACGGGCCCTTGGACACGAAGCATCTTTTGGTAACGAATGTACACAGGCT
TCCAAAAAATTCGCTTTGGATACAGTGTGATGGATTTTCACTTTTCTGTTGACTGTGTCT
GACGTGAAAATTTGTTTTTTCACACACGGACACCGGTGAGGCCGTATGTGAAAAAATGAAA
CAAAATTTTTCTTTTTCTGTGTGTGCTTTTCGGCGGTGAGCAAGTTCTCTTAGTGACGCCA
AAAAATGCTTATGCCTTATTGTTTCGACGATGATTTGTGTTTGCTTTTACTGCAGAGCGTG
TTTGCCTTCCGACGAGAAGATATTCGGAGTTTATAAACAGGTGTTAGTTTCAGCTTTGT
GAATACATCGGTCCCGACTTATGGCCCTTGGGAATGAAAGATCAGTTTCTTTTATAGGA
TATCCAAATCTGTGGCTTTTGTCTGTATCGGACTTAGAAAGGCGGGTTCCGGATACAAC
TATATCTGTGCGGAAATCTCTCATTTTTGCGGTTTGGCGCCTATCCTGGGACCTCGTGGT
AGACACGTGGTCCCGTAGTACGGGAGCTAAATATTGAAATGCCTGGAAGTGAACCTTCT
TTGCAGCGTTTTCAGGTTCAATAGTCAAGTATGTTTCAAGCGAGTCCCTGTGTTTTTCAGACA
GGTCCCAGGACACACATCTATTTCTTAGTGACAGTATGATATGTATGTGGTGACTTTGCCA
GATTGCTGAGGCATATTATTGAAGTCCACGGTTTCTAAGGCCTTTTTTACCATGTTTTGAT
GAAAACGCCACGAAATCGATCTGCTTTTAAAATTTATGTCAAGGCTGCAACACAGATCG
TACGCTTTATTCGATGCGGTGATCTTTATGTTAGATGCTTTTTGTTTCCGCTTTTTCAGCGA
GCATGTACCCTAATGGGAATGCGATGGCTGTTAGTGCGGGACTTACACATGTTTTATTTA
ACCTGTGATGGTAAAGATACACATGTTGTATGCCTTTACTGCAAACGGCCGTAGAAAAC
TGTTGGGAAAAAATACGGAAATCAAACAAGACCTACATTTTCAAGTGTGCGGAAATTTG
CGTGTGGATTTATCGTCTACGCCAGATTTTTTTTTGAGCAGTGGGTTGTACAATCTAAA

GAAGCGCACTGGACAGTAACGGCGAGTAAATATTTATCTGCATGTATTCGGACCAATAAG
ACAGGTCCTTTGCTTTGCCAGTATAACTGTTTATTTTCAGGATATGATGTGTGTTTTTATA
GCTAATAGGTATAATGTTTCTTATTGGATCGAAGAGTTCGATCCTAATGATTATTGTCTG
GAATATCATGAAGGACTTCTGGACTGCAGTAGATATACGGCCGTGATGTCTGAAGATGGA
CAGCTTGTGAGACAAGCACGTGGAATTGCGTTGACTGACAAAATAAACTTTTTCTTATTAC
ATTCTCGTTACACTGAGAGTGTGAGGAGATGGGTGGAGAGTAAATTTGAAGACGTTGAA
CAGACACAGTTTATTAGGTGGGAAAACAGGATGCTCTGTGAACACATTCATTTACTGCAT
TTGAATTAA

>U75

ATGAAATCCTTGAAAAAATTAAAAGAATTAGAAACCTCAGACGTATTCAACATCTTACAC
GTCCGAACATATCCTAAAGGTGATCAAGATTGATAAGTGTATATCACTCGCACGACATCCG
TTAGTAAATATAACAGTAGGAGATGACGGAATATGGTTCCACTTGGAAGACGGAACCATG
ATCAACGGCCTGGAGTACAAGACCATCTGCGAAAAAGAATTGGGATTTCAAGGCTTCATT
GGCATCATAATACTAGACTCGGAAGATACACTGCAAGAACTGCGACTCAATCCCTTTTCAG
TTCAAACGCAGACTAATCCATATGAAAGTTGATACGCCAGAAGAATTTATGCTTTGCGGA
CTAGTGTTCGCATTAGAAAATCTTCCACTGAAGCAAAGCACCTACACAAACTAATCGCT
AAGCTCGTTCTCTTTCCAGCCCTATCTCCGATAACGAAAATATTATTTAACACGTGTGAC
ACGTTGGTTTGTACACTAAGACACATTTTTTTTTAACGAACACGCCTCAGAAATCCTCCAC
AAGGTGCCTCTGATGATTAGACTATACAATGAAATGAAAAACACGCATATAGAAGTTTTG
GAACTTTATTTTAAACACAAAACGTAGTCACAATTTTCATCAATTTATCATTAGAATCTAGA
CAGCTGCAAGACAGTTCTCTCCAAGTCATACAACCTAGCAACACAATTCGCACAAACTTTT
TATTCGAAGAATGGAGACACATCTAGCTAA

>U76

ATGATGTGGCGAAACCAATCTTCCGGCAGGCGCCCTCCAAGCGTCCTTCTGATAACGAA
GCCACCTTATCTACAATCAACTCTATTTTGGGAGCCGAAGACATGTTAAGCAAAAACCTTA
TTATCGTATCTACCTCCGAACAACGAAGAGATAGACATGATTTATCCCTCGGAGCAGATA
ATGACTTTTTATTGAAATGCTACACGGACACAAAAATTTTTTCAAAGGTGAGACAATCCAC
AACGCTCTCCGAGATTCGGCCGTACTGAAAAAACAAATAGCCTATGGAGTAGCGCAAGCC
CTTTTAAACAGCGTTTCCATTCAACAAATCCACGATGAATGGAAACGTCATGTGAGGTCA
TTTCCGTTCCACAACAAAAAATTTGTCATTTCAAGATTATTTTAGCGTATGGGCCACGCG
ATTAAACAGGTGATCTTGGGTGATATATCTAATATAATTAATTTTATTCTTCAGAGCATA
GACAACAGTCACTATAACAGATATGTAGACTGGATCTGTACTGTAGGAATCGTTCCGTTT
ATGAGAACCACCCCTACGGCCCCGAACCTTATACAATCTCCTACAACAAGTGTATCTAAA
CTAATCCACGACATAGTGCGGCATAAACAAAAATTGTAACCTCCGTTTTACTGGGTCTA
TCTTCCGTGATCATTTCCGGATTTTCACAATATTAAGATTTTCCGAGACCGAAATTCGGAG
CAGATTTCTTGTTTTTAAAAACAAGAAGGCTATAGCGTTTTTTCACGTACAGCACCCCATAT
GTTATTAGAAATCGCTTAATGCTGACTACGCCCTCGCACACCTTTCTCCAGAACTAAAA
AAACACAATAGTCTGCGGCGTCATCAAAAAATGTGCCAATTATTGAATACATTTCCATA
AAGGTGCTAACACAGCAAAGACAGATGTAACCAACAAAAAGATAATGGACATGATAGAA
AAAGAAGAGAAAAGCAGCGACGCGAAAAAAGTTAATCAAATTTTTACTCAACTTATCG
GACAGCAAATCTAAAATAGGCATTAGGGACAGCGTTGAAGGATTTATTCAAGAAATCACT
CCCTCTATAATAGACCAAAATAAACTCATGTTGAACCGAGGACAGTTCCGAAAACGCTCT
GCGATCGACACAGGAGAGAGAGACGTTTCGAGATCTGTTCAAAAAACAAATCATCAAGTGC
ATGGAGGAGCAGATACAAACACAGATGGATGAAATCGAAACACTCAAAAACAACAAACCAA
ATGTTTCGAAAGGAAAATCAAGGATCTACACTCCTTGCTCGAGACGAATAACGATTGTGAT
CGCTATAACCCGAATCTGGACCATGACCTCGAGAATCTCTCTTTATCTCGAGCGCTAAAC
ATAGTTCAAAGGCTCCCATTCAGCTTTCCATAGATGACACCCGATCAGTCGCGAAC
AGTTTTTTTTTACAGTATATCCAGACACGCAATATGCCGACAAACGAATAGACCAGCTA
TGGGAAATGGAATACATGCGAACCTTCCGGTTAAGAAAAAACGTTAATAATCAAGGACAG
GAAGAAAGTATCACATACTCTAACTATAGCATCGAGCTGCTCATCGTGCCTTTCCCTAAGG
CGATTTTAAACATTTACAACCTCGAATCGATTCCCGAAGAATTTCTCTTTCTATCCCTC
GGTGAATCTTTATTAGCCATCTACGAAAGCAGCAAAATCAAGCATTATTTAAGACTCGTA
TACGTCCGAGAGCTAAATCAGATATCAGAAGTGTTTAATTTAACCACAAACGCACCCCGAA
AACAGCGAACCGATTTTCGATTCGAACATTTTCTCCCAAAACCCGGAGAATGAAATCCTT
GAAAAAATTTAAAAGAATTAGAAACCTCAGACGTATTCAACATCTTACACGTCCGAACAT
CCTAAAGGTGATCAAGATTGA

>U77

ATGTCATCTCTTCGTTGTTTCGGAGGTAGATACGATAATAAGTTTTTGGCTTAACATGTCT
TCGGCTCCCAAAATAGAGTTGATTGTAGATAAGGTGGCTTCGTTATCAGAAGGACGCTTG
GAGGGGCGCCTGCCGGAAGATTGGTTTTCCGCACATCATGGATCCCGAAACCGAGTTCAAC
AGCGAATTTGCTGATGCTCTGTGCATAGGTATTGATGAATTCGCTCAACCTCTGCCTTTT
TTGCCGTTTTAAGGCCTTGTTAGTAACCGGTACGGCCGGGGCGGGGAAAACGAACAGCATT
CAGACCTTAGCGGCCAACCTTAGATTGCATCGTTACTGCAACGACATCTATCGCCGCTCAG
AATTTGAGCGTTGTTTTAAATAGAAGTAAATCCGCGCAAGTTAAAACAATTTTTAAGACT
TTTGGGTTTTAATAGTTCACATGTATCTATGAGCGAACGTCAAAGTTATATTGCAAATGAC
GAGAGTTCGATTCAAATTC AACAGAAGCAAGACCTGTCCATCTATTGGAATGTTATCTCC
GACATAGCGGATAGGGCGTTGGGTGCTGTGCGCATGCAAACTAAAGAGTTACCTGATTTG
TGTGAGAGTAGTGTATAGTTATCGACGAGGCGGGTGTAAATATTACGACATATTCTGCAC
ACTGTTGTTTTTTTTTATTGGTTTTATAACGCGTTGTATAAAAACCCCTTTGTATGAGGAC
GGAATTGTTCCGTGTATCGTGTGTGTGGGTTCCGCCACGCAGAGTAATGCTTTGGTGACT
TCATTTAATCCGCTGACTCAAATAAGGACGTGAAGAGAGGAATAGACGTACTTTCCGCC
TTGATTTGCGACGACGTGCTGTCTAAATATTGCGAAGTCGATAACAACCTGGATAATCTTT
GTAACAACA AAAAGGTGTGCGGATCATGCGTTCGGAGATTTTTTAAAGCACATAGAATTT
GGTTTGCCGTTGAAGCCGGAGCTTATAGAGTATGTGATCAGTTTGTGAAGCCCGCTTCG
TACATCAGAAATCCCATGAACGAAATAGAAACGACCCGCTTGTTTTTATCGCATAACGAG
GTAAGAACTATTTCCGTTCACTGCATGAGCAGGTTGAAGTTACCAACAGAAATAACTTG
TTTGTGTTTCCGTTTTATTTCTGATAAAAAACAAGACTTTTCGAGGACTACAAAAGTGAA
ATTGGCAATTTCTCCTTAGAAATTAACCGTGGTTTTAAATCTAATATTCATCGGCTGAAT
ACTTATTCAGTTCGCGGATCAGGATTTATCGAAAACGGTTCAATTGGAGGAGATTGTT
TTGGAAGATGGTTCGGTGAAGAGACTTTGATAACCTGTCATCTGAAACATATCAGAAAC
AGTTCAATAGGTGTAACGTCCAAGATCAAGGCTTCGACCGTGGGTTTTTCCGGAACCTAT
GAAAAATTTGTAGAACTGTTACAGAGCGATTTATTTATCGAAAAACGTCATGTGATCAA
ACTATCCACGCATATCTTTTTTATCCGGTCTTATGTTCCGGAGGGATGTATTCTTTTTGC
TGTTCGAAATTCACCACCCCTGAAGTGTTAATGGAAATCAAAAACATAAAGATGCCAGT
ATTGAATTCCTTAGAAAGTGAAATGTGAGGATGTCTCCGGACGTACAAACCGTAGAGACA
GATGAAAGGTATGATTTCCGATTAGTCGATGATGGGTTATCGGATGTGGATCTATTAGAG
ATCGATCCTTCCGGAGATCCATTTTTTACTAGGTATAGTAAGTTGCCTTTAACTAATTCT
CTGTCGTTCCGAGGAAATCAGTTTGTCTTATACGACGTTTAAAGGATATCTTTATATCTAGG
TTTGCTATTTGCGAAGCATACAAAAGGGAAGTTCCGAAAGACTCTCCTAGTTACTTAT
AACAGAAACAACGTTTTCTAGAAAACAGTCCGGTGAGATCTATTCCATTTAAAGAGTTTT
TATGGTATGCTCACTTACGCTATCCCTGCCAATAATTATACTCTGGAGGGCTATACAAAT
GATAACGTAGTTCAATTTAGGAACGGATAAAACAATTACCGCAGATATTATATAAAAAAGGA
TTACCTAGATTAGTTATAAAAAGATGAAATGGGGTTCATTTCTGTTTTAGATAATAACGTG
TCGAAATTTATCGACGTCGTAATGGTCAGAGTTTTTCATCTATGTACAACGGTGGATTAC
GCCACTGTTTTCTAAAGTTTCTATGACCATCACGAAAAGTCAAGGGCTTTCCATACAAAAG
GTCGCAATAGATTTCCGGAGTGATCCGAAAAATTTAAAATTGAGTTCTATTTATGTTGGG
ATGTCACGAGTAACCGACCCGAATAATTTAATAATGAATGTTAATCCTTTGAGACTGAAT
TATGAGAATGATAACTTTATTGCTCCGCATATTGTAAAAGCGTTAAAGAACGAAAATACC
ATGTTAATATTTTAG

>U79

ATGTATGCTGAAGAACGTGCATATGGATCATTTGACGATGTTATGGAAGTTTATCAACAA
ATTGTTACCGAGTCTATTCAGCTGAAAAGGCTTCATTTCCGCAGTGGCTGTTTAATTGAG
TTCTTAGGTGACTCTGGAACGTGTGAAACATTTTTGTAGTGGATGGATATCAATGATTTGT
TGGACATCCGATACAAATTC AATGGGTTCTTTAACGGTGGATATTGGGATAGATGATGGA
AAATATAAACGTACAGTAGTCGGGGTTTTTTACTGTGTTTCAAATCGATCACGTCGATT
TCTCAGAATACGAAGGTAGAGATAGGATTTTTGACTGTTTTCCAAGAAGGCAACAACTT
CAAATAACGTTTTGTACTCTCACTAAAGTTTTCAAAGAGCAGGATATACGAAATCTGGGC
GATGTGAAATGTATGGAAGTTTGA AAAAGAATGTGAGCTTTAGACAGAAAAAACAC
GACGATGAACACAGAAAACGCTCCGGGAAGCAAAAAGAAAAAGAAAAGTGGAGGACACG
GACAAAAAAAAGATGATGACAGACGGAAACAAGAAGAAAAGAAAACGAAACGACGAAGAC
AAGCAACCGGACAAAAGGAGGAATCGGATGAACTTCTAAGGAGAAGAGACCGAAGTAC
CATGACATGAAACGTAATTTGGAAGAACAAGTACGAGGATGGCATAACACCAACTTCT

ACGACACTCGTAAATGGAGCGGTTGAGGGTGCCTTACCGCCCTGTATTTCTATTGATAAT
CACGAAGATCAACAACATGATGAATTAGACAAGCGCGCTTATGCGCAGGGGACGAACCGA
GAAGTCTATCCAATGAAGATAATTATGGGAATTTTCGGCTTAATAAGTCTTTAGAACAG
CTTAGGGCCAAACTTGTGCAAGCAGCGGCGACATTGTAGAAAAGATCGCTTTTGAAATTG
AAAGAGTGTGGATAATGTAAGGATAATTTAATAAAAAATGAATGTGCTGATGTTACT
GGTCCAAGTAAATGTTAAGTAAAACAAAACATATCGAGCCAAAGAAAACAGATAGTGT
TCTGATTGTGTCAGGCCGGTACCCGTGTGTGAGATTAAGCCGTTTATCGATGTGCGACTA
TTTGAACGGCGCAGAACCCAAGAAGACTTCGACAACGGACCAGGACAATCGTCGGATCA
ACAGATGGTGCATCGAGCAACAGCGTGTATTTTCGGGACAAAACCGAGGCAGAGCACGA
GGACGTGGCCGAGGGAGAGCTCCCAGGAGACGAAATTCGAATATAAATAAATCAAGGACA
CAAAC TACAATTGTTATAGACGATAGCAGCGAAGCCGAAAAC TTTGAAAATGAGGGGAGT
TTAATGAAGATTTACTGGCAACCACAATATTGAAAACACTGTAA

>U81

ATGGCCCTACTACAGTGGATGTTAGATCACGTTTACAGGATGAAGAAAAAACTGTGAAAAC
CTTCAATAGATGATCAGCATTCACCTTTCGGCATAAACAGAGACTGGTTGAGTTTCTCG
CAACTCTCCAAGTTAGAAATTACACACCTTAAACATGTATATAAATTTGGTGGACAACGAT
AGGGCTCATCTAACGGTCCACCCCTCTCCGGGTAACGTCCACGCTTGGAGTTTTTTATGC
AAACCCGACGATGTAAGGTTGTGATTTTGGGACAAGATCCGTATCCCAGCGCAGAGGT
TGCGGCTTGGCCTTCGGTACAGTGAAGAGTGTCTCGATTCCAGAATCTCTAAAGAATATA
TTAAAGAACTGGAAAGAAGCGTGCCGAATTTTCCCCCTGACAACGGCTGTTTAAAC
TCTGGTGTAGTGAAGGAGTCCATTGCTAAACTCGATATTCAGTGTAGTTTATGGATTA
CCAATGTCTCACGAGGCATTTGGCTGGCAAAACACTGAGCTATAAGATTATCAGCAAACTA
TCGATAGAAATGGACTCTCTCGTTTTCTTATTGTGGGAAACCATGCCCGAAACTCTCT
TATCTAATAGACGCACGAAACATCTCGTCTTAGAAAGTGCACATCCATCGCCCAAAGTG
AAATCTGCAAGGATGCCATTTATTGGTTGCAATCATTTCGTGCGAGCAAATTTATTTCTT
ACGGAGCACGGGAAAGACCAATCAATTGGAACATTTCTGAACGAATAA

>U82

ATGGAAC TTTTACTATTTGTAATGTCACTCATATTA CTAACCTTCTCAAAGCAATACCT
CTTTTCAACCATAATTC TTTCTATTTTGAAAAC TTTGACGACTGTATCGCAGCTGTAATA
AACTGCACGAAATCCGAAGTACCTCTATTA CTAAGAACCAATCTACCAGCCTCCGGTATAT
AATGAAGACGTTATGTCAATAC TGTACAACCCCGACAAAAAAAAGCCTTTTAGCCGT
ATAATGGTAACCGATGAATTTCTCAGCGACTTCTTACTCCTACAGGATAACCCGGAACAA
CTACGCACATTTGTTTGCCTAATAAGGGACCCAGAATCTCGGGACAAC TGGCTAAACTTT
TTCAATGGCTTCCAGACATGTTACCTTCCGTGCGAATAACAACCTGCATCAGAGATAAC
TGTAAGAAAATATTCGCCGAAAAAATTA CGTACGTCAATAACTTTTTTTGTTGATAACATT
GCGGGTCTCGAGTTTAAACATTTTCAGAAAACACAGACAGTTTTTTACAGCAACATTGGTTTT
TTATTATAC TTTGGAGAATCCTGCTAGAGGCGTCAAAAAGATTATTAGGTTCCCTTTAAC
TCTTTGACTCTTTTTGATACGATTTTGAATTGTTTAAAGTATTTCCACTTGAAAACCGGA
GTAGAGCTCGACCTGCTAAAACATATGGA AACCTACAATTTCTAAACTACCTTTCCGAAGT
TCCCGCCCTACAATTTCTGATTAGAAACACATAA

>U83

ATGGCTATCGGATTTATATGTAGTTCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT
ATTTCTGCTTCTGTCTTGTTAGGATGTTTGTGTGTTGCATGGATTGGTCCGCTGCCGTA
CCTGGGAAAACAGAGTCTTTTAGAAAAC TTTTGTATGCAGTCATGATTA AAAAGCTAAAA
AGTTGTTCTGCTGCTTACCCGTCTGATTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT
GCATCGCCGACAAGTCTTGAATTAGGATTGTGCA AATTAGACAAAAGAAATCATGA

>U84

ATGAAAGACAGAGGAAATAAAAACACCTCTAAAACAAGACATACCGTCAAATCTCGCCGA
ACTAGAAAAACAACGAAACTGTTTCGTAGAACGGAAAAAAAATCTACCCACTAAAACGACT
ATACGAAATAATCGAATACAAAACAATCTATCTCCGCAAACCGCCCAACCGAAAAACCC
GTCGATTATGCAACTGCAAGAGTCTTTCCCCAGCACACCCCATCCGATATAAAAATTTAAT
AAGGAAAAATATCGAAATAACGCGATCTACGCCATTCCA AACTACGGA AAAATCTCAAAC
GCCGTGTTTTCGTCAAGAAAAGAAATTAATCGTGT TACTGTATTTACGACAAAACCTCTC
GCCGACCTGTCATTC AACATGCCGTGTA AATACATCGACAACGGAAAAAAAATGAAATTGTC

AGGTCTACGGAATGCCAAAGGCAAGATCGAAATACAGTCAACGCTTTCAAATTTTTCAA
TCTTACGAAACATCTCCACCTTACTCCCCGTGCTCACCTAAAAATTTTATGTGAGAAATA
TACAGATTATTCGGCTGCTACAACGCTTCAATTGTGCAAATACAGGTTTATTTCCCGAAAT
AAGATCTTAATAAACGCCATGCAAGAAAAATTGATGTCCATCGGCAACATGGTTATTAAC
GTGGGCGAGCAGATAATTTCTGAATCGGTTACCATGAACCGGTAGTCGCTATTGCAATC
AATCAGTTTTTTCAGAGGGACCGTCCCACAAAACGGACGCACTCCGAAAAACGCTGTAAC
CCTATACCGGTAAACGCCCTCAAAAAAAAAGCTTGTGCGCATTTGCTCTCTCATGATAAAC
GCGCCAACGAAAAGGATTTATATGCGCCGCTCACGTTTGTATCAATTTTTCGATCTGC
TATTCCAACGAGATTACTTTGAATATCGCAAAGCTAACACAACCGATGCTCCTTCAAGAA
CACATCACCATATACAACCTTTTCAACAAATATATTTATTGGGATGATTATGGCAAAC
TTAAACTGA

>U85

ATGAGTTCTCTGATGCTTCGGTTATTGCCGCTTCTATATATCATAAGTGCACATTTTGTA
CTCCACCCGGAGACGAGTCCCTTCTTTAATATACGAAATAGGGTCAATTGTCACTTTCCAT
TGTGCGCTCAACACGACCACAAACATCCGAAGCGTCTCGTGGTACAATAAAAGCCGACTT
ATATCGAATCATGAGGTACAAAACATGGACAACCTTATCGTTCACTGACGATGGCCACGTC
CTTATACACGAACTAAATAAGATCAACAATCTCGCCGTGGACAGTAAATTATATTTCCAC
ATAAACACAATAGAACTACATCTCTGTTGAAAATTAAGCCAAATCAGCATATGACGCG
ACATGTTTAACTGTACATTCACAGTAGATAACGAGAAAACGCTGCGACTAGTTGCTTA
AACTCATTATGAAACCCATCGTCTGTTTTGTACTTTTCGTTACTTAGACAATTTTTTAGAC
GTTACGTGTACCGTGACGCTTATCCAAAACCAAATGTGGTTATTAATTTCTTGGAGAA
GTATATAAGAGAGACATACCAATAGTTAGACAGAATGAGAACGGATCGTCAACCGTTAGT
GTGAACTTTACTTTCAAAGACGAACGAAGCTGGAATTTGTCGGAAAACCGTCTCATGT
CTAGCGTTCGAGTTGGTTTACCAACCAAAGGCTAGCGCTCTTGTAACTTCCGGCGAACAC
ACTGTGCAGAATCAGACGAACACAGTAAAGAAGCGGTAAGAGGCTCAAATCTGACGAA
ACCGTATTTACCTGGATCGTACCATTAATTTCTGATATTGATATCTGTTATAGTGCTACTG
ATTTCCGGTGTGTATCGTTGCATTTAAAGTTAA

>U86

ATGGAGCCAGCAAACCCCTCTGAAACAACATGGGATCCAATGATGAACGTATGCAAGAC
TACCGTCCCGATCCAATGATGGAAGAATCTATCCAACAGATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCAAATGAAGACTTG
AAAACTTTAGCTGAACACAAATGCAAACAAGGAAATGATAACGACGTCATTCAAAGCGCC
ATGAAATTTATCCGGACTATACCTGTGATGCGGACATCACACACACAACCGTTGAGTGAT
AATACGCATCAAGATCCTATATATTTCTCAAGAATCCAGAATATTTTTCAAAAACCTATCCAG
GATCCACGGATAGTAGCTCAGACCCACAGACAATGCACATCCTCTGCTTCCAATCTACCA
TCAAACGAAAGCGGTTTCGACCCAGGTGCGCTTTGCTTCGGAACCTACCGAATCAACTCCTA
CAGCCCATGTACACGTACATAACCAAATGCAAATCTACAGAACAATTTACCTCGTTA
CCGTATCAACCTTATCATGATCCATACAGAGATATAGAGAGTTCTACAGAGAATCTAGA
AATACCAATAGAGGATATGACTATAACTTTAGACATCATTCTGATAGGCCAGAGGAGGA
AATGGAAAATATAATTACTACAATCCAAACTCAAAGTATCAACAGCCATACAAAAGATGT
TTTACCAGAACTTATAACAGACGCGGAAGAGGCCACAGATCATACGATTGCTCAGACAGA
TCCGCGGATTTACCATATGAACATTACACGTATCCAAACTACGAACAACAGAATCCAGAC
CCCCGTATGAATAACTATAAAGATTTTACACAACCTAACAAACAAGTTCAATTTTGGAGCC
AATGACTATAGCATGGCATTTTCAACAGACTCCACACACGTACAATCTGACAATTACAAT
CATCCAACAAAAGCACAGACAATACCAGAACTACAAAAACTAAAAAATGAGGCTACT
AAGGATAACGAAACATCCAGAGAAAATCAAGTTTTAACATCTAATGACGCGATCAGTTTA
TCATACAGACCTAGTCCCACAAAACCTGAATATTTAAAAAATATATGACACTGACGTT
ATTCGTTACCGAAGGAAGCGTTAACAGGCAATGGCTCAAATCGTGACGTGGATATTCAG
AAGTACAAAAAGCACATATCCGTTGTAGATCAGTCCAGAAGAAAAAAGAACGCAGTTCT
CAAACAAACAAACATGATGAAAACCATGCCTCATCACGTTCTGATCTTAAAGAGAGAAAA
TCTAATGAGCATGAAGATAAAGCCGTTACTAAAGCTAGAGATTTTTCAAATTTAGATCCC
CTATTATCTCCTCTTCCGTTGACACCTGAACCCGCCATAGACTTTGCGGACCATACAGAT
AAATTCATTCGACACCAGAATTCAACCAAATAAAAACAAAATCTACACCGATCAAAAAC
TCTCTACAAGACACAGTTCCCATATCAAACATACTCCAAGAGCCCCGACCAAAGATAAT
TCTTACAAAAACATCATGACTCAAAGACAATTATCTTAAATGAAACATTTCTCCTGGA

CGTACAACATCTAAAAAATACTACAAACTCCAACGGACACCAAACTTTAAGGAAGTA
TCCGTTAAGAAATGTTTCAGGGAAAGCCACATCTACATCACCAAAATCAAAAACACACCAT
TATTCAGTAGCTCTGATGAAGAAGGTCAATACAAATCTCCGGTAAAAACTATTATCCCA
TCACCTTCTCCATATTGCAAGTTAAAAAATCCTTCAATTATGGACAAAAACAGTGCTAAA
AACCACACTGCAAGCGCCGACAAAAATTTAACAGACAATTCACCTATTCGGAGTAATTTA
AATCCTACTGCTTTTAAACAAATCCAACAATAACAAAAGCATCACGAATAGCACTTCTAAC
AGTGATGAGTGTACTGACAAGAAACCTAACTGCAATTCACCAAAAAATGAATCTAAAGAT
CCTAACCGAACTTGTGGCAAAAACTCTGATAAACATCTGAGTAAAAGTTGCACCATGGCT
TCTAAAAGGGCTCCCAGCAGAGCCTCAAGCAGAACCTCCAGCAGGGATTCTAGCAGGGCA
TCAAGCAGAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGAGCCTCCAGCAGAGCC
TCCAGCAGGGATTCTAGCAGAGCCTCCAGTAGAGCCTCCAGCAGGGCCTCAAGCAAAGCC
TCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGGGCC
TCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGAGCCTCCAGCAGGGAT
TCTAGCAGGGCCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGAGCC
TCAAGTAAAGCCTCAAGAAAAGCCTCCAGCAGAGCCTCCAGCAGGGCTTCCAGCAGGGCC
TCAAGCAAAGCCTCCGGTAAAGCCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGTAGGAAT
TCTAGCAGGGCCTCAAGCAGAGCCTCAAGCAGAGCCTCCAGCAGGGATTCTAGCAGAGCC
TCCAGCAGAGCCTCCAGTAGGGATTCTAGCAGGGCCTCAAGTAAAGCCTCAAGAAAAGCC
TCCAGCAGAGCCTCCAGCAGGGCTTCCAGCAGGGAGCTGAGGCAGATTTATTGCGATTCT
AATAAAAGACAAACGCCCTCCACATGACACCTCAATTAATACAAAATTTGAAATATCAGAA
ATAAAATTTTCGATGTGGAGAAGATCTAAATTTCTACAAAATACTGCAGCTCGACTCCAG
TGCTTTAATCACAATGATCAATTTTACAACCCCGATTTCAGACCTCATATAAGAACCAAC
AGAAAAAATCAGAATCCACCAATGACACCGATTTCAGAAAAGTTCTATGTCTAGATGCAAA
TCTCACTGTGCAAAATTCACCAGATTCCTGACCATTGTAAGAAGAAAAAACACAAATCT
GGATCATCAAGCATCTCTTCCAGTATTGAAGAAAATTGCAGATCTAATTCACATATAGTA
ACCGAAAAGAGAAGTTTACACCATTTTACTATCAGTCTTCCAGAACTAGATCATCATCA
TCATCATCAACATCATCATCAGCTTCTTTATCATGTTCCAAATCCACATTAAAAACATGC
CGCAAACTCAATATAAAGACAATAAACAAATAAAATCTAAATCAGATTCAAAACACAAA
ACTACAAACATGTCTTCAGATTACGAATCGAATAGACATGCAGATGTGTTTAGAAACTCC
CCGGAAGCAGGTGAAAAATTTCCCTACACAATTTCTAGTCCATTTAACACACATGAACAA
TCCAATCATTCGGAAAACGCCATTTGATGAAGAACAAAAAAGGCACCTAATATCACGACT
TCTCACCTACATCAAAAACAAAATGTCAAATTAACAACACAAAAAATGCAAAAAA
CGCCCCGCGATGATGATTTCAGATTCCAGCATAAAAAATTTCTGCAAGAAAAGAATATCC
GGAGCGCAGAAAACAGAGTCTGAAGTAAGTGAACCTGACGATTTGTGTTATAGAGATTAT
ATTAGACTAAAAGAAAAGAAAAGTATCAGAAAAATTTAAAATTCACCGCGGGAGGGTGGCT
ACAAAAGATTTTCAAAAACATTTTAGAAATACTATGAGAGCATTGTAATACAAACAGATC
CCAAAAAACCTTGCAGTGAGAAAAATCTAAAAGAAGCTGTATACAGCATCTGCTGTAAT
GGGCTCTCTAACAACGCAGCCATCATTATGTACTTTACAAGATCAAAAAAAGTAGCTCAA
ATCATAAAGATAATGCAAAAAGAATTAATGATTTCGACCAACATCACTGTTTCAGAGGCA
TTTAAAATGAATCATGCACCGCCTAAGTATTATGACAAAGATGAAATTAACGATTTATC
CAACTACAAAACAAGGACCGCAAGAATTATGGGATAAATTTGAAAACAACACAACACAT
GATCTATTTACGAGACATTCGGACGTAAAGACTATGATAATTTACGCCGCTACGCCCATC
GACTTTGTAGGAGCTGTTAAAACATGCAATAAGTATGCCAAAGATAACCCTAAAGAAATT
GTCTTCGGGTGTGTTCAATCATCGACGGTGACAATCCTATCTCCATCTATAATCCTATC
AGCAAAGATTTTAAATCGAAATTTAGTACACTTTCAAAATGTTAA

>U90

ATGGAGCCAGCAAACCCCTCTGGAAACAACATGGGATCCAATGATGAACGTATGCAAGAC
TACCGTCCCGATCCAATGATGGAAGAATCTATCCAACAGATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCAAATGAAGACTTG
AAAACTTTAGCTGCTCAGTCCGCTGGGAATTGCATCCAAGTATTGGTGCATCAGTGAAG
GCTGCCATGAAACAGGAACAGTCTGACATGGAAGACAAATTAATTAAGTGCCTGGCTTA
TTAACGCAGCAACAATCCATGTTTCATAGGACTAGGATTAGAACAATTATCACAGCTGATA
AACATTAATTTATTAAGTAGTGTCTCAACTAAATATGTTGAATCTTATTCAAAAATGTTG
CATGGAAAAGAGTTGGATTTTTTTAATTTGGTGCGAACCTAGATTCAATGTTTTTGCATGT
GATAAATTCGATGGGTTAGTGA AAAAGGTGGCATCAGAATCCAGAACTTGCTCCTGGAT
TTGAGAGCTAATATGAACAATGACATTATTAAGGCTGTCAAAGACATTTTTTCTAAAGCT

ACTGTGACATTAGATTGTCAGAACTTAACCAGGGCGCCACCATGTTGATGATGATGGCA
CACAACAAAGAGATGTCCAATCCTGACATCTCCAGTAAGGACTTTTTGTGAAAAAATAAAT
GTACTTAAACAGACATTACTGGAAGGTAAAAATGAGATAGTAGAAACAAATGCTAAAAAT
ATGCAAAATTTTACAGACATTTGCTATAAAGCAGATGAATCAAATCTTCATGGATGGTTGT
GATAAAGCATTCTTAAACTTAATGTGAATTTGAAAAATTTAATTACTGCCGCCAAAAAT
CTTGCCAATGCTATTCTCCAGAGTATTATTATATGCAGCAATGAATTTTCATGGCAACAC
CTCAAACCTTGAGACAGGGATTTAAAGCTACCATACTGACTATGATTACCCAAGCATGT
GAACGCTTAGAATCAAGTTATGATGATACTGGTTTTAATTAACCTTTAGCTCCCCTTCAA
ATAATGGAAGGCTACATTAATATGGATAAAAAATAGACAGTCTTCCATTTGTGATGGAAAC
ACAGACCCATCAGACTCAATGATATTAGACTTAGCTGATTTTGATGACCATGGAAGATAT
TCAGAGGAATCTTCTATCGAATCTATCCATGAAGATGATGACAATAAAATGTATCCATGC
ACTCCTTCTCCAGAAGTGCCAGGGAAATCCAAATATGTAGGAACATTCCTGAAAACAGC
AGACAGTCAGGGGACGAACAAACAAACCCCTAACTGTGTAGGTACAGCCTCAGTGACAGAT
CTGGGCGGCCCTGATAACTTGAATTCATCAGCGGCCTCCAGAGTTGTAAAAACATGTTG
CTAGAAAGACTACTTGACACTCAATGTGATTGAGTAGTTGAAGGGACAGAACAAGATGGT
TCCAATGGCAACCTTTAATATCAGAAATGTTGATGTTTGGTTATGAAACTGATCACTCG
GCACCCTATGAGAGTGAAGCGATAACAATGATGAAATTGATTACATTGCAAACCTGAT
TCAGCTGCGAGGACTAATAACATTCATATGAATAGCACTGATGAGAATACACCATTTAGC
AATTCCTATATGTTACCACCAGAGGTCACCCCTTCTAAGAAAAATGTCAAACCTAAGTCT
ATGACCCCGGGAAGCAAACCTAAAAAGAGGGTAGCTAAAAGGAAACATGTTTCAAGTAAA
TCACCCAAACATAAAAAAGATTAAGACAGATAAGTTACCTAAGGCTGCAGACGTAATTGTA
ATAAGTTCAGAGAGTGAAGATGAAGAGGATGGTGACAATATTATTGGTAACTCAATCTTA
ATAAAAGCTATTAATCAGAATCTGATTCTGAATCGAGCAGCGAGTCAGATGACTGTACA
TCTGAACATGAACAGTTACATTTAAGTGATTATGATAAAGTTACCAATAATGGTGATTGT
CCAAGTAAAGGTTTTCCAACCTCAGTGTTTACCATCCCAATAAGGTCAATGCAGGGTATC
AGTGATCAAAATCCAAATTTATTCCAGAAAAAATGGATTTGGTATATGAAAAAACT
CACCAGTAGACAATTTGCTTATTCACAGTTCTGAAAATGTGGATGCAAAAGATGACAAT
GATGGGACAGAAGCCAAACACTGCTTTATGAATCACTTTGTCCCATCAAACCGGATGAT
GAAGACTATGATAAGGACAATGTCTCATATATTTATAATAAAATACAGAATAGCAAAATA
GACTCGGGAGTCATAACACCAACTAAGAGACTTATCATAGATATGGTTATGGATAATTTT
ATGGATCTAAATGATATCATTTAAACAGGGCATAATCAAACATTGCCAGGATCTTTGCAAT
AAATATAATGTTGTCACCCCTACTACATGTGAGAAAGATTTAAATGTGGCAAATTTCCAA
ACTTTTGCAACAACCGCAACTCAAGTGTTTGACCCACCGGTAACAGGAAATAATTCCTCC
ATACTAAATATTATCAATGACACAACAAGTCAAAATGATGAGAACAGATGCACTGAAGGT
ACAAGTAATTTCTAATGAAAAATGTACCAATATAAGTGATTGCAACAGTGATGGTACAGAG
GCCTTTAAATTTAGATGGCTACCCCTTCTGATTATGATCCGTTTCATAGAAAATGCTCAAATC
TATTAG

>U91

ATGGGGAAAAATGTCATCTACTGGAAATGGAAAACTAATCTAAAGATACTGGCATGTATG
CTATTGATATTTCTAATAGCAACAATATCTTTGTTAATACTGGAAATTATATCGGGCCAA
AGATATTCCAATGATGATTCCGAAGGTGTGACTGCGGCGCTGAAGCATGTAAGCACACCT
ACTACCAACTGCCTGAAACCCTACTCCTGATTTGTAACGAGCCAAGCCACGGGAAAC
AAAGAATCAATGAAAAAAATGAAGGGGAACCTCCAGTTTGGATTTCAGGCTTTAACTACA
ACATTATCAATTTATCCATTGGTTTGCATAATAATGGCATGCATTATATGTTCCCGGACC
ACGGAAGAAGAGAAGAGCGGAATGCAATCTTCTGCTTCTTCAGTAGAGACTTTACAGTCT
TTAAATGAGGCTATTTTACCAGAAAGGTGAAATGAATGTGTAA

>U94

ATGTTTTCCATAATAAATCCGAGTGATGATTTCTGGACTAAGGACAAATATATCATGTTG
ACTATCAAAGGCCCATGGAGTGGGAGGCAGAAATCCCTGGAATATCTACGGATTTTTTTC
TGTAATTTCTTAACGTGTCCGTGCCACATTTTCAGGGATATGCACTCACCGGGAGCGCCC
GATATTAATGGATAACTGCATGTACCAAAATGATCGATGTCATACTCAATTAATGGAAT
AATAAACTGCCGTCCCACCCCTGCAAAGTGGTACGCTCAAGCGGAGAATAAAGCTGGC
AGACCTCCCATAATATTATTGATAGCTTTAGATGGAATTTCCAGCGCAACGATAGGAAAA
CACACAACGGAATCAGGGGCGTATTAATTAAGATTTCTTCGACGGGAACGCCCTAAA
ATAGATGATTTGGTGCACGTATGCCAAAACAAAGAAGAATGGTGGCGGAACCCAGGTCTTC
AGTCTAAGTTACATCCCTTTGCCCTTCTTCAAATTTATTAGACCACAGTTCCAATGGGCG

TGGACAAATATTAACGAACTGGGAGACGTATGCGATGAAATACATCGAAAACACATCATA
TCCCATTTTAAATAAAAAACCTAATGTCAAACCTTATGCTGTTTTCCAAAGGATGGGATTAAC
GGGATATCTTTAAAAATCTAAATTTCTGGGAACCATCGAATGGCTGTCTGATCTTGGAATA
GTCACGGAAGACGCGTGGATACGAAGAGACATTAGATCATACATGCAATTGTTGACGCTA
ACACACGGGGACGTGCTAATCCATAGGGCTCTATCTATAGCTAAAAAAGAATAAGAGCA
ACTAGGAAAAGCTATCGATTTTATAGCGCACATAGACACTGACTTTCAGATTTATGAAAAT
CCGGTTTACCAGCTGTTCTGTCTGCAGTCTTTTGACCCTATATTAGCAGGAACCATATTA
TATCAGTGGCTAAGCCACAGAGGAGGAAAAAAAACACCGTTAGTTTTATCGGCCACCC
GGATGTGAAAAATCGATGTTAACGGGAGCCATTCTTGAAAATATACCGTTACATGGAATA
TTACATGGATCTTTGAATACAAAAATTTAAGAGCTTACGGACAGGTTTTAGTCTTGTGG
TGGAAAGACATAAATATCAACTTTGAAAATTTAATATTATAAAATCCCTCCTTGAGGT
CAAAAAATAATATCCCAATTAATGAAAACGACCATGTACAGATAGGACCGTGTCCCATC
ATAGCCACATCTTGCGTTGATATACGCTCGATGGTACATTCCAATATCCACAAAATAAAT
CTATCACAGAGGGTGTATAATTTTACATTTGATAAAGTCATCCCTCGCAATTTTCTGTGA
ATTCAGAAAAGATGACATAAATCAATTTCTGTTTTGGGCCAGGAACCGTTCTATAAATTGT
TTTATTGACTACACGGTTCGAAAATTTTATAA

>U95

ATGTCCTGCAAGTCCGGACGATTTCTTTTTGGCAACAAATTCTCAATATGGATCCGACGGAA
CTTTTGTCCGACGATGCAATCTCGTCCACGTGTGACGAAAACACTGCAGCTGGAACCTGAT
ATTATGCAGCCTTTGTGCGTTGATATGTCTGTCCAACCGACGACATCGGGGTGTTATACC
GGAACGATGATGTCAGACTCGTATTCCCTCCAATGACACACCCGGATCAAAACTGCCAATCA
CATCCGACTTTGACGGACCTTCGCCGGAAGACATGAACAATCAAATCAATGTTCAACCC
AATCAGATGACCTTTCAACCCATTTCCCCACCGATACAGGATCGAAATTATGGCTATAGC
AATAATATGATTAATCCCATCAAACCTGCATCAATTATCAAATTACATGGTCATTTCGATA
GGAGAAATGACATCTTCAGACCTTTCCTCTCATGTTAATGCTCAACACTTTGTCCAACAG
TCTAATCCTAAATCTTTCTAAATATGCATCCGTGCATGACCGCTAACTCACAAGGACCC
GGGAAATGTCAGTCTGTCCATAATCAATCTTCTGGAAGCGGGTCTAATTTCTTATGACATT
ACTACGGCACCCCGGATGAATGGATATTGACTACATCCGGCGGAGAATCCTGGATTCTG
AAACGGAACACCCCTAATCCTCCAAATAGCAGAATAAATTTAATTTTCAACAACGCACAA
CCGATTTTCCATATCCAACCTCAAGTATCTGGGTTTAGCGACTGTTCTATCAGAGGGGT
GGTCTCAATTCACGTGAGTATGTAAATTTGGCAGGTTACACTCCCGTGAGCAACATGCAG
AATATGAAAACTCATGGGCGACACAACCTAGGAGCATCCATTGACTGCATGACTGCGTCA
CCAACCTGGTATAGATGCGAATGTCAACTCTGCTTTCCAGCCGGTGGGGGTGGTATCAAAT
AATCGGGGTGCTAATGTTCCAATGGGAGGCATGGGTAATTATATAGAGAATAACAGTCCA
TGGAACCAATACTATAAATCGGAAATGGCTAATAATTCATAAATGTCAACGAGTCAAT
GGTTATGGCCATGTGGGGTGTGTTAATGACGCCATTTCCGATAAAACAAGGTGGCACCGCT
AACGTCCCATCTTCTCTATTGAATCCGGATCACCAAGATTGGATGAGGATGACTGGTACG
AGCACACATATGTTAAACAATATAAATGCTATAACAGAAATGAAAAACTATGATTTCCCG
GAAAATGATGGCAATGTCCACGGAGCCACCGGTGTATATTTATCATCAACCCTGTCCAAT
GGTCAGCAGCATAGCGCATTTGAAGGAAATGGAACATTTCCCGTGGTAAATACCGGCAAT
CCAAGAACGACGTTGAACCTGTGCTCCTCAGTTGTTGTGGCTTGGAGGAAACCCGACAATT
GCACATCCAATGAATGGGAATAATGAAACCCAGATAAAAAATGTATGCAAGCCTACCCCC
AGCACAGGTGTAATAAAAAAATAAACCTTTGACTGTGTTGACCGAGGGGAAGGAATAGGA
AATTTTCCCAAACCTGGCGTCTTTTGTAGCTATAAGAGAAAAAATGTCAAACACGGTCAAT
CCTTCTTGTGGAACCTTTCTTAGCACAGTTGGAGGGATTACGCCAGCAGAATTCATTTGAA
AAAAATATATCCCTTACTGATGTACAGTTTTCTGAAGAAGATGATGTGCTTTCCAGTGCA
TCCTCGGTGAGCTGTAATAGCAACTGCGACATGAAAATCGGAGTTTTCGCAGCAAGGTACC
ACTGTGTCTAATTTACAACAAGGGTTTAGGCAACAAATCAACATGAACTCATCCGTGGTT
AAAATGGACGGGAGCTATAAAACACAAGAGATGTCAAATGACTGCGCGACAGACATCACG
GATAATGCAGGCATCATCAAGAAAATAAACGTATACACTCTCAAATTTGAGAATCCAGAG
GATGGCCGGGTACGGGAGGGGGGTGTTTCAGACGACAAAAATCAACCTCCGCGGAAAAGT
GCGGTATTCATAACATGAAGTCCGAAGGCGTGACTTGCAGTATGTGCATGACGGCTGCG
GATTCACACGGCAGGATGCGTCTGGAGGTTCAAGCAGCGGAACCAAGAATGAATATTAC
GATGATGAATCGGAATTAACCGACTGTTCGGATACAGACAGCGACAACGAAGTCCAATGC
TGTCAAAAAGTAACAAAGGTCGGTCCCAAGACGTACTCCTCAGAAAATTTAATCCTGAT
TATCGCCAAGCCAAGAGATTGCTTGGCGACATTCCATACAGAGGATGGATTCCAGACCCA
GTCAATATGGAAGAAAACGACGGTCCGTTTCAATAGTGACTAGACCACCCACTGTG

TGTATGAACGGTAGACGCCGGAGGACTTTCTTCCGGCAATGTGTTACCGCTTTCCGGTCCT
CTGAGTAAGTTAATTTATTTCAAAGAATTACTGCAGAGCTACGTGTCCAAAAATAACAAT
AGTTATTTAAGTATCAGTTGGCCTCCCAAACACGGGGTGTATGTCATGTCAGAGAAGAAA
TTAGGGTATGAACATATACCTACCTTGAAGGACAAGTTTCCACTGCCATGTGGGTGGATG
ATGGTCTTGGAGTTGTCCGGTGCAGAACCGCCCGGGCTCTGAATAAAAAACATGGTGATT
CTCCTGTGTGAAAATAGATGGGTGCTTCTGCATAATTACTCAAACCTCTACACACGAGCTG
TTTTTAGCCGCCCTCGGATTTGAAACAGTTTATGGAGGAGGGTCTGTCTAGATGTGATCCT
ATATATGAAGAACCAACAGTGCCGTACGGTGTGGCCATGGAAAACCTCTCTTCGAGATTTT
TTGAGAAACAGTAGAACGTTTCAGGATCTTATGGATCAAAGGGACAACATGCACGGATGT
AACTGGACCTTTAACGGAATGCCTGGTAGACTAGGTGACAGAGTCATACATCTCTGCAAT
CCTGAATCGGTTGATTCATACCAGGAGACGAGGCCGTCTGTATGTGAAGGAAGACCCTTG
TATTTCTTTGCATATGTTACCACCTTCAAATCAAACCCCGCTACTAAGGCTACCGTGTTA
ATAGCCGCGGATAAAGATTTAAAAATATATGGCTATCATAAAGGTGCCCCAGGATACGG
TACTTGTGTAAGAACGTAAAAACCTTCTTTAAAGCCGGGGCTAGAAAGTTCTATTTAGAT
TTTCAGATTACACCTAAAAGCTTGCTTGCCGTCAACGAAGAGGAGTATCTGAATACTCTA
CAAAACGCTCCTTGTCTGCTCTTAAAACCATCTGTGTTTAGAAAGACTTACTCTCAGGAA
GGTAAATAA

>U100

ATGGCAACCGCAAGGTTGGGTGTTATGAAACCACCGAGATCTTGTGCGCTGATATTTTTG
TGTGCGTTTTTCGATGGCGACGGCTCCGACGAACGCTACGGCCACCGACGTGCGGGAACC
GTGAAATCCACACCGCCGAGAGGACAAACATAACTATACCGCAAATACTATGACAAA
GACATTTATTTTAACATTTACGAAGGTAGAAATTTCTACGCCGAGGAGACGGACGCTATCG
GAAATCATCTCTAAATTTTCTACCTCCGAAATGTTAAGTTTAAAAAGAGTTAAAGCTTTC
GTCCCGTAGATGAAAACCCGACTACTACCTTAGAAGATATAGCAGACATCCTGAACTAC
GCCGTTTTCGATGACAATAGCTGCGGTTGTACCATAGAGACACAAGCACGCATCATGTTT
GGAGATATCATAATCTGCGTTCTCTCTCGGCCGACAACAAAGGAGTGAGAAATTTTAAA
GATAGAATAATGCCTAAGGGTCTATCGCAAATTTCTCTTAGCAGCCTCGGCCTTACCTG
TCTCTGCTTTATGGCGCTTTTGGGAGCAACTATAACAGTTTGGCCTATATGAGGGCGCTG
AAACCACCTGACCGGATGACGGCGATAGCTTTCTGTCCAATGACTACTAAATTAGAGCTA
CGCCAAAACACAAAGTAAAAGAACTCTATGTGAACTGATCGTCAGTATCGAGATTTTG
AAAATACGAAACAATGGCGGACAGACTATGAAGACGTTAACCTCGTTTTCGATCGTTTCGA
AAAGACAACGACGGGACGAGATTGGGAAACCTGCACGCGCTTTGCGCCCGTTAATATAGAG
GACATATTGAGGTATAAACGTGTGCTAATGACACGTGTTGTGTCACCGAGACGTGCAG
CATGGTCGTGACGTTAGAATCTTCTAACTCGTGGACTCAGACTCAATATTTTGAACCG
TGGCAAGATATCGTTGACGTATACGTGCCTATAAACGATACTCATTGTCCAAACGATAGC
TACGTCGTGTTTGAACACTGCAAGGTTTTCGAATGGTGTTCACGTTTAAATAAAAAACGAG
ACGAAGAATTATTTATCTTCCGTACTAGGGTTTAGAAATGCACTATTTGAAACTGAGGAA
CTGATGGAAACCATCGCAATGCGACTTGCCTCACAAATATTGTCAATGGTTCGACAGCAA
GGAACAACATACGCGACATAGATCCTGCCATCGTAAGCGCATTGTGGCACTCTTTGCCA
GAAAACCTGACAACGACTAATATAAAAATACGATATAGCCAGTCCGACACACATGGCTCCC
GCATTATGCACCATATTTGTACAGACCGGAACCTTCAAACAACGTTTTTAGAAATGCGGGA
CTACTGATGGTTAACAACATCTTACAGTCCAAGGGAGATATACTACGCAAATATGTTT
GAAAAAAAAGAATATGTATATAAACACCTCGGACAAGCTTTATGTGAGGACGGTGAATTT
CTGTTTCAAATGAAGGCCAGAAGTTTTGTGCGACCTCTCACGGACAATAGAACAATTGTA
TATACAATGCAAGACCAAGTTCAAACAACCTTGTCCGTAACGTGGATGGATTTCAATCTA
GTCATTAGTGACTACGGAAGAGATGTGATAAACAATTTAACCAAAAGTGCTATGTTGGCA
CGCAAAAACGGGCAAGATATCTACAGATGGAAAACGGGCAAGATATCTACAGATGGAA
ACGTTTATATCGGATTTATTCGACATGAATGTTATCAAGATAATTATTATGTGTTAGAC
AAAAAACACAAATGTTTTATCCGACTACTCATAGTAACGAATTACTTTTTTATCCCTCG
GAAGCTACGTTGCCCTCACCTTGGCAGGAACCTCCGTTTTCTCACCTTGGCCGGAACCT
ACGTTTCCCTCACCTTGGTATTGGCTCTTATTAATAACAAAAATTA

7.3 Nucleotide Sequence Alignment of Suspected Chromosome 17p13.3 CI-HHV-6A Integration Site Junctions

Clustal W2 [Larkin et al., 2007] alignment with manual adjustments of the nucleotide sequences of the 12 amplified CI-HHV-6A integration site junctions, as described in section 4.2 and summarised in figure 4.2. Grey shading indicates readily identifiable subtelomeric sequence. Green shading indicates readily identifiable viral DR_R sequence.

CI-HHV-6A 10593	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 15929	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 17172	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 23090	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 23117	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 25533	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 27355	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 38862	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 40672	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 42815	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 43670	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT
CI-HHV-6A 5055	GTTTAAGGTAGAATTTTCATAATTTGGTTTAGAGATTATTCATTCCTAGCATATACATAT

CI-HHV-6A 10593	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 15929	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 17172	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 23090	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 23117	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 25533	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 27355	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 38862	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 40672	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 42815	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 43670	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG
CI-HHV-6A 5055	AAAATGGAATGTTTGGCTAGGCACCCGGGCTCATACCTGTAACCCAAGCAGGTTGAGAGG

CI-HHV-6A 10593	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 15929	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 17172	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 23090	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 23117	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 25533	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 27355	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 38862	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 40672	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 42815	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 43670	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
CI-HHV-6A 5055	CTGAGGAAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT

CI-HHV-6A 10593 -----
CI-HHV-6A 15929 AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA-GGGTTAGGGTTAGGGTTAGGGTTAG
CI-HHV-6A 17172 AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA-GGGTTAGGG-TTAGGG-TTAGGGTTAG
CI-HHV-6A 23090 AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAAGGGTTAGGGTTAGGGTTAGGGTTAG
CI-HHV-6A 23117 -----
CI-HHV-6A 25533 AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA-GGGTTAGGGTTAGGGTTAGGGTTAG
CI-HHV-6A 27355 AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA-GGGTTAGGGTTAGGGTTAGGGTTAG
CI-HHV-6A 38862 AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA-GGGTTAGGG-TTAGGG-TTAGGGTTAG
CI-HHV-6A 40672 -----
CI-HHV-6A 42815 -----
CI-HHV-6A 43670 AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA-GGGTTAGGGTTAGGGTTAGGGTTAG
CI-HHV-6A 5055 -----

CI-HHV-6A 10593 -----
CI-HHV-6A 15929 GG-TTAGGG-TTAGGG-TTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-
CI-HHV-6A 17172 GGGTTAGGGTTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT
CI-HHV-6A 23090 GG-TTAGGG-TTAGGG-TTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-
CI-HHV-6A 23117 -----
CI-HHV-6A 25533 GG-TTAGGG-TTAGGGTTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-
CI-HHV-6A 27355 GG-TTAGGG-TTAGGG-TTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-
CI-HHV-6A 38862 GGGTTAGGG-----TTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-
CI-HHV-6A 40672 -----
CI-HHV-6A 42815 -----
CI-HHV-6A 43670 GG-TTAGGG-TTAGGG-TTAGGG-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG-
CI-HHV-6A 5055 -----

CI-HHV-6A 10593 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 15929 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 17172 TAGGGTAGGGTTAGGG-----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 23090 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 23117 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 25533 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 27355 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 38862 ----TAGGGTTAGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 40672 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 42815 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 43670 -----TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTT
CI-HHV-6A 5055 -----TTGGGGTTGGGGTTGGGGTTGGGGTT

CI-HHV-6A 10593 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 15929 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 17172 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 23090 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 23117 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 25533 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 27355 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 38862 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 40672 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 42815 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 43670 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT
CI-HHV-6A 5055 GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGCCGTTGGTT

CI-HHV-6A 10593 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 15929 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 17172 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 23090 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 23117 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 25533 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 27355 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 38862 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 40672 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 42815 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 43670 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT
CI-HHV-6A 5055 TCCTTTGGCACCGTGCCAGGCGACCGTCCGTTCCAGACGGTGGCACGTCGCGCGCGATGT

CI-HHV-6A 10593 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 15929 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 17172 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 23090 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 23117 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 25533 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 27355 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 38862 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 40672 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 42815 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 43670 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC
CI-HHV-6A 5055 CGTTTTGAATGTGCTTCAAGTCTCACCCCGGTATGCCCCATCTACGTACCCACCCGCGAC

CI-HHV-6A 10593 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 15929 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 17172 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 23090 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 23117 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 25533 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 27355 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 38862 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 40672 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 42815 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 43670 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT
CI-HHV-6A 5055 TACCACATGGGTGGGTTGTGCGTGTCTGTGTGTCTGCGTGTGTACGCGTCCGTGGTAGAT

CI-HHV-6A 10593 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 15929 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 17172 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 23090 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 23117 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 25533 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 27355 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 38862 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 40672 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 42815 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 43670 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA
CI-HHV-6A 5055 ACGCGGTGACAACGGATTACGGAGGTGACTGTTGCGTTGGGTGGGACGGCGAGGGGGTGA

CI-HHV-6A 10593 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 15929 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 17172 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 23090 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 23117 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 25533 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 27355 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 38862 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 40672 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 42815 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 43670 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC
CI-HHV-6A 5055 AGAGGTGTTTCCCGTATAGGTGTTTTTGGAAACATGGTTTTTTTTCTTATACTTGCCATGC

CI-HHV-6A 10593 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 15929 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 17172 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 23090 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 23117 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 25533 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 27355 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 38862 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 40672 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 42815 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 43670 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT
CI-HHV-6A 5055 TAGCGGTGCGCCTCTGTCTTCGGTCCGCGTGACCGTGTGCGTGTGCACTTCCGCATACTTT

CI-HHV-6A 10593 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 15929 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 17172 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 23090 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 23117 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 25533 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 27355 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 38862 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 40672 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 42815 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 43670 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC
CI-HHV-6A 5055 TGTTTTTGC GCCTCCATACACCCCGTTGATTGGATGATCGCCACATGTGTATTTCGATGTC

CI-HHV-6A 10593 ATAGTCTG
CI-HHV-6A 15929 ATAGTCTG
CI-HHV-6A 17172 ATAGTCTG
CI-HHV-6A 23090 ATAGTCTG
CI-HHV-6A 23117 ATAGTCTG
CI-HHV-6A 25533 ATAGTCTG
CI-HHV-6A 27355 ATAGTCTG
CI-HHV-6A 38862 ATAGTCTG
CI-HHV-6A 40672 ATAGTCTG
CI-HHV-6A 42815 ATAGTCTG
CI-HHV-6A 43670 ATAGTCTG
CI-HHV-6A 5055 ATAGTCTG

7.4 Alignments for Determination of Practical Level of Read Coverage for Accurate Sequence Resolution

The parameters used for both variant call methods required any variants to be present in a minimum of 5 reads for variant calling to occur, a value chosen to minimise false positive results. This defined a theoretical read coverage limit required for accurate sequence generation from the mapped assemblies. The work sequencing HHV-6A strain AJ, shown in section 3.2 and table 3.1, indicated that a minimum read coverage of 16 was sufficient to accurately call variants across the CDS one of the most diverse genes in the HHV-6A genome, U100. However, the fact that a number of genes from the CI-HHV-6A samples had been subjected to Sanger capillary sequencing for alternative lines of work described in Tweedy et al., 2015a, using the primer sequences outlined in section 2.3.3, provided further means to ascertain a practical level of read coverage required for accurate variant calling in these samples. For this, read coverage was calculated across each of these genes, while the nucleotide identity of the consensus sequence generated by NGS was compared to that obtained via Sanger capillary sequencing, table 7.1.

CDS	CI-HHV-6A 2284			CI-HHV-6A 5055			CI-HHV-6A 5814		
	Mean Read Coverage	Minimum Read Coverage	NGS % Identity with Sanger	Mean Read Coverage	Minimum Read Coverage	NGS % Identity with Sanger	Mean Read Coverage	Minimum Read Coverage	NGS % Identity with Sanger
U38	119.46	44	100	7710.53	2614	100	1566.01	628	100
U46	202.74	182	100	7.04	4	99.61	4124.52	3774	100
U47	184.77	51	100	7.61	2	98.16	3707.86	1396	100
U51	18.06	10	100	3.57	0	99.67	1.04	0	99.56
U65	100.41	72	100	5.64	0	97.72	40.03	26	100
U83	407.04	336	100	1158.37	1031	100	8417.23	7788	100

Table 7.1 Determination of practical level of read coverage for accurate genome sequence resolution. Pairwise comparisons of the nucleotide sequence of genes subjected to both NGS and Sanger sequencing. Differences in sequence identity are indicated in red. Full alignments of these genes are below.

This indicated that a minimum read coverage of 10, seen in the U51 gene from CI-HHV-6A sample 2284, still permitted accurate variant calling. While U51 represents a relatively conserved gene amongst HHV-6A, there are clear differences in the gene between strain U1102 and the Sanger amplification from CI-HHV-6A 2284 (99.67% nucleotide identity between the sequences, variation in this gene is also discussed further in section 5.2). All these sites of variation were accurately called by the NGS protocol used here, full alignment of these sequences alongside the U1102 reference used for the mapping assembly is shown below.

To further define this level, the read coverage required for the correct identification of SNPs in the U47 gene, the variable glycoprotein O, was further investigated for CI-HHV-6A isolate 5055, alignment shown in appendix 7.4.7. This showed accurate resolution at a read depth of greater than 7 reads utilising the read assembly and variant analyses methods here. Thus for the purposes of the work with these CI-HHV-6A genomes described below, a minimum read coverage of 8 or greater across a genes CDS was applied as a cut-off for inclusion in further analysis. Table 4.3 highlights 14 genes where one or more of CI-HHV-6A genomes possessed read coverage below this defined limit. Therefore while reads covered these CDS features, sequencing errors had the potential to confound these sequence, as such these genes were excluded from downstream analyses.

7.4.1 U38 partial CDS fragment

CI-HHV-6A Isolate 2284

```
2284_U38_Sanger      AGCATTATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA 60
2284_U38_NGS         AGCATTATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA 60
*****

2284_U38_Sanger      GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGATGAGACTCATCGGTTT 120
2284_U38_NGS         GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGATGAGACTCATCGGTTT 120
*****

2284_U38_Sanger      GTGAAGCCTTGATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG 180
2284_U38_NGS         GTGAAGCCTTGATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG 180
*****

2284_U38_Sanger      AGACGAGAAGTGAAGGCGGAGATGCAGAAGTTCGGATCCGATGATGAAACTTCTTCTG 240
2284_U38_NGS         AGACGAGAAGTGAAGGCGGAGATGCAGAAGTTCGGATCCGATGATGAAACTTCTTCTG 240
*****

2284_U38_Sanger      GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACAGGAGCG 300
2284_U38_NGS         GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACAGGAGCG 300
*****

2284_U38_Sanger      GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGCTTGAAGGGAG 360
2284_U38_NGS         GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGCTTGAAGGGAG 360
*****

2284_U38_Sanger      ATGCTTTGTTCCACGGTGGATTATGTTAATCCAAGATGCAGTCCGAGCAATTTTTTTC 420
2284_U38_NGS         ATGCTTTGTTCCACGGTGGATTATGTTAATCCAAGATGCAGTCCGAGCAATTTTTTTC 420
*****

2284_U38_Sanger      GAGGAATTTGGTTTAAACGTCATCAGATTTTACTGGTGAT 459
2284_U38_NGS         GAGGAATTTGGTTTAAACGTCATCAGATTTTACTGGTGAT 459
*****
```

CI-HHV-6A Isolate 5055

```
5055_U38_Sanger      AGCATTATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA 60
5055_U38_NGS         AGCATTATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA 60
*****

5055_U38_Sanger      GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGATGAGACTCATCGGTTT 120
5055_U38_NGS         GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGATGAGACTCATCGGTTT 120
*****

5055_U38_Sanger      GTGAAGCCTTGATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG 180
5055_U38_NGS         GTGAAGCCTTGATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG 180
*****

5055_U38_Sanger      AGACGAGAAGTGAAGGCGGAGATGCAGAAGTTCGGATCCGATGATGAAACTTCTTCTG 240
5055_U38_NGS         AGACGAGAAGTGAAGGCGGAGATGCAGAAGTTCGGATCCGATGATGAAACTTCTTCTG 240
*****

5055_U38_Sanger      GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACAGGAGCG 300
5055_U38_NGS         GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACAGGAGCG 300
*****

5055_U38_Sanger      GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGCTTGAAGAGAG 360
5055_U38_NGS         GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGCTTGAAGAGAG 360
*****

5055_U38_Sanger      ATGCTTTGTTCCACGGTGGATTATGTTAATCCAAGATGCAGTCCGAGCAATTTTTTTC 420
5055_U38_NGS         ATGCTTTGTTCCACGGTGGATTATGTTAATCCAAGATGCAGTCCGAGCAATTTTTTTC 420
*****

5055_U38_Sanger      GAGGAATTTGGTTTAAACGTCATCAGATTTTACTGGTGAT 459
5055_U38_NGS         GAGGAATTTGGTTTAAACGTCATCAGATTTTACTGGTGAT 459
*****
```

CI-HHV-6A Isolate 5814

5814_U38_Sanger AGCATTATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA 60
5814_U38_NGS AGCATTATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATA 60

5814_U38_Sanger GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTGGGGGATGAGACTCATCGGTTT 120
5814_U38_NGS GCTGGATTGTCAGAGAGTGACATCTTAACCGTGAAGTGGGGGATGAGACTCATCGGTTT 120

5814_U38_Sanger GTGAAGCCTTGATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG 180
5814_U38_NGS GTGAAGCCTTGATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAG 180

5814_U38_Sanger AGACGAGAAGTGAAGCGGAGATGCAGAACTGTTCCGATCCGATGATGAAACTTCTTCTG 240
5814_U38_NGS AGACGAGAAGTGAAGCGGAGATGCAGAACTGTTCCGATCCGATGATGAAACTTCTTCTG 240

5814_U38_Sanger GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGGTGACGGTGTACAGGAGCG 300
5814_U38_NGS GATAAAAAGCAGCTCGCTCTGAAAACAACATGTAACTCGGGTGACGGTGTACAGGAGCG 300

5814_U38_Sanger GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGCTTTGGAAGGGAG 360
5814_U38_NGS GCGCACGGGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGCTTTGGAAGGGAG 360

5814_U38_Sanger ATGCTTTGTTCCACGGTGGATTATGTTAATCCAAGATGCAGTCCGACAATTTTTTTCG 420
5814_U38_NGS ATGCTTTGTTCCACGGTGGATTATGTTAATCCAAGATGCAGTCCGACAATTTTTTTCG 420

5814_U38_Sanger GAGGAATTTGGTTTAAACGTCAATCAGATTTTACTGGTGAT 459
5814_U38_NGS GAGGAATTTGGTTTAAACGTCAATCAGATTTTACTGGTGAT 459

7.4.2 U46

CI-HHV-6A Isolate 2284

2284_U46_Sanger ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTACATT 60
2284_U46_NGS ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTACATT 60

2284_U46_Sanger CTTGTTTTTGGCGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTTCGCATACC 120
2284_U46_NGS CTTGTTTTTGGCGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTTCGCATACC 120

2284_U46_Sanger TATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATG 180
2284_U46_NGS TATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATG 180

2284_U46_Sanger TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAGACTTTTGCTTCAGAGACG 240
2284_U46_NGS TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAGACTTTTGCTTCAGAGACG 240

2284_U46_Sanger GTGAAGGGTTATTAG 255
2284_U46_NGS GTGAAGGGTTATTAG 255

CI-HHV-6A Isolate 5055

5055_U46_Sanger ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTACATT 60
5055_U46_NGS ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTCCATT 60

5055_U46_Sanger CTTGTTTTTGGCGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTTCGCATACC 120
5055_U46_NGS CTTGTTTTTGGCGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTTCGCATACC 120

5055_U46_Sanger TATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAAATCTTTTTTTATTTG 180
5055_U46_NGS TATGAGATCGTCTTGAATTCGTTTCTTCGATCTGGCTTTTGATAAAATCTTTTTTTATTTG 180

5055_U46_Sanger TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAGACTTTTGCTTCAGAGACG 240
5055_U46_NGS TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAGACTTTTGCTTCAGAGACG 240

5055_U46_Sanger GTGAAGGGTTATTAG 255
5055_U46_NGS GTGAAGGGTTATTAG 255

CI-HHV-6A Isolate 5814

5814_U46_Sanger ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTACATT 60
5814_U46_NGS ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTACATT 60

5814_U46_Sanger CTTGTTTTTCGCGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACC 120
5814_U46_NGS CTTGTTTTTCGCGCCGCTACGGAAGTGGACTTTTATTCTCCAGAGTGCCATTCGCATACC 120

5814_U46_Sanger TATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAAATCTTTTTTTATTTG 180
5814_U46_NGS TATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAAATCTTTTTTTATTTG 180

5814_U46_Sanger TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAGACTTTTGCTTCAGAGACG 240
5814_U46_NGS TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAGACTTTTGCTTCAGAGACG 240

5814_U46_Sanger GTGAAGGGTTATTAG 255
5814_U46_NGS GTGAAGGGTTATTAG 255

7.4.3 U47

CI-HHV-6A Isolate 2284

2284_U47_Sanger ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAGCAAATACTATTTTAGTCACACA 60
2284_U47_NGS ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAGCAAATACTATTTTAGTCACACA 60

2284_U47_Sanger TTCTACTTGTACAAATTCATTTGTGGTCAACTCACCCGACATGCTTCACATCTCGCAACTC 120
2284_U47_NGS TTCTACTTGTACAAATTCATTTGTGGTCAACTCACCCGACATGCTTCACATCTCGCAACTC 120

2284_U47_Sanger GGCCTTTTTCGGGCCCTTTTCGCGATAGTCATGCACCTCCGTAATCTAATAAAATACACG 180
2284_U47_NGS GGCCTTTTTCGGGCCCTTTTCGCGATAGTCATGCACCTCCGTAATCTAATAAAATACACG 180

2284_U47_Sanger TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATTGGAGCGACGAACAAGA 240
2284_U47_NGS TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATTGGAGCGACGAACAAGA 240

2284_U47_Sanger GAACATTTTACGACCTCCGAAACTTGATATACAAGTTTTTGTGAGACAAATCTATCCCTC 300
2284_U47_NGS GAACATTTTACGACCTCCGAAACTTGATATACAAGTTTTTGTGAGACAAATCTATCCCTC 300

2284_U47_Sanger GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC 360
2284_U47_NGS GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC 360

2284_U47_Sanger AAGTCGCGGTAACTTGTCTCCATTACAGAATTTACCGCGACAGAAACCAAATAGTG 420
2284_U47_NGS AAGTCGCGGTAACTTGTCTCCATTACAGAATTTACCGCGACAGAAACCAAATAGTG 420

2284_U47_Sanger CTATCGTCGACGGCAGCAAACAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT 480
2284_U47_NGS CTATCGTCGACGGCAGCAAACAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT 480

2284_U47_Sanger CTGTTTGCCACACTATCTACATATACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT 540
2284_U47_NGS CTGTTTGCCACACTATCTACATATACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT 540

2284_U47_Sanger AAAATTCCTGGGAGCACAGTTATTTGACATAAAAACGACCGAAGAATCCTTGGCGATGGCA 600
2284_U47_NGS AAAATTCCTGGGAGCACAGTTATTTGACATAAAAACGACCGAAGAATCCTTGGCGATGGCA 600

2284_U47_Sanger ATGAGCAAATTTTCCAACCTCGAATCTGACACGGTCATTGACTTCGTTTACGTCAGAAACC 660
2284_U47_NGS ATGAGCAAATTTTCCAACCTCGAATCTGACACGGTCATTGACTTCGTTTACGTCAGAAACC 660

2284_U47_Sanger TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720
2284_U47_NGS TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720

2284_U47_Sanger TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTTGGGA 780
2284_U47_NGS TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTTGGGA 780

2284_U47_Sanger CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT 840
2284_U47_NGS CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT 840

2284_U47_Sanger CTCACATCACCATATATCCCTAGTACCAACATACCAACCCAGAACCTCCCCCGTAACA 900
2284_U47_NGS CTCACATCACCATATATCCCTAGTACCAACATACCAACCCAGAACCTCCCCCGTAACA 900

2284_U47_Sanger AAAAATCTACAAAACCTGCATACAGACACCATAAAAAGTTACACCGAACACACCACCATA 960
2284_U47_NGS AAAAATCTACAAAACCTGCATACAGACACCATAAAAAGTTACACCGAACACACCACCATA 960

2284_U47_Sanger ACAACGCAAACAACGGAAAGCATCAAAAAAATAGTTAAACGTTTCCCTCGACCC 1020
2284_U47_NGS ACAACGCAAACAACGGAAAGCATCAAAAAAATAGTTAAACGTTTCCCTCGACCC 1020

2284_U47_Sanger ATGTACACCCCAACCGACATTTCCAACCTTTACAATCCGTCTTAACGCCACTATTAACC 1080
2284_U47_NGS ATGTACACCCCAACCGACATTTCCAACCTTTACAATCCGTCTTAACGCCACTATTAACC 1080

2284_U47_Sanger GAACAAAACACCGAAAACCCAAAAGTCCACCAAAAACCAACAAATTTTGAAAATACCACA 1140
2284_U47_NGS GAACAAAACACCGAAAACCCAAAAGTCCACCAAAAACCAACAAATTTTGAAAATACCACA 1140

2284_U47_Sanger ATCAGAATTTCCCAAAAACCTTTGAGAGCGCTACAGTAACAACAACAACAACCTCAAAGCTC 1200
2284_U47_NGS ATCAGAATTTCCCAAAAACCTTTGAGAGCGCTACAGTAACAACAACAACAACCTCAAAGCTC 1200

2284_U47_Sanger GAAAGCACCACTTCGCAACAATAGTAATCAAGGAAATAGCGGCAATACCTATTCTTCA 1260
2284_U47_NGS GAAAGCACCACTTCGCAACAATAGTAATCAAGGAAATAGCGGCAATACCTATTCTTCA 1260

2284_U47_Sanger CCAAAAACCTCTATTTATCTTAAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC 1320
2284_U47_NGS CCAAAAACCTCTATTTATCTTAAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC 1320

2284_U47_Sanger GCCGAACACACCACTCCGATTTTAAAGTTTACCCTTGCCAAAACACGGCAGCACATAC 1380
2284_U47_NGS GCCGAACACACCACTCCGATTTTAAAGTTTACCCTTGCCAAAACACGGCAGCACATAC 1380

2284_U47_Sanger ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAAATCCAGAGGACAACCTTGAAA 1440
2284_U47_NGS ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAAATCCAGAGGACAACCTTGAAA 1440

2284_U47_Sanger TCCTCAAACGAGCTACCTACCATTGACAGCTTGTCTGTCTCCTCAAACAAAACCTACCG 1500
2284_U47_NGS TCCTCAAACGAGCTACCTACCATTGACAGCTTGTCTGTCTCCTCAAACAAAACCTACCG 1500

2284_U47_Sanger TCGAATGTAACCTGCCAAAACCTGAAGTACACATAAATAACAATGCTTTACCATCTAGTAAT 1560
2284_U47_NGS TCGAATGTAACCTGCCAAAACCTGAAGTACACATAAATAACAATGCTTTACCATCTAGTAAT 1560

2284_U47_Sanger TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACCTAGAATGTCAGCACCC 1620
2284_U47_NGS TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACCTAGAATGTCAGCACCC 1620

2284_U47_Sanger ACTCAGGAAGAGATAAACCCACACAGAAATAGCACGAATGACACCAATTCCTTAACGCTCAC 1680
2284_U47_NGS ACTCAGGAAGAGATAAACCCACACAGAAATAGCACGAATGACACCAATTCCTTAACGCTCAC 1680

2284_U47_Sanger ACATCGGAAAAAATACACCTCAACGGCCCTTCACCGCTGAAACGTTCTTAACGACATCT 1740
2284_U47_NGS ACATCGGAAAAAATACACCTCAACGGCCCTTCACCGCTGAAACGTTCTTAACGACATCT 1740

2284_U47_Sanger TCAAAGCCTACCATCCTAACCTGGTCAAACCTACTAACAACAACGCCTAAGGAACCATTA 1800
2284_U47_NGS TCAAAGCCTACCATCCTAACCTGGTCAAACCTACTAACAACAACGCCTAAGGAACCATTA 1800

2284_U47_Sanger ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA 1860
2284_U47_NGS ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA 1860

2284_U47_Sanger ACTCAATCAGCCAAACTATCAAAGCTAACATCTCGTCACAAACGACTAACATCTACCCC 1920
2284_U47_NGS ACTCAATCAGCCAAACTATCAAAGCTAACATCTCGTCACAAACGACTAACATCTACCCC 1920

2284_U47_Sanger CAAACAATCACGGGACGATCTACAGAGGTTTAA 1953
2284_U47_NGS CAAACAATCACGGGACGATCTACAGAGGTTTAA 1953

CI-HHV-6A Isolate 5055

5055_U47_Sanger ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA 60
5055_U47_NGS ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA 60

5055_U47_Sanger TTCTACTTGTACAAAATTCATTTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC 120
5055_U47_NGS TTCTACTTGTACAAAATTCATTTGTGGTCAACTCACCCGACATGCTTCACATCTCGCGACTC 120

5055_U47_Sanger GGCCTTTTCTGGGCCTTTTCGCGATAGTCATGCACCTCCGCTAATCTAATAAAATACACG 180
5055_U47_NGS GGCCTTTTCTGGGCCTTTTCGCGATAGTCATGCACCTCCGCTAATCTAATAAAATACACG 180

5055_U47_Sanger TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATFGGAGCGACGAACAAAGA 240
5055_U47_NGS TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATFGGAGCGACGAACAAAGA 240

5055_U47_Sanger GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTGACACAAATCTATCCCTC 300
5055_U47_NGS GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTGACACAAATCTATCCCTC 300

5055_U47_Sanger GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC 360
5055_U47_NGS GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTCACAATGC 360

5055_U47_Sanger AAGTCTGCGGTAACTTGTCTCCATTACAGAATTTACCGGGACAGAAACCAAATAGTG 420
5055_U47_NGS AAGTCTGCGGTAACTTGTCTCCATTACAGAATTTACCGGGACAGAAACCAAATAGTG 420

5055_U47_Sanger CTATCGTCGACGGCAGCAAACAAATCTATCATCGCTAGCAGTTTTCTCTATTTTATCTT 480
5055_U47_NGS CTATCGTCGACGGCTGCAAACAAATCTATCATCGCTAGCAGTTTTCTCTATTTTATCTT 480

5055_U47_Sanger CTGTTTGCCACACTATCTACATATAACCGAGATCCGCCATGCGTAGAGTTACTACCATTT 540
5055_U47_NGS CTGTTTGCCACACTATCTACATATAACCGAGATCCGCCATGCGTAGAGTTACTACCATTT 540

5055_U47_Sanger AAAATCTGGGAGCACAGTTATTTGACATAAACTGACCGAAGAATCCTTGCGGATGGCA 600
5055_U47_NGS AAAATCTGGGAGCACAGTTATTTGACATAAACTGACCGAAGAATCCTTGCGGATGGCA 600

5055_U47_Sanger ATGAGCAAATTTTCCAACCTCGAATCTGACACGGTCATTGACTTCGTTACGTCAGAAACC 660
5055_U47_NGS ATGAGCAAATTTTCCAACCTCGAATCTGACACGGTCATTGACTTCGTTACGTCAGAAACC 660
*****.***.*

5055_U47_Sanger TTCCTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720
5055_U47_NGS TTCCTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720

5055_U47_Sanger TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTGGA 780
5055_U47_NGS TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTGGA 780

5055_U47_Sanger CGAACCATAGTAAACTTTGATTCGATACTAATACTACGACACCATCATCGAGTCAGCGTCT 840
5055_U47_NGS CGAGCCATCGTAAACTTTGATTCGATACTAATACTACGACACCATCATCGAGTCAGCGTCT 840
..*****

5055_U47_Sanger CTCACATCACCACATATCCCTAGTACCAACATACCAACCCAGAACCTCCCCCGTAACA 900
5055_U47_NGS CTCACATCACCACATATCCCTAGTACCAACATACCAACCCAGAACCTCCCCCGTAACA 900
*****.*****.***

5055_U47_Sanger AAAAAGCTTACAAAAGTCATACAGACACCATAAAAAGTTACACCGAACACACCACCATA 960
5055_U47_NGS AAAAAGCTTACAAAAGTCATACAGACACCATAAAAAGTTACACCGAACACACCACCATA 960
***** **

5055_U47_Sanger ACAACGCAAAACAACGAAAGCATCAAAAAATAGTTAAACGTTTCAGATTTTCCTCGACCC 1020
5055_U47_NGS ACAACGCAAAACAACGAAAGCATCAAAAAATAGTTAAACGTTTCAGATTTTCCTCGACCC 1020

5055_U47_Sanger ATGTACACCCCAACCGACATTTCAACTCTTACAATCCGTCTTAACGCCACTATTAACC 1080
5055_U47_NGS ATGTACACCCCAACCGACATTTCAACTCTTACAATCCGTCTTAACGCCACTATTAACC 1080

5055_U47_Sanger GAACAAAACACCGAAAACCCAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA 1140
5055_U47_NGS GAACAAAACACCGAAAACCCAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA 1140

5055_U47_Sanger ATCAGAATTTCCAAAACCTTTGAGAGCGCTACAGTAACAACAACAACTCAAAGCTC 1200
5055_U47_NGS ATCAGAATTTCCAAAACCTTTGAGAGCGCTACAGTAACAACAACAACTCAAAGCTC 1200
*****.*****.*** **

5055_U47_Sanger GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATAGCGGCAATACCTATTCTTCA 1260
5055_U47_NGS GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATAGCGGCAATACCTATTCTTCA 1260
*****.*****.*****

5055_U47_Sanger CCAAAAACTCTATTTATCTTAAAGAGCAAATCACAGCAGGTACAACAAAATTCACCGAC 1320
5055_U47_NGS CCAAAAACTCTATTTATCTTAAAGAGCAAATCACAGCAGGTACAACAAAATTCACCGAC 1320
*****.*****

5055_U47_Sanger GCCGAACACACCCTCCGATTTTAAAGTTTACCCTTGCCAAAACACGGCAGCACATAC 1380
5055_U47_NGS GCCGAACACACCCTCCGATTTTAAAGTTTACCCTTGCCAAAACACGGCAGCACATAC 1380

5055_U47_Sanger ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCAGAGGACAACCTTGAAA 1440
5055_U47_NGS ATGAGTCACAACACAGAAGTACAAAACATGACCGGCAAATTCAGAGGACAACCTTGAAA 1440

5055_U47_Sanger TCCTCAAACGAGCTACTTACCATTAGACGTTGTCTGTCTTCCAAAACAAAACCTACCG 1500
5055_U47_NGS TCCTCAAACGAGCTACTTACCATTAGACGTTGTCTGTCTTCCAAAACAAAACCTACCG 1500

5055_U47_Sanger TCGAATGTAAGTCCAAAACCTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT 1560
5055_U47_NGS TCGAATGTAAGTCCAAAACCTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT 1560

5055_U47_Sanger TCATCACAATCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACCTAGAATGTCAGCACC 1620
5055_U47_NGS TCATCACAATCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACCTAGAATGTCAGCACC 1620
***** **

5055_U47_Sanger ACTCAGAGAGATAAACACACAGAAATAGCACGAATGACACCAATCTTAAACGCTCAC 1680
5055_U47_NGS ACTCAGAGAGATAAACACACAGAAATAGCACGAATGACACCAATCTTAAACGCTCAC 1680
*****.*****

5055_U47_Sanger ACATCGGAAAAA---ACTACACCTCAACGGCCCTTACCCTGAAACGTTCTTAACGACA 1737
5055_U47_NGS ACATCGGAAAAATCAACTACACCTCAACGGTCCTTACCCTGAAACGTTCTTAACGACA 1740

5055_U47_Sanger TCTTCAAAGCCTACCATCTAACCTGGTCAAACCTACTAACAACAACGCCTAAGGAACCA 1797
5055_U47_NGS TCTTCAAAGCCCGCCATCTAACCTGGTCAAATTTATATCAACAACGCCTAAGGAACCA 1800

5055_U47_Sanger TTAACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAAT 1857
5055_U47_NGS TTAACGAATACAAGTCTAAGGTGGACAGATCATATCACTACACAGCTAACGACTAGCAAT 1860

5055_U47_Sanger AGAATCAATCAGCCAAACTATCAAAAAGCTAACATCTCGTCACAAACGACTAACATCTAC 1917
5055_U47_NGS AGAATCAATCAGCCAAACTAACAAAAGCTAACATCTCGTCACAAACGACTAACATCTAC 1920

5055_U47_Sanger CCCCAAACAATCACGGGACGATCTACAGAGTTTAA 1953
5055_U47_NGS CCCCAAACAATCACGGGACGATCTACAGAGTTTAA 1956

CI-HHV-6A Isolate 5814

5814_U47_Sanger ATGCATTTGGAGGTCATCGTCCAATCTATAAAAAAGCAAATACTATTTTAGTCACACA 60
5814_U47_NGS ATGCATTTGGAGGTCATCGTCCAATCTATAAAAAAGCAAATACTATTTTAGTCACACA 60

5814_U47_Sanger TTCTACTTGTACAAATTCATTTGGTCAACTACCCGACATGCTTCACATCTCGCAACTC 120
5814_U47_NGS TTCTACTTGTACAAATTCATTTGGTCAACTACCCGACATGCTTCACATCTCGCAACTC 120

5814_U47_Sanger GGCCTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAATACACG 180
5814_U47_NGS GGCCTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAATACACG 180

5814_U47_Sanger TCTGATCCCTTAGAAGCTTTCAAACCCTCAACCGCCACAATTTGGAGCGACGAACAAAGA 240
5814_U47_NGS TCTGATCCCTTAGAAGCTTTCAAACCCTCAACCGCCACAATTTGGAGCGACGAACAAAGA 240

5814_U47_Sanger GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTGACAGCAAATCTATCCCTC 300
5814_U47_NGS GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGTGACAGCAAATCTATCCCTC 300

5814_U47_Sanger GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTACAATGC 360
5814_U47_NGS GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATTACAATGC 360

5814_U47_Sanger AAGTCTGCGGTTAACTTGTCTCCATTACAGAATTTACCGGGACAGAAACCAAATAGTG 420
5814_U47_NGS AAGTCTGCGGTTAACTTGTCTCCATTACAGAATTTACCGGGACAGAAACCAAATAGTG 420

5814_U47_Sanger CTATCGTCGACGGCAGCAAACAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT 480
5814_U47_NGS CTATCGTCGACGGCAGCAAACAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT 480

5814_U47_Sanger CTGTTTGCCACACTATCTACATATAACCGCAGATCCGCATGCGTAGAGTTACTACCATT 540
5814_U47_NGS CTGTTTGCCACACTATCTACATATAACCGCAGATCCGCATGCGTAGAGTTACTACCATT 540

5814_U47_Sanger AAAATCTGGGAGCACAGTTATTTGACATAAACTGACCGAAGAATCCTTGCGGATGGCA 600
5814_U47_NGS AAAATCTGGGAGCACAGTTATTTGACATAAACTGACCGAAGAATCCTTGCGGATGGCA 600

5814_U47_Sanger ATGAGCAAATTTCCAACCTCGAATCTGACACGGTCATGACTTCGTTACAGTCAGAAACC 660
5814_U47_NGS ATGAGCAAATTTCCAACCTCGAATCTGACACGGTCATGACTTCGTTACAGTCAGAAACC 660

5814_U47_Sanger TTCTTTAATTACACAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720
5814_U47_NGS TTCTTTAATTACACAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT 720

5814_U47_Sanger TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTGGA 780

5814_U47_NGS TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTGGA 780

5814_U47_Sanger CGAACCATAGTAAACTTTGATTTCGATACTAATACGACACCATCATCGACGTCAGCGTCT 840
5814_U47_NGS CGAACCATAGTAAACTTTGATTTCGATACTAATACGACACCATCATCGACGTCAGCGTCT 840

5814_U47_Sanger CTCACATCACCACATATCCCTAGTACCAACATACCAACCCGAGAACCCTCCCCCGTAACA 900
5814_U47_NGS CTCACATCACCACATATCCCTAGTACCAACATACCAACCCGAGAACCCTCCCCCGTAACA 900

5814_U47_Sanger AAAAAGTCTACAAAAGTGCATACAGACACCATAAAAAGTTACACCGAACACACCACCATA 960
5814_U47_NGS AAAAAGTCTACAAAAGTGCATACAGACACCATAAAAAGTTACACCGAACACACCACCATA 960

5814_U47_Sanger ACAACGCAAAACAACGGAAAGCATCAAAAAATAGTTAAACGTTTACAGATTTTCTCGACCC 1020
5814_U47_NGS ACAACGCAAAACAACGGAAAGCATCAAAAAATAGTTAAACGTTTACAGATTTTCTCGACCC 1020

5814_U47_Sanger ATGTACACCCCAACCGACATTTCAACTCTTACAATCCGTCTTAACGCCACTATTAAACC 1080
5814_U47_NGS ATGTACACCCCAACCGACATTTCAACTCTTACAATCCGTCTTAACGCCACTATTAAACC 1080

5814_U47_Sanger GAACAAAACACCGAAAACCCAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA 1140
5814_U47_NGS GAACAAAACACCGAAAACCCAAAAGTCCACCAAAACCAACAAATTTTGAAAATACCACA 1140

5814_U47_Sanger ATCAGAATTCCCAAAAACCTTTGAGAGCGCTACAGTAAACAACAACAACAACAACAACA 1200
5814_U47_NGS ATCAGAATTCCCAAAAACCTTTGAGAGCGCTACAGTAAACAACAACAACAACAACAACA 1200

5814_U47_Sanger GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA 1260
5814_U47_NGS GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA 1260

5814_U47_Sanger CCAAAAACTCTATTTATCTTAAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC 1320
5814_U47_NGS CCAAAAACTCTATTTATCTTAAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC 1320

5814_U47_Sanger GCCGAACACACCACCTCCGATTTTAAAGTTTACCACCTGGCAAAAACACGGCAGCACATAC 1380
5814_U47_NGS GCCGAACACACCACCTCCGATTTTAAAGTTTACCACCTGGCAAAAACACGGCAGCACATAC 1380

5814_U47_Sanger ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAAATTCAGAGGACAACCTTGAAA 1440
5814_U47_NGS ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAAATTCAGAGGACAACCTTGAAA 1440

5814_U47_Sanger TCCTCAAACGAGCTACCTACCATTAGACGTTGTCTGTCTCAC TCCAAAACAAAACCTACCG 1500
5814_U47_NGS TCCTCAAACGAGCTACCTACCATTAGACGTTGTCTGTCTCAC TCCAAAACAAAACCTACCG 1500

5814_U47_Sanger TCGAATGTAAGTGCAAAACCTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT 1560
5814_U47_NGS TCGAATGTAAGTGCAAAACCTGAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT 1560

5814_U47_Sanger TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACCTAGAAATGTCAGCACCC 1620
5814_U47_NGS TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACCTAGAAATGTCAGCACCC 1620

5814_U47_Sanger ACTCAGAAAGAGATAAACACACAGAAATAGCACGAATGACACCAATCTTAAACGCTCAC 1680
5814_U47_NGS ACTCAGAAAGAGATAAACACACAGAAATAGCACGAATGACACCAATCTTAAACGCTCAC 1680

5814_U47_Sanger ACATCGGAAAAAAGTACACCTCAACGGCCCTTACCGCTGAAACGTTCTTAAACGACATCT 1740
5814_U47_NGS ACATCGGAAAAAAGTACACCTCAACGGCCCTTACCGCTGAAACGTTCTTAAACGACATCT 1740

5814_U47_Sanger TCAAAGCTTACCATCTAACCTGGTCAAACCTACTAACAACAACGCCTAAGGAACCATTA 1800
5814_U47_NGS TCAAAGCTTACCATCTAACCTGGTCAAACCTACTAACAACAACGCCTAAGGAACCATTA 1800

5814_U47_Sanger ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA 1860

5814_U47_NGS ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA 1860

 5814_U47_Sanger ACTCAATCAGCCAAACTATCAAAGCTAACATCTCGTCACAAACGACTAACATCTACCCC 1920
 5814_U47_NGS ACTCAATCAGCCAAACTATCAAAGCTAACATCTCGTCACAAACGACTAACATCTACCCC 1920

 5814_U47_Sanger CAAACAATCACGGGACGATCTACAGAGGTTTAA 1953
 5814_U47_NGS CAAACAATCACGGGACGATCTACAGAGGTTTAA 1953

7.4.4 U51

CI-HHV-6A Isolate 2284

2284_U51_Sanger ATGGGAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTATGGTTGGGTA 60
 2284_U51_NGS ATGGGAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTATGGTTGGGTA 60

 2284_U51_Sanger TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTGTGCTGACGGTTAGGTTAATAGT 120
 2284_U51_NGS TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTGTGCTGACGGTTAGGTTAATAGT 120

 2284_U51_Sanger TTTAAAGTCGGTCGAGAATACGCTGTGTTACATTTGCCGGGATGAGCTTTAATTTGTTTT 180
 2284_U51_NGS TTTAAAGTCGGTCGAGAATACGCTGTGTTACATTTGCCGGGATGAGCTTTAATTTGTTTT 180

 2284_U51_Sanger TTACTTCCGATAAAGATGGGATTTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC 240
 2284_U51_NGS TTACTTCCGATAAAGATGGGATTTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC 240

 2284_U51_Sanger GCAATCTTCTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGTFT 300
 2284_U51_NGS GCAATCTTCTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGTFT 300

 2284_U51_Sanger ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC 360
 2284_U51_NGS ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC 360

 2284_U51_Sanger AAGGCTCTGGCAAAGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT 420
 2284_U51_NGS AAGGCTCTGGCAAAGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT 420

 2284_U51_Sanger CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTCCTC 480
 2284_U51_NGS CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTCCTC 480

 2284_U51_Sanger GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTGGTTTATTCGGTCATGATTCCA 540
 2284_U51_NGS GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTGGTTTATTCGGTCATGATTCCA 540

 2284_U51_Sanger GTTATGCTGGTGTTTATTTACCTATACAGCAAAAATTTTCATGCTTTTGAAAGATGAACCT 600
 2284_U51_NGS GTTATGCTGGTGTTTATTTACCTATACAGCAAAAATTTTCATGCTTTTGAAAGATGAACCT 600

 2284_U51_Sanger TCGACCGTGAAGTACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG 660
 2284_U51_NGS TCGACCGTGAAGTACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG 660

 2284_U51_Sanger AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTTGC CGCATATTTTT 720
 2284_U51_NGS AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTTGC CGCATATTTTT 720

 2284_U51_Sanger ATGGCGCTTCCGTGTTTTAAAGTTTTATATATACCCGTTATGGCCTATTGCATGGCATGT 780
 2284_U51_NGS ATGGCGCTTCCGTGTTTTAAAGTTTTATATATACCCGTTATGGCCTATTGCATGGCATGT 780

 2284_U51_Sanger GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA 840
 2284_U51_NGS GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA 840

```

*****
2284_U51_Sanger      TGTTTTTCGTGTACGCGGCCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA 900
2284_U51_NGS         TGTTTTTCGTGTACGCGGCCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA 900
*****

2284_U51_Sanger      AAATGA 906
2284_U51_NGS         AAATGA 906
*****

```

CI-HHV-6A Isolate 5055

```

5055_U51_Sanger      ATGGGGAAAGAAACGAAGTC'TTTGGCTTGGCCAGCCACTGCGGAGTTTATGGTTGGGTA 60
5055_U51_NGS         ATGGGGAAAGAAACGAAGTC'TTTGGCTTGGCCAGCCACTGCGGAGTTTATGGTTGGGTA 60
*****

5055_U51_Sanger      TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTACTGACGGTTAGGTTAATAGT 120
5055_U51_NGS         TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTACTGACGGTTAGGTTAATAGT 120
*****

5055_U51_Sanger      TTTAAAGTCGGTCGAGAATACGCTGTGTTACATTTGCCGGGATGAGCTTTAATTTGTTT 180
5055_U51_NGS         TTTAAAGTCGGTCGAGAATACGCTGTGTTACATTTGCCGGGATGAGCTTTAATTTGTTT 180
*****

5055_U51_Sanger      TTACTTCCGATAAAGATGGGATGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTC 240
5055_U51_NGS         TTACTTCCGATAAAGATGGGATGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTC 240
*****

5055_U51_Sanger      GCAATCTTCTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGT 300
5055_U51_NGS         GCAATCTTCTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGT 300
*****

5055_U51_Sanger      ATGGCGATCGAGCGTATCAACCACTTCTGTTACAGCACACCGTGTGTTGAACGAGAAT 360
5055_U51_NGS         ATGGCGATCGAGCGTATCAACCACTTCTGTTACAGCACACCGTGTGTTGAACGAGAAT 360
*****

5055_U51_Sanger      AAGGCTCTGGCAAAGTGTGTTTCCCGATAGTTGGATCATATCGGGAGTGCAGCTCT 420
5055_U51_NGS         AAGGCTCTGGCAAAGTGTGTTTCCCGATAGTTGGATCATATCGGGAGTGCAGCTCT 420
*****

5055_U51_Sanger      CAGATGCTGAATAATTACAAAGCAACAGCTTTCGAGAACGAGACCCACAGTGTTC 480
5055_U51_NGS         CAGATGCTGAATAATTACAAAGCAACAGCTTTCGAGAACGAGACCCACAGTGTTC 480
*****

5055_U51_Sanger      GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGAT 540
5055_U51_NGS         GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGAT 540
*****

5055_U51_Sanger      GTTATGCTGGTGTATTTACCTATACAGCAAAAATTTTCATGCTTTTGAAAGATGAA 600
5055_U51_NGS         GTTATGCTGGTGTATTTACCTATACAGCAAAAATTTTCATGCTTTTGAAAGATGAA 600
*****

5055_U51_Sanger      TCGACCGTACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCT 660
5055_U51_NGS         TCGACCGTACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCT 660
*****

5055_U51_Sanger      AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTTGC CGCATAT 720
5055_U51_NGS         AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTTGC CGCATAT 720
*****

5055_U51_Sanger      ATGGCGCTTCCGTGTTTTAAAAGTTTATTATATACCCGTCATGGCCTATTGCATGG 780
5055_U51_NGS         ATGGCGCTTCCGTGTTTTAAAAGTTTATTATATACCCGTCATGGCCTATTGCATGG 780
*****

5055_U51_Sanger      GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTA 840
5055_U51_NGS         GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTA 840
*****

5055_U51_Sanger      TGTTTTTCGTGTACGCGGCCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATG 900
5055_U51_NGS         TGTTTTTCGTGTACGCGGCCGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATG 900
*****

```


5055_U51_Sanger AAATGA 906
5055_U51_NGS AAATGA 906

CI-HHV-6A Isolate 5814

5814_U51_Sanger ATGGGAAAGAAACGAAGTCCTTGGCTTGGCCAGCCACTGCGGAGTTTATGGTTGGGTA 60
5814_U51_NGS ATGGGAAAGAAACGAAGTCCTTGGCTTGGCCAGCCACTGCGGAGTTTATGGTTGGGTA 60

5814_U51_Sanger TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTACTGACGGTTAGGTTAATAGT 120
5814_U51_NGS TTTATTTTTTCGTCCATACAGTTGTGTACAATGGTTTACTGACGGTTAGGTTAATAGT 120

5814_U51_Sanger TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTAAATGTTTT 180
5814_U51_NGS TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTAAATGTTTT 180

5814_U51_Sanger TTACTCCGATAAAGATGGGATTTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC 240
5814_U51_NGS TTACTCCGATAAAGATGGGATTTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC 240

5814_U51_Sanger GCAATCTTCTCTACATCGATGATTTTTCTATCTATTCTCCTCGTGGTCATTGGTGT 300
5814_U51_NGS GCAATCTTCTCTACATCGATGATTTTTCTATCTATTCTCCTCGTGGTCATTGGTGT 300

5814_U51_Sanger ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC 360
5814_U51_NGS ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC 360

5814_U51_Sanger AAGGCTCTGGCAAAGTGTGTTTCCCGATAGTTGGATCATATCGGGAGTCAAGCTCTT 420
5814_U51_NGS AAGGCTCTGGCAAAGTGTGTTTCCCGATAGTTGGATCATATCGGGAGTCAAGCTCTT 420

5814_U51_Sanger CAGATGCTGAATAATTACAAAGCAACAGCTTTCGAGAACGAGACCCACAGTGT 480
5814_U51_NGS CAGATGCTGAATAATTACAAAGCAACAGCTTTCGAGAACGAGACCCACAGTGT 480

5814_U51_Sanger GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGATTCCA 540
5814_U51_NGS GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGATTCCA 540

5814_U51_Sanger GTTATGCTGGTGT 600
5814_U51_NGS GTTATGCTGGTGT 600

5814_U51_Sanger TCGACCGTACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG 660
5814_U51_NGS TCGACCGTACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG 660

5814_U51_Sanger AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTGCGCGATATTTTT 720
5814_U51_NGS AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTGCGCGATATTTTT 720

5814_U51_Sanger ATGGCGCTTCCGTGTTTTAAAAGTTTATTATATACCCGTTATGGCCTATTGCATGGCATGT 780
5814_U51_NGS ATGGCGCTTCCGTGTTTTAAAAGTTTATTATATACCCGTTATGGCCTATTGCATGGCATGT 780

5814_U51_Sanger GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA 840
5814_U51_NGS GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA 840

5814_U51_Sanger TGTTTTTCGTGTACGCGGCGGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA 900
5814_U51_NGS TGTTTTTCGTGTACGCGGCGGAAAAGGAGTCGGATTTAGAGGTTGGAATAAAAATGTTA 900

5814_U51_Sanger AAATGA 906
5814_U51_NGS AAATGA 906

7.4.5 U65

Ci-HHV-6A Isolate 2284

```
2284_U65_Sanger      ATGGCGATATCGACTTTTCAATTGGCGATTGGGTTATCTTAGAAATTTCTGCAGAAT 60
2284_U65_NGS         ATGGCGATATCGACTTTTCAATTGGCGATTGGGTTATCTTAGAAATTTCTGCAGAAT 60
*****

2284_U65_Sanger      GAATGTAACGGTTCAGAATTTGTAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120
2284_U65_NGS         GAATGTAACGGTTCAGAATTTGTAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120
*****

2284_U65_Sanger      ACATCGTCTCCTACATTTCTCGTTGAATAATAAACCTAAGAAATTTTGCTTGCATTGTGAG 180
2284_U65_NGS         ACATCGTCTCCTACATTTCTCGTTGAATAATAAACCTAAGAAATTTTGCTTGCATTGTGAG 180
*****

2284_U65_Sanger      ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTTCAGCCTTGCGGTAAACGGCATACAT 240
2284_U65_NGS         ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTTCAGCCTTGCGGTAAACGGCATACAT 240
*****

2284_U65_Sanger      TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAATCAGTTCGGGAAGGGCTC 300
2284_U65_NGS         TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAATCAGTTCGGGAAGGGCTC 300
*****

2284_U65_Sanger      TATTACTATATATTGGAATTGGAAGCATAACCCTTATCGATTGAACTTTATCCCAGAGA 360
2284_U65_NGS         TATTACTATATATTGGAATTGGAAGCATAACCCTTATCGATTGAACTTTATCCCAGAGA 360
*****

2284_U65_Sanger      TATAATTCCGACTGTGTTACAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT 420
2284_U65_NGS         TATAATTCCGACTGTGTTACAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT 420
*****

2284_U65_Sanger      TGCTCTATTGTGTGTCCCGAAGAGGCAATTCGCCTCACAGTAAAAGGGTCCGGGGACAAT 480
2284_U65_NGS         TGCTCTATTGTGTGTCCCGAAGAGGCAATTCGCCTCACAGTAAAAGGGTCCGGGGACAAT 480
*****

2284_U65_Sanger      AAATTGACTCCCTTAGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540
2284_U65_NGS         AAATTGACTCCCTTAGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540
*****

2284_U65_Sanger      ATCTATACTTTTGCACCTCTTACGATCTTTTCTTAAGTGTATGACAAATCCATCTTT 600
2284_U65_NGS         ATCTATACTTTTGCACCTCTTACGATCTTTTCTTAAGTGTATGACAAATCCATCTTT 600
*****

2284_U65_Sanger      CCATCTCTGGCAAAAATTTATTTTGTATGATAGCTTGCGAATCCGAAGATTGTGTCTTT 660
2284_U65_NGS         CCATCTCTGGCAAAAATTTATTTTGTATGATAGCTTGCGAATCCGAAGATTGTGTCTTT 660
*****

2284_U65_Sanger      TGTAAGATCACAACAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTAAT 720
2284_U65_NGS         TGTAAGATCACAACAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTAAT 720
*****

2284_U65_Sanger      CAAGAACTGTTTTTGTCTACACCGTGTAAAGAAAAAATGGCTAATATTAACAATCCG 780
2284_U65_NGS         CAAGAACTGTTTTTGTCTACACCGTGTAAAGAAAAAATGGCTAATATTAACAATCCG 780
*****

2284_U65_Sanger      GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATATGTATCCCGAA 840
2284_U65_NGS         GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATATGTATCCCGAA 840
*****

2284_U65_Sanger      AGAAAAGCATCGTTATCACTGGACATTAATTTTACGCTCATGGGTACTCCGGCGACGAC 900
2284_U65_NGS         AGAAAAGCATCGTTATCACTGGACATTAATTTTACGCTCATGGGTACTCCGGCGACGAC 900
*****

2284_U65_Sanger      CCTTATGCGTTAAAAATGTGTTAATTTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG 960
2284_U65_NGS         CCTTATGCGTTAAAAATGTGTTAATTTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG 960
*****

2284_U65_Sanger      CTGATTATTTTGTGCTGTCCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008
2284_U65_NGS         CTGATTATTTTGTGCTGTCCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008
*****
```

CI-HHV-6A Isolate 5055

5055_U65_Sanger ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATTTTCTGCAGAAT 60
5055_U65_NGS ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATTTTCTGCAGAAT 60

5055_U65_Sanger GAATGTAAC TGGTTCAGAATTTGTAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120
5055_U65_NGS GAATGTAAC TGGTTCAGAATTTGTAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120

5055_U65_Sanger ACATCGTCTCCTACATTTCTCGTTGAATAATAAACCTAAGAAATTTTGCATACATTGTGAG 180
5055_U65_NGS ACATCGTCTCCTACATTTCTCGTTGAATAATAAGCCTAAGAAATTTTGCATGCATTGCGAG 180

5055_U65_Sanger ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTTCAGCCTTGCAGTAAACGGCATAACAT 240
5055_U65_NGS ATTGTAATACTCAAGCGAAGTGAAGAATTTATGTTTCAGCCTTGCAGTAAATGGCATAACAT 240

5055_U65_Sanger TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAAGCAGTCCGGAAGGGCTC 300
5055_U65_NGS TTTGGGCAGTTTTTAACCGGAAAAATGAAATTTAATAAGAAAGCAGTCCGGAAGGGCTC 300

5055_U65_Sanger TATTACTATATATTGGAATTTGGGAAGCATAACCCCTTATCGATTGAACTTTATCCCGAGA 360
5055_U65_NGS TATTACTATATATTGGAATTTGGGAAGCATAACCCCTTATCGATTGAACTTTATCCCGAGA 360

5055_U65_Sanger TATAATTCGACTGTGTACAAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT 420
5055_U65_NGS TATAATTCGACTGTGTACAAACATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT 420

5055_U65_Sanger TGCTCTATTGTGTGCCGAAGAGGCAATTCGCTCAGATAAAGGGTCCGGGGACAAT 480
5055_U65_NGS TGCTCTATTGTGTGCCGAAGAGGCAATTCGCTCAGATAAAGGGTCCGGGGACAAT 480

5055_U65_Sanger AAATTGACTCCCTTAGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540
5055_U65_NGS AAATTGACTCCCTTAGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540

5055_U65_Sanger ATCTATACTTTTGCAC TCGCTTACGATCTTTTCTTA ACTTGTATGACAAATCCATCTTT 600
5055_U65_NGS ATCTATACTTTTGCAC TCGCTTACGATCTTTTCTTA ACTTGTATGACAAATCCATCTTT 600

5055_U65_Sanger CCATCTCTGGCAAAAATTTATTTTGTATGATAGCTTGC GAATCCGAAGATTGTGTCTTT 660
5055_U65_NGS CCATCTCTGGCAAAAATTTATTTTGTATGATAGCTTGC GAATCCGAAGATTGTGTCTTT 660

5055_U65_Sanger TGTAAAGATCACAACAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTAAT 720
5055_U65_NGS TGTAAAGATCACAACAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTAAT 720

5055_U65_Sanger CAAGAAACCTGTTTTGCTACACCCGTGTAAGAAAAAATGGCTAATATTAACAATCCG 780
5055_U65_NGS CAAGAAACCTGTTTTGCTACACCCGTGTAAGAAAAAATGGCTAATATTAACAATCCG 780

5055_U65_Sanger GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTTATGTATCCCGAA 840
5055_U65_NGS GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTTATGTATCCCGAA 840

5055_U65_Sanger AGAAAAGCATCGTTATCACTGGACATTAATTTCTTACGCTCATGGGTACTCCGGCGACGAC 900
5055_U65_NGS ATAAAAGCATCGTTATCACTGGACATTAATTTCTTACGCTCATGGGTACTCCGGTGCACGAC 900
* *****

5055_U65_Sanger CCTTATGCGTTAAAAATGTGTTAATTTGGATACCAAGTCAGGATCAGCGCAGCTCTGAGCAGG 960
5055_U65_NGS CCTTATGCGTTAAAAATGTGTTAATTTGGATACCAAGTCAGGATCAGCGCAGCTCTGAGCAGG 960

5055_U65_Sanger CTGATTATTTTGTGCTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008
5055_U65_NGS CTGATTGTTTTGTGCTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008

CI-HHV-6A Isolate 5814

5814_U65_Sanger ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATTTCTGCAGAAT 60
5814_U65_NGS ATGGCGATATCGACTTTTTCAATTGGCGATTGGGTTATCTTAGAAATTTCTGCAGAAT 60

5814_U65_Sanger GAATGTAAC TGGTTCAGAATTTGTA AAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120
5814_U65_NGS GAATGTAAC TGGTTCAGAATTTGTA AAAAAACATTCTATCGCGAATATCGTAGCGTTGCG 120

5814_U65_Sanger ACATCGTCTCCTACATTTCTCGTTGAATAATAAACCTAAGAAATTTTGCATACATTGTGAG 180
5814_U65_NGS ACATCGTCTCCTACATTTCTCGTTGAATAATAAACCTAAGAAATTTTGCATACATTGTGAG 180

5814_U65_Sanger ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTTCAGCCTTGGCGTAAACGGCATAACAT 240
5814_U65_NGS ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTTCAGCCTTGGCGTAAACGGCATAACAT 240

5814_U65_Sanger TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAAGCAGTTCCGGAAGGGCTC 300
5814_U65_NGS TTTGGGCAGTTTTTGACCGGAAGAATGAAATTTAATAAGAAAGCAGTTCCGGAAGGGCTC 300

5814_U65_Sanger TATTACTATATATTGGAATTGGGAAGCATAACCCCTTATCGATTGAACTTTATCCCAGAGA 360
5814_U65_NGS TATTACTATATATTGGAATTGGGAAGCATAACCCCTTATCGATTGAACTTTATCCCAGAGA 360

5814_U65_Sanger TATAATTCGACTGTGTTACAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAAT 420
5814_U65_NGS TATAATTCGACTGTGTTACAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAAT 420

5814_U65_Sanger TGCTCTATTGTGTGCCGAAGAGGCAATTCGTCTCACAGTAAAGGGTCCGGGGACAAT 480
5814_U65_NGS TGCTCTATTGTGTGCCGAAGAGGCAATTCGTCTCACAGTAAAGGGTCCGGGGACAAT 480

5814_U65_Sanger AAATTGACTCCCTTAGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540
5814_U65_NGS AAATTGACTCCCTTAGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT 540

5814_U65_Sanger ATCTATACTTTGCACTCGCTTACGATCTTTTCCTAACTGTTATGACAAATCCATCTTT 600
5814_U65_NGS ATCTATACTTTGCACTCGCTTACGATCTTTTCCTAACTGTTATGACAAATCCATCTTT 600

5814_U65_Sanger CCATCTCTGGCAAAAATTTATTTTGTATGATAGCTTGCGAATCCGAAGATTGTGCTTTT 660
5814_U65_NGS CCATCTCTGGCAAAAATTTATTTTGTATGATAGCTTGCGAATCCGAAGATTGTGCTTTT 660

5814_U65_Sanger TGTAAGATCACAACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGGCTCTCTAAT 720
5814_U65_NGS TGTAAGATCACAACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGGCTCTCTAAT 720

5814_U65_Sanger CAAGAAACCTGTTTTTGTCTACACACCGTGTAAAGAAAAAATGGCTAATATTAACAATCCG 780
5814_U65_NGS CAAGAAACCTGTTTTTGTCTACACACCGTGTAAAGAAAAAATGGCTAATATTAACAATCCG 780

5814_U65_Sanger GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA 840
5814_U65_NGS GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA 840

5814_U65_Sanger AGAAAAGCATCGTTATCACTGGACATTAATTCCTACGCTCATGGGTACTCCGGCGACGAC 900
5814_U65_NGS AGAAAAGCATCGTTATCACTGGACATTAATTCCTACGCTCATGGGTACTCCGGCGACGAC 900

5814_U65_Sanger CCTTATGCGTTAAAATGTGTTAATTTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG 960
5814_U65_NGS CCTTATGCGTTAAAATGTGTTAATTTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG 960

5814_U65_Sanger CTGATTATTTGTCTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008
5814_U65_NGS CTGATTATTTGTCTGTCCTGTGTGTAAGCGTGTGGTAATGGACTAA 1008

7.4.6 U83

CI-HHV-6A Isolate 2284

```
2284_U83_Sanger      ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT 60
2284_U83_NGS        ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT 60
*****

2284_U83_Sanger      ATTTTCGTCTTCTGTCTTGTAGGATGTTTGTGTGTGTCATGGATTGGTCCGCTGCCGTA 120
2284_U83_NGS        ATTTTCGTCTTCTGTCTTGTAGGATGTTTGTGTGTGTCATGGATTGGTCCGCTGCCGTA 120
*****

2284_U83_Sanger      CCTGGAAAACAGAGCCTTTTAGAAAACCTTTTGTATGCAATCATGATTA AAAAGCTAAAA 180
2284_U83_NGS        CCTGGAAAACAGAGCCTTTTAGAAAACCTTTTGTATGCAATCATGATTA AAAAGCTAAAA 180
*****

2284_U83_Sanger      AGTTGTTCTGTCTGCTTACCCGCTGTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT 240
2284_U83_NGS        AGTTGTTCTGTCTGCTTACCCGCTGTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT 240
*****

2284_U83_Sanger      GCATCGCCGACAAGTCTTGAATTAGGATTGTCGAAATTAGACAAAGAATCATGA 294
2284_U83_NGS        GCATCGCCGACAAGTCTTGAATTAGGATTGTCGAAATTAGACAAAGAATCATGA 294
*****
```

CI-HHV-6A Isolate 5055

```
5055_U83_Sanger      ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT 60
5055_U83_NGS        ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT 60
*****

5055_U83_Sanger      ATTTTCGTCTTCTGTCTTGTAGGATGTTTGTGTGTGTCATGGATTGGTCCGCTGCCGTA 120
5055_U83_NGS        ATTTTCGTCTTCTGTCTTGTAGGATGTTTGTGTGTGTCATGGATTGGTCCGCTGCCGTA 120
*****

5055_U83_Sanger      CCTGGAAAACAGAGCCTTTTAGAAAACCTTTTGTATGCAATCATGATTA AAAAGCTAAAA 180
5055_U83_NGS        CCTGGAAAACAGAGCCTTTTAGAAAACCTTTTGTATGCAATCATGATTA AAAAGCTAAAA 180
*****

5055_U83_Sanger      AGTTGTTCTGTCTGCTTACCCGCTGTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT 240
5055_U83_NGS        AGTTGTTCTGTCTGCTTACCCGCTGTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT 240
*****

5055_U83_Sanger      GCATCGCCGACAAGTCTTGAATTAGGATTGTCGAAATTAGACAAAGAATCATGA 294
5055_U83_NGS        GCATCGCCGACAAGTCTTGAATTAGGATTGTCGAAATTAGACAAAGAATCATGA 294
*****
```

CI-HHV-6A Isolate 5814

```
5814_U83_Sanger      ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT 60
5814_U83_NGS        ATGGCTATCGGATTTATATGTAGTTCCCCCGATGCGGAGCTGTTTTCCGAAAAATCACGT 60
*****

5814_U83_Sanger      ATTTTCGTCTTCTGTCTTGTAGGATGTTTGTGTGTGTCATGGATTGGTCCGCTGCCGTA 120
5814_U83_NGS        ATTTTCGTCTTCTGTCTTGTAGGATGTTTGTGTGTGTCATGGATTGGTCCGCTGCCGTA 120
*****

5814_U83_Sanger      CCTGGAAAACAGAGCCTTTTAGAAAACCTTTTGTATGCAATCATGATTA AAAAGCTAAAA 180
5814_U83_NGS        CCTGGAAAACAGAGCCTTTTAGAAAACCTTTTGTATGCAATCATGATTA AAAAGCTAAAA 180
*****

5814_U83_Sanger      AGTTGTTCTGTCTGCTTACCCGCTGTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT 240
5814_U83_NGS        AGTTGTTCTGTCTGCTTACCCGCTGTTGGAGCAGGGCTCGATGTGTGATATGGCAGAT 240
*****

5814_U83_Sanger      GCATCGCCGACAAGTCTTGAATTAGGATTGTCGAAATTAGACAAAGAATCATGA 294
5814_U83_NGS        GCATCGCCGACAAGTCTTGAATTAGGATTGTCGAAATTAGACAAAGAATCATGA 294
*****
```


7.5 CI-HHV-6A Conserved Gene Set Nucleotide Sequences

U3

>2284_U3

```
ATGGAAACGACAAAGAACCGGCGTGTGCGATTTACTGGAGGCCTTCATGAGCCAGCGCAA
GATTCCTACGATGTGTGGGAATGCTGAGCTGCAGTCTCGGGAGAGGACTAAGACACGTCAA
CGTGTCTATAACGAGTCGAAGAGAGATCTCGTAGATCTCTTGCGCGTCTATACGTGCCTG
AATAGCGTTTCAATGTTTACGTTTCGTAAATTTTCGGCAGGCGAATTCCTTTGGCTTGGCCT
GAAGGATACCAGCTGGTCATCAATAACTGTAGAGACGAGAATGAATATAACCGATGAAAAG
TTGGACGAAC TGGCCGGTCTGGTGTGTTGCCCGGAGCGCCTCGTCTACTGGGATACGTG
AAGAAAAGGGGGGAATCGGATCGGAGCCGTGTTACTGGTTTTTCGAAAGGTGACATAGTC
GTTCTGCTCGGCGGCGCCGGTCTGTGTTTATGCGCACACGTGGTTATTGCCGCAGCAACTG
TGCCCGTCTCGGTGACACCATTGACGATTTTCTACGTAAAGGTCTGAAGAGGTTCTATTAC
GCACATCAGTTGGTGTCTCATCGCTGCAGTTTGTGGTGGAGGACCCCGAGGTGGACGGCGTC
CGCAGTTGTCTGGGACGTGCTGTGCTTTTCGAGACCGGCACATAGGTCTGTTCTTTTCGCCCTG
AGGTGGCCCAAGAACGAGACTCTGCTCTTTTCATCGGCGGCGGGATTCTCTTGCCTTCGTT
ATGTGCGACGAGGCGGGCGCAGGCTCGCGGAGATGTGTTTCTTCGGCAGCTTCGGATAC
AAGTACTTTGGGGACGAAGGTAGGATTTCTCTCTACGTGACCGACGACGGCCAGATTTTC
GGATTCAACGACAACGACGAAGACGGACGGCCGCGCTTCGTGGCGGAAGATTTTCAGCAG
TTTAGAAGGATAGGAGTCACTCAGTATTATAAGAGCTACGTTTTCCGTAGGGAGCAGCCC
GAATGGGCGCTTCTTCCCTACCTGTCTATCTATCGCAATGTTTTTATCAGAAAACCTGGAGG
CAGGCCGACGCTGATGTGCTGTTTTTAATAGCTAGGAACGACGAGCGCAAGAAATTTCCG
GACGCTGGTACTAGGAGTGCCGTACACGACGTGGTTCGATGA
```

>5055_U3

```
ATGGAAACGACAAAGAACCGGCGTGTGCGATTTACTGGAGGTCTTCATGAGCCAGCGCAA
GATTCCTACGATGTGTGGGAATGCTGAGGTGCAGTCTCGGGAGAGGACCAAGACACGTCAA
CGTGTCTATAACGAGTCGAAGAGAGATCTCGTAGATCTCTTGCGCGTCTATACGTGCCTG
AATAGCGTTTCAATGTTTACGTTTCGTAAATTTTCGGCAGGCGAATTCCTTTGGCTTGGCCT
GAAGGATACCAGCTGGTCATCAATAACTGTAGAGACGAGAATGAATATGCCGATGAAAAG
TTGGACGAAC TGGCCGGTCTGGTGTGTTGCCCGGAGCGCCTCGTCTACTGGGATACGTG
AAGAAAAGGGGGGGATCGGATCGGAGCCGTGTTACTGGTTTTTCGAAAGGTGACATAGTC
GTTCTGCTCGGCGGCGCCGGTCTGTGTTTATGCGCACACGTGGTTATTGCCGCACCAACTG
TGCCCGTCTCGGTGACACCATTGACGATTTTCTACGTAAAGGTCTGAAGAGGTTCTATTAC
GCACATCAGCTGGTGTCTCATCGCTGCAGTTTGTGGTGGAGGACCCCGAGGTGGACGGCGTC
CGCAGTTGTCTGGGACGTGCTGTGTTTTTCGAGACCGGCACATAGGTCTGTTCTTTTCGCCCTG
AGGTGGCCCAAGAACGAGACTCTGCTCTTTTCATCGGCGGCGGGATTCTTTTGCCTTCGTC
ATGTGCGACGAGGCGGGCGCAGGCTCGCGGAGATGTGTTTCTTCGGCAGCTTCGGATAC
AAGTACTTTGGGGACGAAGGTGGATTTCTCTCTACGTGACCGACGATGGACAGATTTTC
GGATTCAACGACAACGACGAAGACGGACGGCCGCGCTTCGTGGCGGAAGACTTTTCAGCAG
TTTAGACGGATAGGAGTCACTCAGTATTATAAAAGCTACGTTTTCCGTAGGGAGCAGCCC
GAATGGGCGCTTCTTCCCTACCTGTCTATCTATCGCAATGTTTTTATCAGAAAACCTGGAGG
CAGGCAGACGCTGATGTGCTGTTTTTAATAGCTAGGAACGACGAGCGCAAGAAATTTCCG
GACGCTGGTACTAGGAGTGCCGTACACGACGTGGTTCGATGA
```

>5814_U3

```
ATGGAAACGACAAAGAACCGGCGTGTGCGATTTACTGGAGGCCTTCATGAGCCAGCGCAA
GATTCCTACGATGTGTGGGAATGCTGAGCTGCAGTCTCGGGAGAGGACTAAGACACGTCAA
CGTGTCTATAACGAGTCGAAGAGAGATCTCGTAGATCTCTTGCGCGTCTATACGTGCCTG
AATAGCGTTTCAATGTTTACGTTTCGTAAATTTTCGGCAGGCGAATTCCTTTGGCTTGGCCT
GAAGGATACCAGCTGGTCATCAATAACTGTAGAGACGAGAATGAATATAACCGATGAAAAG
TTGGACGAAC TGGCCGGTCTGGTGTGTTGCCCGGAGCGCCTCGTCTACTGGGATACGTG
AAGAAAAGGGGGGAATCGGATCGGAGCCGTGTTACTGGTTTTTCGAAAGGTGACATAGTC
GTTCTGCTCGGCGGCGCCGGTCTGTGTTTATGCGCACACGTGGTTATTGCCGCAGCAACTG
TGCCCGTCTCGGTGACACCATTGACGATTTTCTACGTAAAGGTCTGAAGAGGTTCTATTAC
GCACATCAGTTGGTGTCTCATCGCTGCAGTTTGTGGTGGAGGACCCCGAGGTGGACGGCGTC
```

CGCAGTTGTCGGGACGTGCTGTGCTTTTCGAGACCGGCACATAGGTCGTTCTTTTCGCCCTG
AGGTGGCCCAAGAACGAGACTCTGCTCTTTTCATCGGCGGGGATTCTCTTGCCTTCGTT
ATGTGCGACGAGGCGCGGCGCAGGCTCGCGGAGATGTGTTTCTTCGGCAGCTTCGGATAC
AAGTACTTTGGGGACGAAGGTAGGATTTCTCTCTACGTGACCGACGACGGCCAGATTTTC
GGATTCAACGACAACGACGAAGACGGACGGCCGCGCTTCGTGGCGGAAGATTTTCAGCAG
TTTAGAAGGATAGGAGTCACTCAGTATTATAAGAGCTACGTTTTCCGTAGGGAGCAGCCC
GAATGGGCGCTTCTTCCCTACCTGTCATCTATCGCGAATGTTTTATCAGAAAACCTTGGAGG
CAGGCCGACGCTGATGTGCTGTTTTTAATAGCTAGGAACGACGAGCGCAAGAAATTTCCG
GACGCTGGTACTAGGAGTGCCGTACACGACGTGGTTTCGATGA

U4

>2284_U4

ATGGAGTTATTGGACCACGACATTTATAAGGGACCCGTCCGAGAACGGGTCACGTATAACC
ATTCCCAACCATCCCTATCTCTCGCTTACCCTTCATCATTCCAGGGAACCTTGATGTTGAT
CTGAAAGACATAACGGAGGAGATGATTATAGATAGCGGCACGTTGACCGCAGAGGATCTG
TTTATGACCCGCGGCTTACGATTTTTGTGACGATAGTGTCTCTGGGCGGCTCTGGCGGAG
AAAGTTTCTGTTGAATTCCAGAGGCGTGAAAGACGACCATGGATCTCTCGGCTTTTATG
TCTCTCTTAGAACGGTTAACTTTTTGAGGACGAAACGTCAGTTTTTTATCGACTCCTGACC
AAGTGCACCGCGCTGGCGCACTTCTCGATTCTGGAGTACATCGTCGATGGGGAGAAGAGG
GATACTATATCGGCGCATTTTAACCGTCTTTTTGGAGTTGTTGGACTCCCTCTTCTGCAG
TTCTCATGTTGAGGAACAGGAGCGAGTCCAATACTCTGTTGACCCTGTTTCGATATTTTG
CCCAATCCCAAGGAGATCTATGGCGACGCGGCGTCCCCGACGAGTGTGCTCTTGAAACAC
TTTTTGTCCGATCACATCTTTGTGCGGTACGCTTCTTCTTATCGCTTTGTGTCTTCTGTG
CTCTGTAGTTGCGATCAGTGCCGGGCGTCTCTCTTTTCGGATCTATCTGGGTAAGAAAGAT
TCCCGGGCCGTGCACATATCCGATATTTCAGAATTCGATCCGGCTGATCTCATTTTTAGGC
CAGCTGCACCTCGACGCGCGTGAGGCGGTGCGTCTCCGGGCGGAGATCGACCGAGATCTC
GGCTCTTCTCTCATCTTAACCTCTATGAAAGGCGGCATTTGCCATTCTTCATGATAAG
CAGTTGGGGAGAGGGCATTTTGAACGTAACATCTTGAAGGTGTAAGCAACATCGTTCTG
TGCTGTTTTCTACTGAGACGCGTCAAACAGCGCATCGTGAGTGATCTCACTGCCATTGCT
CGTTTTCTTCGTAAGCGATCTCCGAAATGGAGCGTTGCATCGATAACGTTTCTGGCCTC
CGCGGTGTGAGGATCAAACGTTAGTTGTGAGTTCTCAGTTTCGAGATTAAGATCCCATG
CGCGTCCCTAGGTTGTGTTTTTTCTTTCTTGGAGATCGTGGCGATTCTCGTATCGGGTTAT
GACAAGGAACGGCATCGCGTTGCTGCTCTTTAGAGCATCTGTGCCTTAAAAAGATCTGT
CTCGGCGACCTGTGTGCATCCATGCGACTCGCTCGCGAGATAAGGCACGACGTCCTGGAG
GGCTTCTCAATACTCTGGAGTTATCTTACATGTGCAATCCGGTTAGGTTTTTTTCGTTTT
CACGACTGCAACCTTTACGGGACCCATATGTGGAGTGGGGTGTATCCGAACGTCGTTCTC
TACGCTGTGCGGGCGGCGGCTCGTTTTGACATCAGGCTGCAGGAGCACCGGAGACGTCAG
AAGCGAATAGTCTGTCAGACGTAGGCTTATAAGACGCGTGCGGCAGCGAGGGTTAGACATT
CGTGAAGCGCGCCGCTTCTAAGCTGGCTTGTGTGACCTTTATCTGA

>5055_U4

ATGGAGTTATTGGACCACGACATTTATAAGGGACCCGTCCGAGAACGGGTCACGTATAACC
ATTCCCAACCATCCCTATCTCTCGCTTACCCTTCATCATTCCAGGGAACCTTGATGTTGAT
CTAAAAGACATAACGGAGGAGATGATTATAGATAGCGGCACGTTGACCGCAGAGGATCTG
TTTATGACTCGCGGCTTACGATTTTTGTGACGATAGTGTCTCTGGGCGGCTCTGGCGGAG
AAAGTTTCTGTTGAATTCCAGAGGCGTGAAAGACGACCATGGATCTCTCGGCTTTTATG
TCTCTCTTAGAACGGTTAACTTTTTGAGGACGAAACGTCAGTTTTTTATCGACTCCTGACC
AAGTGCACCGCGCTGGCGCACTTCTCGATTCTGGAGTACATCGTCGATGGGGAAAAGAGG
GATACTATATCGGCGCATTTTAACCGTCTTTTTGGAGTTGTTGGACTCCCTCTTCTGCAG
TTCTCATGTTGAGGAACAGGAGCGAGTCCAATACTCTGTTGACCCTGTTTCGATATTTTG
CCCAATCCCAAGGAGATCTATGGCGACGCGGCGTCTCCGACGAGTGTGCTCTTGAAACAC
TTTTTGTCCGATCACATCTTTGTGCGGTACGCTTCTTCTTATCGCTTTGTGTCTTCTGTG
CTCTGTAGTTGCGATCAGTGCCGGGCGTCTCTCTTTTCGGATCTATCTGGGTAAGAAAGAT
TCCCGGGCCGTGCACATATCCGATATTTCAGAATTCGATCCGGCTGATCTTATTTTTAGGC
CAGTTACACCTGGACGCGTGTGAGGCGGTGCGTCTCCGGGCGGAGATCGACCGAGATCTC
GGCTCTTCTCTCATCTTAACCTCTATGAAAGGCGGCATTTGCCATTCTTCATGATAAG

CAGTTGGGGAGAGGGCATTGTTGAACGTAACATCTTGAAGGTGACTGCAACATCGTTCTG
TGCTGTTTTCTACTGAGACGCGTCAAACAGCGCATCGTGAGTGATCTCACTGCCATTGCT
CGTTTTCTTCGTAAAAGCGATCTCCGAAATGGAGCGTTGCATCGATACCGTTTTCTGGCCTC
CGTGGTGTCAAGATCAAACGTGTTAGTTGTCAGTTCTCAGTTTCGAGATTAAGATCCCATG
CGCGTCCCTAGGTTGTGTTTTTTCTTTCTTGAGATCGTGGCGATTCTCGTATCGGGTTAT
GACAAAGAACGGCATCGCGTTTCGTGCTCTCTTAGAGCATCTGTGCCTTAAAAAGATCTGT
CTCGGCGACCTGTGTGCATCCATGCGACTCGCTCGCGAGATAAGGCACGACGTCCTGGAG
GGCTTCCTCAATACTCTGGAGTTATCTTACATGTGCAATCCGGTTAGGTTTTTTCTGTTTT
CACGACTGCAACCTTACGGGACCCATATGTGGAGTGGGGTGTATCCGAACGTCGTTCCCT
TACGCTGTGCGCGGGCGGGCTCGTTTGACATCAGGCTGCAGGAGCACCGGAGACGTCAG
AAGCGAATAGTTCGTAGACGTAGGCTTATAAGACGCGTGCAGCAGCGAGGGTTAGACATT
CGTGAAGCGCGCCGCGTTCCTAAGCCGGCTTGTGTGACCTTTATCTGA

>5814_U4

ATGGAGTTATTGGACCACGACATTTATAAGGGACCCGTCGAGAACGGGTCACGTATAACC
ATTCCCAACCATCCCTATCTCTCGCTTACCCTTCATCATTCCAGGGAACCTTGATGTTGAT
CTGAAAGACATAACGGAGGAGATGATTATAGATAGCGGCACGTTGACCGCAGAGGATCTG
TTTATGACCCGCGGCTTACGATTTTGTGACGATAGTGTCTCTGGGCGGCTCTGGCGGAG
AAAGTTTCTGTTGAATTCCAGAGGCGTGGAAAGACGACCATGGATCTCTCGGCTTTTATG
TCTCTCTTAGAACGGTTAACTTTTGGAGACGAAACGTCAGTTTTTTATCGACTCCTGACC
AAGTGCACCGCGCTGGCGCACTTCTCGATTCTGGAGTACATCGTTCGATGGGGAGAAGAGG
GATACTATATCGGCGCATTTTAAACCGTCTTTTGGAGTTGTTGGACTCCCTCTTCTGCGAG
TTCTCATGTTGAGGAACAGGAGCGAGTCCAATACTCTGTTGACCCTGTTTCGATATTTTG
CCCAATCCCAAGGAGATCTATGGCGACGCGGCGTCCCGACGAGTGTGCTCTTGAAACAC
TTTTTGTCCGATCACATCTTTGTGCGGTACGCTTCTTCTTATCGCTTTGTGTCTTCTGTG
CTCTGTAGTTGCGATCAGTGCAGGCGGTCTCTCTTTTCGGATCTATCTGGGTAAGAAAGAT
TCCCGGGCCGTGCACATATCCGATATTTAGAAATTCGATCCGGCTGATCTCATTTTAGGC
CAGCTGCACCTCGACGCGGTGAGGCGGTGCGTCTCCGGGCGGAGATCGACCGAGATCTC
GGCTCTTCTCTCATCTTAACTCTATGGAAGGCGGCATTTGCCATTCTTCATGATAAG
CAGTTGGGGAGAGGGCATTGTTGAACGTAACATCTTGAAGGTGACTGCAACATCGTTCTG
TGCTGTTTTCTACTGAGACGCGTCAAACAGCGCATCGTGAGTGATCTCACTGCCATTGCT
CGTTTTCTTCGTAAAAGCGATCTCCGAAATGGAGCGTTGCATCGATACCGTTTTCTGGCCTC
CGCGTGTCAAGATCAAACGTGTTAGTTGTCAGTTCTCAGTTTCGAGATTAAGATCCCATG
CGCGTCCCTAGGTTGTGTTTTTTCTTTCTTGAGATCGTGGCGATTCTCGTATCGGGTTAT
GACAAGGAACGGCATCGCGTTTCGTGCTCTCTTAGAGCATCTGTGCCTTAAAAAGATCTGT
CTCGGCGACCTGTGTGCATCCATGCGACTCGCTCGCGAGATAAGGCACGACGTCCTGGAG
GGCTTCCTCAATACTCTGGAGTTATCTTACATGTGCAATCCGGTTAGGTTTTTTCTGTTTT
CACGACTGCAACCTTACGGGACCCATATGTGGAGTGGGGTGTATCCGAACGTCGTTCCCT
TACGCTGTGCGCGGGCGGGCTCGTTTGACATCAGGCTGCAGGAGCACCGGAGACGTCAG
AAGCGAATAGTTCGTAGACGTAGGCTTATAAGACGCGTGCAGCAGCGAGGGTTAGACATT
CGTGAAGCGCGCCGCGTTCCTAAGCTGGCTTGTGTGACCTTTATCTGA

U22

>2284_U22

ATGGTGCCTCAGGGTTGTAGTTTAGTATGGGTGAGTGCATTATACGTATCGGTGATAGCG
TCTTTGCATATTATTAACAATGAAAATTCAGTTTTTCATCGCAACACACAGTGAAACGGAG
TTAAGGCATTGGTTGATTTTTGTAAAAATGGCTCAGAGAAACGGAAGTCTTGGTGGCGG
ATGGCTAGTGTGCCATTATAATGCTTATTTTTGAAAGAGATATAGCCTTTTTTGTAAACCCT
CGATGCGTTATCGAGACCGCCATGGGGTCTAAGATATTATGTCGATAACAATAAAATATT
GGTGTGTTGTTTCGTGGACAATGATACTAAGTGTAAATGTTTCTTTCCGAGCGGTGTGCAG
TTACAATTACTAAATCAATCGGTGATGGAGTCAATTAGAACTAAAACGTATGTGGTGGAT
TACGCTAGAAAAACTACAGAGAGGGGTGATTGCTTTATTTCTGTAGCATTCTGCCGAAAA
GAGAGACGACGCTTTTTATCCCGCTGTGAGCGCTTTGTTTATTACTGTATCTCGGTTTAC
CTTTTTGCTGTTGTGGTTTTTTGTTCTTGCTGGTTTTGCTTTGGATCCGTTGTTTAAATATG
TGGGCGTAA

>5055_U22

ATGGTGCCTCAGGGTTGTAGTTTAGTATGGGTGAGTGCATTATACGTGTCCGGTGATAGCG
TCTTTGCATATTATTAACAATGAAAATTCAGTTTTTATCGCAACACACAGTGAAACGGAG
TTAAGGCATTGGTTGATTTTTGTAAAAATGACTCAGAGAAACGGAACCTGCTTGGTGGCGG
ATGGCTAGTGTGCCTATTAATGCCTATTTTCGAAAGAGATATAGCCTTTTTGTTTAAACCT
CGATGCGTTATCGAGACCGCCATGGGGTCTAAGATATTATGTCGATACAATAAAAAATATT
GGTGTGTGTTCGTGGACAATGATACTACGTGTAATGTTTCCTTTCCGAGCGGTGTGCAG
TTACAATTACTAAATCAATCGGTGATGGAGTCAATTAGAACTAAAACGTATGTGGTAGAT
TACGCTAGAAAAACTACAGAGAGGGGTGATTGCTTCATTTCTGTAGCATTCTGCCGAAAA
GAGAGACGACGCTTTTTATCCCCTGTGAGCGCTTCGTTTATTATTGTATCTCGGTTTAC
CTTTTTGCTGTTGTGGTTCCTTTGTTCTTGCTGGTTTTGCTTTGGATCCGTTGTTTAAATATG
TGGGCGTAA

>5814_U22

ATGGTGCCTCAGGGTTGTAGTTTAGTATGGGTGAGTGCATTATACGTATCCGGTGATAGCG
TCTTTGCATATTATTAACAATGAAAATTCAGTTTTTATCGCAACACACAGTGAAACGGAG
TTAAGGCATTGGTTGATTTTTGTAAAAATGGCTCAGAGAAACGGAACCTGCTTGGTGGCGG
ATGGCTAGTGTGCCTATTAATGCCTATTTTGAAGAGATATAGCCTTTTTGTTTAAACCT
CGATGCGTTATCGAGACCGCCATGGGGTCTAAGATATTATGTCGATACAATAAAAAATATT
GGTGTGTGTTCGTGGACAATGATACTAAGTGTAAATGTTTCCTTTCCGAGCGGTGTGCAG
TTACAATTACTAAATCAATCGGTGATGGAGTCAATTAGAACTAAAACGTATGTGGTGGAT
TACGCTAGAAAAACTACAGAGAGGGGTGATTGCTTTATTTCTGTAGCATTCTGCCGAAAA
GAGAGACGACGCTTTTTATCCCCTGTGAGCGCTTTGTTTATTACTGTATCTCGGTTTAC
CTTTTTGCTGTTGTGGTTTTTTTTGTTCTTGCTGGTTTTGCTTTGGATCCGTTGTTTAAATATG
TGGGCGTAA

U24A

>2284_U24A

ATGTTGCTTATAAGCAAACCTCTATTCCCCATTCCCTAGAGGTGCCGATGAATTCTGGGAA
GGAACATACATTCTCGTAGTTGCATTCGTGCTCGCATTTCTCGTTTACTCCGATTTTCTT
AGTAATTTAAGTCCGTTTGGAGAAATTTTATCCTCGTCTTGTATTTCAACGTGA

>5055_U24A

ATGTTGCTTATAAGCAAACCTCTATTCCCCATTCCCTAGAGGTGCCGATGAATTCTGGGAA
GGAACATACATTCTCGTAGTTGCATTCGTGCTCGCATTTCTCGTTTACTCCGATTTTCTT
AGTAATTTAAGTCCGTTTGGAGAAATTTTATCCTCGCTTGTATTTCAACGTGA

>5814_U24A

ATGTTGCTTATAAGCAAACCTCTATTCCCCATTCCCTAGAGGTGCCGATGAATTCTGGGAA
GGAACATACATTCTCGTAGTTGCATTCGTGCTCGCATTTCTCGTTTACTCCGATTTTCTT
AGTAATTTAAGTCCGTTTGGAGAAATTTTATCCTCGTCTTGTATTTCAACGTGA

U25

>2284_U25

ATGGGTTTGTCTTCAAGCAAAGACATCGATTTGCTGGTAAATTTTATCGAAAATCGCCAG
GGCGCCACTTTAGCTTTGCCCTGGCCCGAGGATTATCGGTTGACTCTGCATAGTGTAGAA
AACATCCCAGAAATTTACGTGAAGATGTTCAAAATGGGGCAAGACCTATATGTGTTGC
GGAGGCGAATTGGTTGTGGTTGGGGTTATACATAAGGCAAAAACGAGAACCTGCAGGGGA
CCGATTGTCTTTCAGGGGGCGCGGGGGCACATCTATGTGTATAATGGTTTTTTTTGACAGA
AGTCTTTATCACGTGGCTTCTTCTTTTACGATTTGTTTTCTAATGGTTTTGCGGTTTTTT
TATCCGATTTATGAGACGTGTGATTATGCACTGGACAGTACTGTGGCGCTGGATATGATT
GCTCACAGTAAATCGTTTTCTGAGCTCTTACATTACCGAAATGAAAGAAAGAATGCTTTT
TTTACCTTGAAAACGTATCCGTACAAGACCTTTGTTTCGCTTTTGCAATCTGAGCATGACT
GAGTTTTCCAGTAGACATTTGATTCAGTGGCGAAGGAACTGCAGACGAGTCTACTAGAT
GTAGTGTATTATTGTACAGCATAATTTTTTTCGGTAACTGGAGGGAGCTCGTAGTAGTGT

GACGGCCACGGTATGCTTTTTTGCCTCGACCGTGAGGAGAGCCTTCTCTTCATAGCTCGC
AACATGTCCGATTTTTTAAAAATTGGATGTCTGCGTTATAATGAGAATCGTCGGTTGCAT
ACGCAATGGTTTACGCAAGATACGGACTATATTAAGCAGGTGGATGAGATGTTTTCTAGA
GACGTCTCTGTCTCTCCGTGAGCACTGTCTAGAGGTCTGCGCCGAGAGCGAGGATTATTA
AAGATATGTACGTCTCTAGTCAGGGGGATTAATTGTATTGAAAGAGGGTAG

>5055_U25

ATGGGTTTATCTTCAAGCAAAGACATCGATTTGCTGGTAAATTTTATCGAAAATCGCCAG
GGCGCCACGTTAGCTTTACCCCTGGCCCGAGGATTATCGATTGACTCTGCATAGTGTAGAA
AACATCCCAGAAATTTACAGTGAAGATGTTCAAAATTTGGGCGAAGACCTATATGTGTTGC
GGAGGCGAATTTGGTTGTGGTTGGGGTTATACATAAGGCAAAAACGAGAACCTGCAGGGGA
CCGATTGTCTTTCAGGGGGCGCGGGGGCACATCTATGTGTATAATGGTTTTTTTTGACAGA
AGTCTTTATCACGTGGCTTCTTCTTTTACGATTTGTTTTCTAATGGTTTTGCGGTTTTTT
TATCCGATTTATGAGACGTGTGATTATGCACCTGGACAGTACTGTGGCGCTGGATATGATT
GCTCACAGTAAATCGTTTTCTGAGCTCTTACATTACCGAAATGAAAGAAAGAATGCTTTTT
TTTACCTTGAAAACATATCCGTACAAGACCTTTGTTTCGCTTTTGCAATCTGAGCATGACT
GAGTTTTCTAGTAGACATTTGATTCAGTGGCGAAGGAACTGCAGACGAGTCTACTAGAT
GTAGTGTATTATTGTACAGCATAATTTTTTTCGGTGACTGGAGGGAGCTCGTAGTAGTGT
GACGGCCACGGTATGCTTTTTTGCCTCGACCGTGAGGAGAGCCTTCTCTTCATAGCTCGC
AACATGTCCGATTTTTTAAAAATTGGATGTCTGCGTTATAATGAGAATCGTCGGTTGCAT
ACGCAATGGTTTACGCAAGATACGGACTATATTAAGCAGGTGGATGAGATGTTTTCTAGA
GACGTCTCTGTCTCTCCGTGAGCACTGTCTAGAGGTCTGCGCCGAGAGCGAGGATTATTA
AAGATATGTACGTCTCTAGTCAGGGGGATTAATTGTATTGAAAGAGGGTAG

>5814_U25

ATGGGTTTGTCTTCAAGCAAAGACATCGATTTGCTGGTAAATTTTATCGAAAATCGCCAG
GGCGCCACTTTAGCTTTGCCCTGGCCCGAGGATTATCGGTTGACTCTGCATAGTGTAGAA
AACATCCCAGAAATTTACAGTGAAGATGTTCAAAATTTGGGCGAAGACCTATATGTGTTGC
GGAGGCGAATTTGGTTGTGGTTGGGGTTATACATAAGGCAAAAACGAGAACCTGCAGGGGA
CCGATTGTCTTTCAGGGGGCGCGGGGGCACATCTATGTGTATAATGGTTTTTTTTGACAGA
AGTCTTTATCACGTGGCTTCTTCTTTTACGATTTGTTTTCTAATGGTTTTGCGGTTTTTT
TATCCGATTTATGAGACGTGTGATTATGCACCTGGACAGTACTGTGGCGCTGGATATGATT
GCTCACAGTAAATCGTTTTCTGAGCTCTTACATTACCGAAATGAAAGAAAGAATGCTTTTT
TTTACCTTGAAAACGTATCCGTACAAGACCTTTGTTTCGCTTTTGCAATCTGAGCATGACT
GAGTTTTCCAGTAGACATTTGATTCAGTGGCGAAGGAACTGCAGACGAGTCTACTAGAT
GTAGTGTATTATTGTACAGCATAATTTTTTTCGGTAACTGGAGGGAGCTCGTAGTAGTGT
GACGGCCACGGTATGCTTTTTTGCCTCGACCGTGAGGAGAGCCTTCTCTTCATAGCTCGC
AACATGTCCGATTTTTTAAAAATTGGATGTCTGCGTTATAATGAGAATCGTCGGTTGCAT
ACGCAATGGTTTACGCAAGATACGGACTATATTAAGCAGGTGGATGAGATGTTTTCTAGA
GACGTCTCTGTCTCTCCGTGAGCACTGTCTAGAGGTCTGCGCCGAGAGCGAGGATTATTA
AAGATATGTACGTCTCTAGTCAGGGGGATTAATTGTATTGAAAGAGGGTAG

U27

>2284_U27

ATGTGTTGGTCATTTCAATTTGTTTTTTAAAGCACATAAGGCGCGAGTTGGTGCTAGGACT
TCGTTCTTGACAGAGATGGAGCGCGGTAGTCGAGATCATCACCGTGATCACCGTGATCAT
CGGGAACATAGAGACACGAGAGAACCGCCAACCTAGCTTTTTCATATGAAAAGCTGGAAA
ACTATTCATAAATCGCTTAAAGCGTTTGCAAACTGTTAAAAGAGAATACTACAGTGACT
TTCACGCCGACCCGTCGATAATAATTCAGTCTGCTAAAAATCATCTCGTGCAAAAAGCTG
ACTATTCAGGCGAATGTTTTGTTTTTGTCTAGATACGGATCGTTTTTTAACGAAAACCATT
AATAATCATATTCACATGTTTTGAAAGTTTTATGAACATCATCTCGAATCCTGAGGTTACG
AAGATGTACATTCAGCATGATAGTGATCTGTATACGAGGGTTTTGGTAACGGCTTCCGAT
ACATGTACACAGGCGTCGGTTCCCTGTGTGCACGGACAAGAAGTGGTGCGAGACAGCGGG
AGATCGCCGTTGAGGATTGACCTTGATCATTCGACCGTTTTCCGATGTGTTGAAATGGCTT
TCGCCGTAACATAAGACTAAACGCTCTGGTAAATCTGACGCTTTAATGGCGCACATTATA
GTACAGGTTAATCCTCCGACTATAAAATTCGTGACAGAGATGAATGAACTGGAGTTTTTCG

AACAGCAATAAGGTTATATTTTACGATGTTAAGAGCATGCGATTTAATCTATCTGCCAAA
AATTTACAGCAGGCTTTAAGTATGTGTGCTGTAATCAAGACGTCGTGTAGTTTACGTACG
GTAGCCGCTAAGGACTGTAATTTGATCTTAACTTCCAAAAGCACGTTGTTGACGGTGGAA
GCCTTTTTTAACCTCAGGAACAGCTCAAAGAGGAATCTCGCTTTGAGCGAATGGGTAAACAA
GATGACGGAAAAGGGGATAGAAGTCATAAGAACGAAGACGGAAGTGCCTCGGCATCAAAA
CAGGAAATGCAATATAAAAATAACCAACTACATGGTCCCCGCTAAAAACGGAACCTGCGGGC
TCTAGCTTATTTAACGAGAAGGAAGATAGCGAAAAGTGACGATTCCATGCATTTTGAATTC
AGCTCCAATCCCAAGAGGCAGAGATGCGTCTGTA

>5055_U27

ATGTGTTGGTCATTTCAATTTGTTTTTTAAAGCACATAAGACGCGAGTTGGTGCTAGGACT
TCGTTCTTGACAGAGATGGAGCGCGGTAGTCGAGATCATCATCGTGATCACCGTGATCAT
CGGGAACATAGAGAAAACGAGAGAACCGCCGACCCTAGCTTTTCATATGAAAAGCTGGAAA
ACTATTAATAAATCGCTTAAAGCGTTTGCAAACTGTTAAAAGAGAATACTACAGTGACT
TTCACGCCGCAGCCGTCGATAATAATTCAGTCTGCTAAAAATCATCTCGTGCAAAAAGCTG
ACTATTCAGGCGGAATGTTTTGTTTTTGTGATACGGATCGTTTTTTAACGAAAACCATT
AATAATCATATTCACCTTTTTGAAAGTTTTATGAACATCATCTCGAATCCTGAGGTTACG
AAGATGTACATTCAGCATGATAGTGATCTGTATACGAGGGTTTTGGTAACGGCTTCCGAT
ACATGTACACAGGCGTCGGTTCCCTGTGTGCACGGACAAGAAGTGGTGCGAGACACCGGG
AGATCGCCGTTGAGGATTGACCTTGATCATTGACCGTTTTCCGATGTGTTGAAATGGCTT
TCACCTGTAACATAAGACTAAACGCTCTGGTAAATCTGACGCTTTAATGGCGCACATTATA
GTACAGGTTAATCCTCCGACTATAAAAATTTGTGACAGAGATGAATGAACTGGAGTTTTCG
AACAGCAATAAGGTTATATTTTACGATGTTAAGAACATGCGATTTAATCTATCTGCCAAA
AATTTACAGCAGGCTTTAAGTATGTGTGCTGTAATCAAGACGTCGTGTAGTTTACGTACG
GTAGCCGCTAAGGACTGTAATTTGATATTAACTTCCAAAAGCACGTTGTTGACGGTGGAA
GCCTTTTTTAACCTCAGGAACAGCTCAAAGAGGAATCTCGCTTTGAAACGAATGGGTAAACAA
GATGACGGAAAAGGGGATAGGAGTCATAAGAACGAAGACGGAAGTGCCTTGGCATCAAAA
CAGGAAATGCAATATAAAAATAACCAACTACATGGTCCCCGCTAAAAACGGAACCTGCGGGC
TCTAGCTTATTTAACGAGAAGGAAGATAGCGAAAAGTGACGATTCCATGCATTTTGAATTC
AGCTCTAATCCCAATCCCAAGAGGCAGAGATGCGTCTGTA

>5814_U27

ATGTGTTGGTCATTTCAATTTGTTTTTTAAAGCACATAAGGCGCGAGTTGGTGCTAGGACT
TCGTTCTTGACAGAGATGGAGCGCGGTAGTCGAGATCATCACCGTGATCACCGTGATCAT
CGGGAACATAGAGACACGAGAGAACCGCCAACCCTAGCTTTTCATATGAAAAGCTGGAAA
ACTATTCATAAATCGCTTAAAGCGTTTGCAAACTGTTAAAAGAGAATACTACAGTGACT
TTCACGCCGCAGCCGTCGATAATAATTCAGTCTGCTAAAAATCATCTCGTGCAAAAAGCTG
ACTATTCAGGCGGAATGTTTTGTTTTTGTGATACGGATCGTTTTTTAACGAAAACCATT
AATAATCATATTCACCTGTTTTGAAAGTTTTATGAACATCATCTCGAATCCTGAGGTTACG
AAGATGTACATTCAGCATGATAGTGATCTGTATACGAGGGTTTTGGTAACGGCTTCCGAT
ACATGTACACAGGCGTCGGTTCCCTGTGTGCACGGACAAGAAGTGGTGCGAGACAGCGGG
AGATCGCCGTTGAGGATTGACCTTGATCATTGACCGTTTTCCGATGTGTTGAAATGGCTT
TCGCCGTAACATAAGACTAAACGCTCTGGTAAATCTGACGCTTTAATGGCGCACATTATA
GTACAGGTTAATCCTCCGACTATAAAAATTCGTGACAGAGATGAATGAACTGGAGTTTTCG
AACAGCAATAAGGTTATATTTTACGATGTTAAGAGCATGCGATTTAATCTATCTGCCAAA
AATTTACAGCAGGCTTTAAGTATGTGTGCTGTAATCAAGACGTCGTGTAGTTTACGTACG
GTAGCCGCTAAGGACTGTAATTTGATCTTAACTTCCAAAAGCACGTTGTTGACGGTGGAA
GCCTTTTTTAACCTCAGGAACAGCTCAAAGAGGAATCTCGCTTTGAGCGAATGGGTAAACAA
GATGACGGAAAAGGGGATAGAAGTCATAAGAACGAAGACGGAAGTGCCTCGGCATCAAAA
CAGGAAATGCAATATAAAAATAACCAACTACATGGTCCCCGCTAAAAACGGAACCTGCGGGC
TCTAGCTTATTTAACGAGAAGGAAGATAGCGAAAAGTGACGATTCCATGCATTTTGAATTC
AGCTCCAATCCCAAGAGGCAGAGATGCGTCTGTA

U28

>2284_U28

ATGAAGAGAAAAGAGCGTCGTATTAATAAAGATTATGGCTACAACAGAAAATGCGTTTTGT

CATTACGAAGCTTCGCAAAAACGTTTTTGTATTCTCAATATTCATGCGCTTCTGTGTTG
TATGAACGAGTTCGTGACATCGCTAAGATAATCGATCGTCTCGACTCAGGTTTAGATGCT
TGGTGTTCGCGGGATGCCATAATTTCTGTTTTGAGAGCTACACATTGCGTTCCACGAGTG
GATCGAATGTTAGGAAGATGGTATCTGAAGACGAGTGTTCGATTTTTTGTCCGGAC
GATTTAATTTTATCATGCCCAAATGTGATTATGCCAAATGTATTGAACTTTGTAAAGAAG
TATCGAGATTTTATTCGCAGCGTACTCTATAAAGTAAGCGTGAGTTGGAAAAACCAATAT
ATGCCCGGAGTTTTGGCCGCATCTCGCTTTTTAGAAAGAAATTTCCAATTTTAAACGGT
GTCGAAGAGAGTATCCATGCATCTATCTTCGTATGTGTGCTACTTTGACAGAGATCGTG
TTAAGAATTGGTTATCTGCGGGAGATATATCAGGAGAATCCTTACGTGATGTTTCAAGAA
CTGGCGTTTTCTTTTTACTCAAAAATGGGTTTTACCGTTTTCTGTATGACAAATTTG
GGCTCGTGGAAAAGGCTAATAGCACGGTTTTTGACGTGGCGATTTACAATACATGTTTA
TATTCGCTTGTGGATTTTACAACAGTAAACGGCGAACATCTTTTTCCGGCTTTGTTAAAT
GGTTCCTAATAAAGCATGAATTTGACTCGTTATCAACAGGAAGCCAAGAATATTTTTGAG
ATATTGCTGTCTCAAATTCGGGTGGTGGAAACGTGATACCGATAAGACGGTTCAGTTGACA
GTCTATGTGGAAGTCTGGCATGTGTCTGCACCTAATGTGGCTGGACTTGTATGAAGCTTTG
CCTCAGACTTCTCGGGTCACATTTTGCTTGATCATTCTGGGATTTTCATGGATAGGTAC
GAGCTAAAAAGGGCGCAGTGGTCTTTTTTCAGAAAAATATAGCTTTTGAATTGGGAAAG
TGTGACGAGATAACTTTCTCGACAAAATATTTGGAGTTTGAGCGGACAACAGATCATGCG
AAAATTACTATGTCTAGTTTTATAGAAAAAATTTGCCTTTGTTTAAAAGGGGGCAGGATG
GGACTTATTTCCGAAAAACGTATATCAATATTTCTATGATTCCTCATGTGCCTTTGTAT
TGCGGGGTGATTTTTTGGATGTTTTACCCGTCAGAGACGGTATTAACACTTGTGTTGAGA
ATGTTGCTTAATGTGGTACATTTCTTGGGTGACGAAGTAAGCGATGAGCTAACTGAGGAG
ATTGATTTCTGTCGTTTTGCAGTGTAAATTTTTTATGTTTAAATGAATTAAGGCGGGTGGTT
CGAAAAATGGTGTGGTTCGCTAATGCTGTGATTGATTATGCTGTAGAAAAATAAGATTTT
CTGTGTGAGGGAATCGAAGATGGACGAAGTCTAGGTATCTGTGTACGGGTCTTCATTCT
GTTTTCATGACTGTTCGGGTATCCTATGCACATCCAGATGCTCGTCGATTATATCGAATG
ATATGTGAACACATTTATTATACCTTGTGTTTCAACAAGTGTGGATTGTTGCATGAAAGGA
GCAGAGCCATGTAATTTATTTGACAGGTCAAATACGCACTGGGGATGTTATACTTTGAT
CATTTTGAGAATGTCGAATGCACGCTCCCGGAGGAATTTGTGGACAACGCTAAGGAAAGAC
GTGCTGATGCACGGAGTTCGCAATGTTTCACTGCGGACTGCCATGCAGAAAAGAA
TTCGATATCATCAACTCTTCTGAGTCTGTTCTGGCCGATGGAAGACAATAAAATCCTAAGG
CGTCAAATATCAAAGTCTGTAATTTGAAAAGACGGTTTTAAATGATGTCACCTTCAGTTTAT
TCGTCCGAATTTAAATCGTTATATATCCCGTGTATAACAATCTGTTGCTCAACAGATTT
AACAGCATCAGCAATATTTGAAGACTGTGGGATATAGGGTATTAACGCGGACACGAAC
CTATTTACGGACAAGGAGTTGGATGATTTAGCCGTTTTCAAAGATGGTTTTAGTTATCAT
TTGAATGATTTGATTGAGATGTATAAGAGTGGATTGCCTTTTTTTAGATCAGGGGCAGGCA
AATGTGTTTTATTTCAACGATACCTGTGAGTTTGAGACTGTTGCTACCGTTGCTGTATAAAA
ACCGGATTTAAAGTCTGCTATGTATAAGGTTTTGTGTAATTTCTGAGATGTATAAACACCTT
GATTTATCAAATCCGTTACCGTTGATCGGGAATGCTCGGATGGTGTGGTTATGCACGTA
AAGAACATTTTGTAG

>5055_U28

ATGAAGAGAAAAGAGCGTCGTATTAATAAAGATTATGGCTACAACAGAAAATGCGTTTTGT
CATTACGAAGCTTCGCAAAAACGTTTTTGTATTCTCAATATTCATGCGCTTCTGTTTTGT
TATGAACGGGTTCGTGACATCGCTAAGATAATCGATCGTCTCGACTCAGGTTTAGATGCT
TGGTGTTCGCGGGATGCCATAATTTCTGTTTTGAGAGCTACACATTGCGTTCCACGAGTG
GATCGAATGCTAGGAAGATGGTATCTGAAGACGAGTGTTCGATTTTTTGTCCGGAC
GATTTAATTTTATCATGCCCAAATGTGATTATGCCAAATGTATTGAACTTTGTAAAGAAG
TATCGAGATTTTATTCGCAGCGTACTCTATAAAGTAAGCGTGAGTTGGAAAAACCAATAT
ATGCCCGGAGTTCGGCCGCATCTCGCTTTTTAGAAAGAAATTTCCAATTTTAAACGGT
GTCGAAGAGAGTATCCATGCATCTATCTTCGTATGTGTGCTACTTTGACAGAGATCGTG
TTAAGAATTGGTTATCTGCGGGAGATATATCAGGAGAATCCTTACGTGATGTTTCAAGAA
CTGGCGTTTTCCCTTTTTACTCAAAAATGGATTTTACCGTTTTCTGTATGACAAATTTG
GGCTCGTGGAAAAGGCTAATAGCACGGTTTTTGACGTGGCGATTTACAATACATGTTTA
TATTCCTTGTGGATTTTACAATAGTAAACGGCGAACATCTTTTTCCGGCTTTGTTAAAC
GGTTCCTAATAAAGCATGAATTTGACTCGTTATCAACAGGAAGCCAAGAATATTTTTGAG
ATATTGCTGTCTCAAATTCGGGTGGTGGAAACGTGATACCGATAAGACGGTTCAGTTGACA
GTCTATGTGGAAGTCTGGCATGTGTCTGCACCTAATGTGGCTGGACTTGTATGAAGCTTTG

CCTCAGACTTCTCGGGTCACATTTTGCTTGATCATTCTGGGATTTTCATGGATAGGTAC
GAGCTGAAAAGGGCGCAGTGGTCTCTTTTTTCATAAAAATATAGCTTTTGAATTAGGAAAAG
TGTGACGAGATAACTTTTCTCGACAAAATATTTGGAGTTTGAGCGGACAACGGATCATGCG
AAAATTACTATGGCTAGTTTTGTAGAAAAAATTTGCCTTTGTTTTAAAAGGGGCGAGGATG
GGACTTATTTTCCGAAAAACGTATATCAATATTCTATGATTCCTCATGTGCCTTTGTAT
TGCGGGGGTGATTTTTTGGATGTTTTACCCGTCAGAGACGGTATTAACACTTGTATGAGA
ATGTTGCTTAATGTGGTACATTTCTTGGGTGACGAAGTAAGCGATGAGCTAACTGAGGAG
ATTGATTTCTGTCGTTTTGCAGTGTAATTTTTTATGTTTAAATGAATTAAGCGGGTGGTT
CGAAAAATGGTGTGGTTCGCTAATGCTGTGATTGATTATGCCGTAGAAAATAAAGATTTT
TTGTGTGAGGGAATCGAAGATGGACGAAGTCTAGGTATCTGTGTACGGGTCTTCATTCT
GTTTTCATGACTGTCCGGTTATCCTATGCACATCCAGATGCTCGTCGATTATATCGAATG
ATATGTGAACACATTTATTATACCTTGTGTTTGAACGAGTGTGGATTGTTGCATGAAAGGA
GCAGAGCCATGTAATTTATTTGACAGGTCAAATACGCACTGGGGATGTTATACTTTGAT
CATTTTGATAATGTCGAATGCACGCTCCCGGAGGAATTGTGGACAACGCTAAGGAAAAGAC
GTGCTGATGCACGGAGTTCGCAATATTCATTTCAACGCGGTTACTGCCATGCAGAAAAGAA
TTCGATATCATCAACTCTTCTGAATCGTTCTGGCCGATGGAAGACAATAAAATCCTAAGG
CAGTCAAATATCAAAGTCGTAATTGGAAGACGGTTTTAAATGATGTCACCTTCAGTTTAT
TCGTCGGAATTTAAATCGTTATATATCCCGTGTATAACAATCTGTTGCTCAACAGATTT
AACAGCATCAGCAATATTTGAAGACTGTGGGATATAGGGTATTAACGTTGGACACGAAC
CTATTTACGGACAAGGAGTTGGATGATTTAGCCGTTTTTCAAAGATGGTTTTAGTTATCAT
TTGAATGATTTGATTGAGATGTATAAGAGTGGATTGCCTTTTTTATGATCAGGGGCGAGCA
AATGTGTTTTTATTTCAACGATACTGTGAGTTTTGAGACTCTTGCTACCGTTGCTGTATAAA
GCCGGATTTAAAGTCGCTATGTATAAGGTTTTGTGTAATCTGAGATGTATAAACACCTT
GATTTATCAAATCCGTTACCGTTGATCGGGAATGTTCCGGATGGTGTGGTTATGCACGTA
AAGAACATTTTGTAG

>5814_U28

ATGAAGAGAAAAGAGCGTCGTATTAATAAAGATTATGGCTACAACAGAAAATGCGTTTGT
CATTACGAAGCTTCGCAAAAACGTTTTTGTATTCTCAATATTCATGCGCTTCTGTGTTG
TATGAACGAGTTCGTGACATCGCTAAGATAATCGATCGTCTCGACTCAGGTTTAGATGCT
TGGTGTTCGCGGATGCCATAATTTCTGTTTTGAGAGCTACACATTGCGTTCCACGAGTG
GATCGAATGTTAGGAAGATGGTATCTGAAGACGAGTGTTTTTTACGATTTTTGTCCGGAC
GATTTAATTTTATCATGCCCAAATGTGATTATGCCAAATGTATTGAACTTTGTAAAGAAG
TATCGAGATTTTATTCGCAGCGTACTCTATAAAGTAAGCGTGAGTTGGA AAAACCAATAT
ATGCCCGGAGTTTTGGCCGCATCTCGCTTTTTTAGAAGAAAATTTCCAATCTTTAAACGGT
GTCGAAGAGAGTATTCATGCATCTATCTTCGTATGTGTGCTACTTTGACAGAGATCGTG
TTAAGAATTGGTTATCTGCGGGAGATATATCAGGAGAATCCTTACGTGATGTTTGAAGAA
CTGGCGTTTTCTCTTTTTTACTCAAAAATGGGTTTTACCGTTTTCTGTATGACAAAATTTG
GGCTCGTGGAAAAGGCTAATAGCACGGTTTTTACGTTGGCGATTTACAATACATGTTTA
TATTCGCTTGTGGATTTTACAACAGTAAACGGCGAACATCTTTTTCCGGCTTTGTTAAAT
GGTTCATAATATAAGCATGAATTTGACTCGTTATCAACAGGAAGCCAAGAATATTTTTGAG
ATATTGCTGTCTCAAATTCGGGTGGTGAACGTGATACCGATAAGACGGTTCAGTTGACA
GTCTATGTGGAAGTCTGGCATGTGTCTGCACATAATGTGGCTGGACTTGTATGAAGCTTTG
CCTCAGACTTCTCGGGTCACATTTTGCTTGATCATTCTGGGATTTTCATGGATAGGTAC
GAGCTAAAAGGGCGCAGTGGTCTCTTTTTTCAGAAAAATATAGCTTTTGAATTGGGAAAAG
TGTGACGAGATAACTTTTCTCGACAAAATATTTGGAGTTTGAGCGGACAACAGATCATGCG
AAAATTACTATGTCTAGTTTTATAGAAAAAATTTGCCTTTGTTTTAAAAGGGGCGAGGATG
GGACTTATTTTCCGAAAAACGTATATCAATATTCTATGATTCCTCATGTGCCTTTGTAT
TGCGGGGGTGATTTTTTGGATGTTTTACCCGTCAGAGACGGTATTAACACTTGTTTGAGA
ATGTTGCTTAATGTGGTACATTTCTTGGGTGACGAAGTAAGCGATGAGCTAACTGAGGAG
ATTGATTTCTGTCGTTTTGCAGTGTAATTTTTTATGTTTAAATGAATTAAGCGGGTGGTT
CGAAAAATGGTGTGGTTCGCTAATGCTGTGATTGATTATGCTGTAGAAAATAAAGATTTT
CTGTGTGAGGGAATCGAAGATGGACGAAGTCTAGGTATCTGTGTACGGGTCTTCATTCT
GTTTTCATGACTGTCCGGTTATCCTATGCACATCCAGATGCTCGTCGATTATATCGAATG
ATATGTGAACACATTTATTATACCTTGTGTTTGAACGAGTGTGGATTGTTGCATGAAAGGA
GCAGAGCCATGTAATTTATTTGACAGGTCAAATACGCACTGGGGATGTTATACTTTGAT
CATTTTGAGAATGTCGAATGCACGCTCCCGGAGGAATTGTGGACAACGCTAAGGAAAAGAC
GTGCTGATGCACGGAGTTCGCAATGTTTCACTGCGGTTACTGCCATGCAGAAAAGAA

TTCGATATCATCAACTCTTCTGAGTCGTTCTGGCCGATGGAAGACAATAAAATCCTAAGG
CGGTCAAAATATCAAAGTCGTAATTGGAAGACCGGTTTAAATGATGTCACTTCAGTTTAT
TCGTCGGAATTTAAATCGTTATATATCCCCGTGTATAACAATCTGTTGCTCAACAGATTT
AACAGCATCAGCAATATTTGAAGACTGTGGGATATAGGGTATTAACGCGGACACGAAC
CTATTTACGGACAAGGAGTTGGATGATTTAGCCGTTTTTCAAAGATGGTTTTAGTTATCAT
TTGAATGATTTGATTGAGATGTATAAGAGTGGATTGCCTTTTTTTAGATCAGGGGCAGGCA
AATGTGTTTTATTTCAACGATACTGTGAGTTTTGAGACTGTTGCTACCGTTGCTGTATAAA
ACCGGATTTAAAGTCGCTATGTATAAGGTTTTGTGTAATTTCTGAGATGTATAAACACCTT
GATTTATCAAATCCGTTACCGTTGATCGGGAATGCTCGGATGGTGTGGTTATGCACGTA
AGAACATTTTGTAG

U29

>2284_U29

ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAG
AGAAAATATATATCTATCGATGCGGGAACCTTAAATAATGTTGTAGAAAAAGAAAAGAAA
TTTCTGAAACAGTTTTTGTCCGGCAGGCAAAATTTACGTATAGCGGCAAGGGTTTTACT
CCTTGTGAATTACTGGCACCGGAGCTGGAATAATTTGGGAATGTTGATGTATCGTTTTGAG
ACTGATGTGGACAACCCATAAATTTCTGTTTGTGGGTCTGTTTTTCTTTGTTCTAACGCA
TTCAATGTGTCTACTTGTGTGAGAACCGCTTAACTGCCATGTATACAAATCAATGGTT
GATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTT
GGTGTGACGAGCTTAGTTTTCTTGCAGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGA
AATGTCTATGACGTTAACAAGAAAACATTTACGGGCTCACGGTATTAAGAGATTCTA
CTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTT
TACAAGGAGGTTTTAGCAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTTCTAG
AACGAAGATCTGGTAGTTGACATCTTGCAGCAGTATTTGAAAAGGAGCGGTTTTCTGTTT
TTGAACTATCTTATTAATAGCAATACCACGCTGAGTTACTTTGGAAGCGTTCAGCGTATT
GGTCGTTGCGCGACGGAGGACATTAATCGGGTTTTCTGCAGTACCGTGAATTACCCTG
TCAGTGATAAAGCTAGAGAATATATTCGTTGACCTTTCTGAGAAAAAGTGTGTTGTATAG

>5055_U29

ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAG
AGAAAATATATATCTATCGATGAGGGAACCTTAAATAATGTTGTAGAAAAAGAAAAGAAA
TTTCTGAAACAGTTTTTGTCCGGCAGGCAAAATTTACGTATAGCGGCAAGGGTTTTACT
CCTTGTGAATTACTGGCACCGGAGCTGGAATAATTTGGGAATGTTGATGTATCGTTTTGAG
ACTGATGTGGACAACCCATAAATTTCTGTTTGTGGGTCTGTTTTTCTTTGTTCTAACGCA
TTCAATGTGTCTACTTGTGTGAGAACCGCTTAACTGCCATGTATACAAATCAATGGTT
GATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTT
GGTGTGACGAGCTTAGTTTTCTTGCAGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGA
AATGTCTATGACGTTAACAAGAAAACATTTACGGGCTCACGGTATTAAGAGATTATA
CTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTT
TACAAGGAGGTTTTAGCAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTTCTAG
AACGAAGATCTGATAGTTGACATCTTGCAGCAGTATTTGAAAAGGAGCGGTTTTCTGTTT
TTGAACTATCTTATTAATAGCAATACCACGCTGAGTTACTTTGGAAGCGTTCAGCGTATT
GGTCGTTGCGCGACGGAGGACATTAATCGGGTTTTCTGCAGTACCGTGAATTACCCTG
TCAGTGATAAAGCTAGAGAACATATTCGTTGACCTTTCTGAGAAAAAGTGTGTTGTATAG

>5814_U29

ATGAACTCTAAATCGAGTGCACGAGCTGCCATTGTGGATACTGTGGAAGCTGTGAAGAAG
AGAAAATATATATCTATCGATGCGGGAACCTTAAATAATGTTGTAGAAAAAGAAAAGAAA
TTTCTGAAACAGTTTTTGTCCGGCAGGCAAAATTTACGTATAGCGGCAAGGGTTTTACT
CCTTGTGAATTACTGGCACCGGAGCTGGAATAATTTGGGAATGTTGATGTATCGTTTTGAG
ACTGATGTGGACAACCCATAAATTTCTGTTTGTGGGTCTGTTTTTCTTTGTTCTAACGCA
TTCAATGTGTCTACTTGTGTGAGAACCGCTTAACTGCCATGTATACAAATCAATGGTT
GATAACGTGTTATCGATGATCAATACGTGTAGGTACCTGGAAGACAAAGTGTCTTTGTTT
GGTGTGACGAGCTTAGTTTTCTTGCAGCAGCAGTTGTCTGTTGTCATGCGTCATGCAAGGA
AATGTCTATGACGTTAACAAGAAAACATTTACGGGCTCACGGTATTAAGAGATTCTA

CTAGAACCAGACTGGGAACCGCGGCAGCATTCCACCCAGTATGTTTACGTGGTTCATGTT
TACAAGGAGGTTTTAGCAAAGCTCCAGTATGGGATCTACGTAGTTCTGACTAGTTTTTCAG
AACGAAGATCTGGTAGTTGACATCTTGCACAGTATTTTCGAAAAGGAGCGGTTTTCTGTTT
TTGAACTATCTTATTAATAGCAATACCACGCTGAGTTACTTTTGGAAAGCGTTCAGCGTATT
GGTCGTTGCGCGACGGAGGACATTAATTCGGGTTTTCTGCAGTACCGTGGAAATTACCCTG
TCAGTGATAAAGCTAGAGAATATATTCGTTGACCTTTCTGAGAAAAAAGTGTGTTGTATAG

U30

>2284_U30

ATGACGCATGACAACAGACAACCTGCTGCCGCAAGAACTAAGCTCGTCACACCAAACAAA
GACACTTTGTCTTCCAGGTACCTACACGTATTGATCATCGATAACACGTTATCAACCATT
GAGTTTGTATACATGGCAGTTAAAGCGGTTCTCACACAAGTAGACACATTGAATGCGTTA
GAACAAAGGAAAAACAGACCCACAAACAGAATTTTAGGGTTGTCCACATCAGTCTCAAAA
CGATACATCAACATTCCCAAATTTCCAGCTCCGGTGCCAGTAATTCACAAGGAGTGAAA
ACCCTTGCCGTATACGTAAATTTTGCCTGCCGGACAAAACTGTTTCAGAAATTTTCTT
TCTTTTTCTACAACATTATTTAAGGTTCCCGCATCGATAGATATATATTTTTCTCTTCTT
ACAGCTTCCACAGTATCCACAATGGCAGCTCGTGCACCTCGATTTAGAGTTCATCCTTAGT
TCACTCAAAAACAGCAGCTCTCCAGAATCATTACTAGCCGCTACGGCAAAAATCGAACTT
TTATCCCTGGCAGCCGATAGCGTAACACACAACAGAGTAATAGCTTTTATTAATCGACTA
ACCCCGAAAAATATCACTTTGACCTCATCAGAGAGTATGCGGTATTCTACTTTTTAAAC
TCTACGACCTTAACATCAGAAAATAAATTTGCTCTCCGCAGAACTTTTATACGAAGAACTC
AGTCAAATTCCTCTTCGAGTTCGTGCGGCACAGAACTTGAAAATTAATAATGCAGAG
GTCCTTTACGTGTTGATCGAATGCTCAATTCATTAATAATGCTAAAAAATGAGTTATCG
TCCCAATCGGCAAGCTTCGAGCACAACCCGACGCCAATGACCCAGGTACCGATAAAAACG
ATAAAATACCTCAAGATTCACCTCGATCATAAAAAGATAAACAGCTTTACGCGAGAAAAT
TCACTTCTTGCCAACTGCCGCGCCGTGATAGAGCTGATAGACGATCTCTATCGGAACTT
TACTCATGGTTTTATGCAGGTTCTCACTTTTCGAGGACATTCAGTTTTCCAGGCGACACATTT
CTAGACAGGCTGTTGAAAATGGACTACTGTTTTACATATTACCCGTCTTCAAATCGCCAT
CTAATAGATCTGTTGAAAAGACGTTGGACAATCAAACGTTTCACGGACCTCGATAAAATTT
TTTGACACCTCAGGAAACAGCCAGAATCTTTTATCAAAAAACGTTTAGCTTGAAAATA
TTTTCAAAAAATCTCACAGCCCAGGACAATGGATTATATATCTATCCACTACTTAAAACC
GACCTGTCCATCCTCGACTTCCTCGGGACCGAAAATATTCTCTTCCATCGAGGCCTCATC
TATCACATACTTCATCAAAAAACAATCCCTCAAGAACGGGAGAATGATCTCAATAAAAATC
AATCAATTTTTTCGCTACTGTCTATCCAGCAGTTATCGAAACGAAAAGCTCATGCTTACCC
GCATCGTTATCCCAACTGTTAGATACGATTTTTTCACTTCAACAGAATCGGTTTTAAACATG
GAAACGTGCAGAACCCTACATCGAGATCCTCTCAAACCACATGGCAACGCCAGACACCCAA
CCCATAATCAATACATTCACCATTAACCTTAATCCATATTGTTTTTACC GCCCACGTGTTT
TTTATCTGTATGGAAAATTTAGCCCAACATTCATATTCTATAATAGGAAAAAACTCATA
CTAGAACAACAGAGAGCCATTTTGATCATCGAACGAAATGATTATTCCACACTCTGGAAG
CAAATATCTGATCACATCGACTGTCTCTTCAACGTCTCTCTCACCGAATCCTTTTTTAAA
GAATATACCAAAGGAGGAAACGAAGATCAAAAAACGTTTTCTCTACAAAAACCTATTTCGAA
AAATGGGGTAACGTATTTTTCCCTTTACGTAATCCACCTCAAACAATTCTACG
GTACACCATATTACAACCCCTGAGCTAAGGGATATCTGCAAAGAGGTCTATCAGAGCGAT
TCCCGGACGCATATGAATCTTTACTTCCCTACAGCACTCATCCATCTTTCAAACATTA
TTCGTTAAAAATCTACGTTATACCTATGGTTTTACACATTACCAATCTTACCTTTGATAAA
TTACAATCTGATTGCAGACTCTTAACCTAATCCACGCCTGCAAACCTTACTTCCCCTCG
CAACATCTACTACTACATGCGGTGGCTATATGCCTTCTCCATAAACGTGGACCAT
ATAGATCTAGGCATTTTACCCTAATAAAGTCCGTAATCTTTAAAATCGCTGACCACATA
AACGTTATGACGCATACGATCTACTCCCGGAGACAAACCTTCTCGTAAGCATTCTACTA
AACGCGTATACCAACTACCTCAAATAACGTAACCTTGGATCAAACAAACCATAACA
GCAAACCTTCTCCCTGATACAGACGTACATAACTTTTACGAAACAATGCGCCAGCATCTTG
GCGACCAAATGCAACATAAACCTTAGACAATCTATTTATCTACATAACTATCGGAACAGAC
AAAATTGTAACCCTTCGTTTTGCTCCTTTCATCGCTACATGCAGAAATTTAGTACGTCAA
CACGAAGAATTGAAAAATCACTACAAACGATCCAAATTTCCGAAACCACACTCACGGGG
ATGCTACGAAACATAATCAGGAGCGTCTCCTCCTCAAAGAGCTGCTAACCAACGAAGCA
CTACAAAAATTCATCGACACTGTCCAACGCATCTCCAACACGTAAACGAAACATAACCAG

TCTATTTCCGTGAACCTCGAAAAATGTAAAACCTCAAACGACATCCTAATCGAATCCCTG
AAGAAGATCATCTACATAGTTGATGTACTCAGCTCCAATGCAATCCTAAACACGTGCTTA
GCTTCCAGATGTCTGGAAGCCGCCAACCTCGCGGTTTTCGAACGAATCTTTTACAATACTC
GAAATTA AAAAGGACGCAGTTGCCGTTTTTCAAGCCTTTTATAACACA ACTATTTGAAAAGC
ATGAAACCCACCAGGTTTACATAAGAAATTTGATGTCTACCCAAAACTAACCACCGAC
CACATTCATTTCTCGATACATTTGACGATAGGTACAACCTGGTCAGACACGTGGAACGA
CAATTAACCTGGTACGCTGCACATGCTGAGGCAGCGCAGCAAGATCTTATTACTCCTCTT
AAATTTTAA

>5055_U30

ATGACGCATGACAACAGACAACCTGCTGCCGCAAGAACTAAGCTCGTCACACCAAACAAA
GACACTTTGTCTTCCAGGTACCTACACGTATTGATCATCGATAACACGTTATCAACCATT
GAGTTTGTATACATGGCAGTTAAAGCGGTTCTCACACAAGTAGACACATTGAATGCGTTA
GAACAAAGGAAAAACAGACCCACAAACAGAATTTTAGGGTTGTCCACATCAGTCTCAAAA
CGATACATCAACATTTCCAAATTTTCCAGCTCCGGTGCCAGTAATTCACAAGGAGTGA
ACCTTGCCGCTATACGTAATTTCTGCCTGCCGGACAAAACTGTTTCAGAAATTTCTTT
TCTTTTTCTACAACATTTAAGGTTCCCTCATCGATAGATATATATTTTTCTCTTCTTC
ACAGCTTCCACAGTATCCACAATGGCAGCTCGTGCACCTCGATTTAGAGTTTCATCCTTAGT
TCACTCAAAAACAGCAGCTCTCCAGAATCATTACTAGCCGCTACGGCAAAAATCGAAGTT
TTATCCCTGGCAGCCGATAGCGTAACACACAACAGAGTAATAGCTTTTATTAATCGACTA
ACCCGAAAAATATCACTTTGACCTCATCAGACAGTATGCGGTATTCTACTTTTTAAAC
TCTACGACCTTAACATCAGAAAATAAATTTGCTCTCCGCAGAACTTTTATACGAAGAACTC
AGCCAAATTTCCCTCTTCGAGTTCTGTCGGGCACAGA ACTTGAAA ACTTAAATAATGCAGAG
GTCTTTTACGTGTTTCGATCGAATACTCAATTTCCATTA AAAATGCTAAAAAATGAGTTATCG
TCCCAATCGGCAAGCTTCGAGCACAACCCGCCCAATGACCCAGGTACAGATAAAAACG
ATAAAATACCTCAAGATTCAGTCGATCATAAAAAGATAAACAGCTTTACACACGAAAAT
TCACTTCTTGCCAACTGCCGCGCCGTCGTAGAGCTGATAGACGATCTATATCGGAACTT
TACTCATGGTTTCTGCACGTTCTCACTTTTCAAGACATTCAGTTTCCAGGCGACACATTT
CTAGACAGGCTGTTGAAAATGGACTACTGTTTTACATATTACCCTTCTTCAAATCGCCAT
CTAATAGATCTGTTGAAAAGACGTTGGACAATCAAACGTTTACGAACCTCGATAAAATTT
TTTGACACCTCAGGAAACAGCCAGA ACTACTTTATCAAAAAACGTTTAGCTTGAAAATA
TTTTCAAAAAATCTCACAGCCAGGACAATGGATTATATATCTATCCACTACTTAAACC
GACCTGTCCATCTTCGACTTCTCGGAACCGAAAATATTCTCTTCCATCGAGGCCTCATC
TATCACATACTTCATCAAAAAACAATCCCTCAAGAACGGGAGAATGATCTCAATAAAAATC
AATCAATTTTTTCGCTACTGTCTATCCAGCAGGTTATCGAAACGAAAAGCTCATGCTTACCC
GCATCGTTATCCCAACTGCTAGATACAATTTTTTCACTTCAACAGAATCGGTTTAAACATG
GAAACGTGCAGAACCCTACATCGAGATCCTCTCAAACCACATGGCAACGCCAGACACCCAA
CCCATAATCAATACATTCACCATTAACCTAACCATATTGTTTTTACC GCGCACGTGTTT
TTTATCTGTATGGAAAATTTTAGCCCAACATTCATATTCTATAATAGGAAAAAACTCATA
CTAGAACAACAGAGAGCCATTTTGATCATCGAACGAAATGATTATTCCACACTCTGGAAG
CAAATATCTGATCACATCGAATGTCTCTTTAACGTCTCTCTCTCCGAATCCTTTTTTAAA
GAATATACCAAAGGAGGAAACGAAGATCAAAAAACGTTTCTATACAAAAACCTATTTCGAA
AAATGGGGTAACGTATTTTTTCCCTTTTACGTA CTCTCGGTATCCACCTCAAACA ACTCTACG
GCACACCATATTACAACCTTTGAGCTAAGAGACATCTGCAAAGAGGTCTATCAGAGCGAT
TCCCGGACGCATATGAATCTTTACTTCCCTACAGCACTCATCCATCTTTCAAACATTA
TTCGTTAAAATCTACGTTATACCTATGGTTTTACACATTACCAATCTTACCTTTGATAAA
TTACAATCTGATTGCAGACTCTTAACTCTAATCCACGCCTGCAA ACTCTTACTCCCCTCG
CAACATCTACTACTACACTACATGGCGTGGCTATATGCCTTCTCCATAAACGTGGACCAT
ATAGATCTAGGCATTTTACCCTAATAAAGTCCGTAATCTTTAAAATCGCTGACCACATA
AACGTGATGACGCATACGATCTACTCCCGGAGACAAACCTTCTCGTAAGCATTCTACTA
AACGCGTATACCAACTACCTCCAAAAATACGTA AACCTTTGGATCAAACAAACCATAACG
GCAAACCTTCTCCCTGATACAGACGTACATAACTTTTACGAAACAATGCGCCAGCATCTTG
GCGACCAAATGCAACATAAACTTAGACAATCTATTTATCTACATAACTATCGGAACAGAC
AAAATTTGTAACCACTTCGTTTTTGTCTCTTCTCATCGCTACATGCAGGAATTTAGTACGTCAA
CACGAAGAAATCGAAAAATCACTACAAACGATCCAAATTTCCGAAACCACACTCACGGGG
ATGCTACGAAACATAATCACGAGCGTCTCTCTCCAAAGAGCTGCTAACCAACGAAGCA
CTACAAAAATTCATCGACACTGTCCAACGCATCTCCAAACAGTA AACGAAACATAACCAG
TCTATTTCCGTGAACCTCGAAAAATGTAAAACCTCAAACGACATCCTAATCGAATCCCTG

AAGAAGATCATCTACATAGTTGATGTACTCAGCTCCAATGCAATCCTAAACACGTCGTTA
GCTTCCAGATGTCTGGAAGCCGCCAACCTCGCGGTTTTCGAACAAATCTTTTACAATACTC
GAAATTAAGGAGCAGGTTGCCGTTTTCAAGCTTTTATAACACAACCTATTTGAAAGC
ATGAAACCCACCAGGTTTACATAAGAAATTTGATGTCTACCCAAAACTAACACCGAC
CACATTCATTTCTCGATACATTTGACGATAGGTACAACCTGGTCAGACACGTCGAACGA
CAATTAACCTGGTACGCTGCACATGCTGAGGCAGCGCAGCAAGATCTTATTACTCCTCTT
AAATTTTAA

>5814_U30

ATGACGCATGACAACAGACAACCTGCTGCCGCAAGAACTAAGCTCGTCACACCAAACAAA
GACACTTTGTCTTCCAGGTACCTACACGTATTGATCATCGATAACACGTTATCAACCATT
GAGTTTGTATACATGGCAGTTAAAGCGGTTCTCACACAAGTAGACACATTGAATGCGTTA
GAACAAAGGAAAAACAGACCCACAAACAGAATTTTAGGGTTGTCCACATCAGTCTCAAAA
CGATACATCAACATTTCCAAATTTTCCAGCTCCGGTGCCAGTAATTCACAAGGAGTGAAA
ACCCTTGCCGCTATACGTAATTTTGCCGCGGACAAAACTGTTTCAGAAATTTTCTT
TCTTTTTCTACAACATTTAAGGTTCCCGCATCGATAGATATATATTTTTCTTCTTCTC
ACAGCTTCCACAGTATCCACAATGGCAGCTCGTGCACCTCGATTTAGAGTTTCATCCTTAGT
TCACTCAAAAACAGCAGCTTCCAGAATCATTACTAGCCGCTACGGCAAAAATCGAACTT
TTATCCCTGGCAGCCGATAGCGTAACACACAACAGAGTAATAGCTTTTATTAATCGACTA
ACCCGAAAAATATCACTTTGACCTCATCAGAGAGTATGCGGTATTCTACTTTTTAAAC
TCTACGACCTTAACATCAGAAAATAAATTGCTCTCCGCAGAACTTTTATACGAAGAACTC
AGTCAAATTCCTCTTCGAGTTTCGTCGGGCACAGAACTTGAAAACCTAAATAATGCAGAG
GTCCTTTACGTGTTTCGATCGAATGCTCAATTCATTAATAATGCTAAAAATGAGTTATCG
TCCCAATCGGCAAGCTTCGAGCACAACCCGACGCAATGACCCAGGTACCGATAAAAACG
ATAAAATACCTCAAGATTCACCTCGATCATAAAAAGATAAACAGCTTTACGCGAGAAAAT
TCACTTCTTGCCAACTGCCGCGCCGTCGTAGAGCTGATAGACGATCTCTATCGGAACTT
TACTCATGGTTTATGCAGTTCTCACTTTTCGAGGACATTCAGTTTCCAGGCGACACATTT
CTAGACAGGCTGTTGAAAATGGACTACTGTTTTACATATTACCCGTCTTCAAATCGCCAT
CTAATAGATCTGTTGAAAAGACGTTGGACAATCAAACGTTTCACGGACCTCGATAAAATTT
TTTGACACCTCAGGAAACAGCCAGAATACTTTATCAAAAACGTTTAGCTTGAAAATA
TTTTCAAAAATCTCACAGCCCAGGACAATGGATTATATATCTATCCACTACTTAAACC
GACCTGTCCATCCTCGACTTCCTCGGGACCGAAAATATTCTTTCATCGAGGCCTCATC
TATCACATACTTCATCAAAAACAATCCCTCAAGAACGGGAGAATGATCTCAATAAAATC
AATCAATTTTTTCGCTACTGTCTATCCAGCAGGTTATCGAAACGAAAAGCTCATGCTTACCC
GCATCGTTATCCCAACTGTTAGATACGATTTTTCACTTCAACAGAATCGGTTTAAACATG
GAAACGTGCAGAACCACATCGAGATCCTCTCAAACCACATGGCAACGCCAGACACCCAA
CCCATAATCAATACATTCACCATTAACCTAATCCATATTGTTTTACCCGCCACGTGTTT
TTTATCTGTATGGAAAATTTTCAGCCCAACATTCATATTCTATAATAGGAAAAAACTCATA
CTAGAACAACAGAGAGCCATTTTGATCATCGAACGAAATGATTATTCCACACTCTGGAAG
CAAATATCTGATCACATCGACTGTCTCTTCAACGTCTCTCTCACCGAATCCTTTTTTAAA
GAATATACCAAAGGAGGAAACGAAGATCAAAAACAGTTTCTCTACAAAACCTATTTCGAA
AAATGGGGTAACGATTTTTTCCCTTTCACGTACTCGGTATCCACCTCAAACAATTCTACG
GTACACCATATTACAACCCCTTGAGCTAAGGGATATCTGCAAAGAGGTCTATCAGAGCGAT
TCCCGGACGCATATGAATCTTTACTTCCCTACAGCACTCATCCATCTTTCAAACATTA
TTCGTTAAAATCTACGTTATACCTATGGTTTTCACACATTACCAATCTTACCTTTGATAAA
TTACAATCTGATTGCAGACTCTTAACTCTAATCCACGCCTGCAAACCTTACTCCCTCG
CAACATCTACTACTACACTACATGGCGTGGCTATATGCCTTCTCCATAAACGTGGACCAT
ATAGATCTAGGCATTTTACCCTAATAAAGTCCGTAATCTTTAAAATCGCTGACCACATA
AACGTTATGACGCATACGATCTACTCCCGGAGACAAACCTTCTCGTAAGCATTCTACTA
AACGCGTATACCAACTACCTCAAAAATACGTAACCTTGGATCAAACAAACATAACA
GCAAACCTTCTCCCTGATACAGACGTACATAACTTTTACGAAACAATGCGCCAGCATCTTG
GCGACCAAATGCAACATAAACTTAGACAATCTATTTATCTACATAACTATCGGAACAGAC
AAAATTGTAACCCTTCGTTTTGCTCCTTTCATCGCTACATGCAGAAATTTAGTACGTCAA
CACGAAGAAATCGAAAAATCACTACAAACGATCCAAATTTCCGAAACCACACTCACGGGG
ATGCTACGAAACATAATCACGAGCGTCTCCTCCTCAAAGAGCTGCTAACCAACGAAGCA
CTACAAAATTCATCGACACTGTCCAACGCATCTCCAAACACGTAACGAAACATAACCAG
TCTATTTCCGTGAACCTCGAAAAATGTAAAACCTCAAACGACATCTAATCGAATCCCTG
AAGAAGATCATCTACATAGTTGATGTACTCAGCTCCAATGCAATCCTAAACACGTCGTTA

GCTTCCAGATGTCTGGAAGCCGCCAACCTCGCGGTTTTCGAACGAATCTTTTACAATACTC
GAAATTA AAAAAGGACGCAGTTGCCGTTTTCAAGCCTTTTATAACACA ACTATTTGAAAAGC
ATGAAACCCACCACGAGTTTACATAAGAAATTTGATGTCTACCCAAAACTAACACCGAC
CACATTCATTTCTCGATACATTTGACGATAGGTACAACCTGGTCAGACACGTGGAACGA
CAATTA AACTGGTACGCTGCACATGCTGAGGCAGCGCAGCAAGATCTTATTACTCCTCTT
AAATTTTAA

U31

>2284_U31

ATGAAAATCATTACCAGTAGCACAAAACAAAATGACTCTAAATACGGACCGAGAGCCGGA
AAACAATGCATGTGGAACAGCTTTTCATTTCTCCATACCGTTTACTTAAACGGAATTAAC
AATTTCTCTCAATGCGGGTACTATAGACGCCATAATGGAAGAAGGCTATCACTTAGACACG
GCCGGCACCTTAGCCCTAATGTTAAATAACTCAGATTCGCAAGACTACAGACTCCTCACC
GAGATCCCCAAGAGGATCCATTTCTAGATACGGTGTACACAGCATGAACTCTCACGGCCA
TTTAAACGGAACCCCTAGACACACAAAAAATTGACAATGAGGTATATCTCGGACTCATAGAT
TTCATACTGTATGGCAAAGCAAGAATGGTCTACTTTTTGCAGTCATCACGATCGGCGTG
CTATCGCGAGCACTTTTTTTTTCTTAAACAATACTCTCTACTTATTTCGACTCTCACCCCTACT
GAACGAGAAGCCACAGCAGCCATCTACATCTGTCAAAAACATTGAAGAAGTTTATGAACTG
CTCACACCCACGGCACCGAAGGCTTCTACTACGACGCCAGCTTCATTTTTTTTCATAGAA
ACCTCCAATTTGTGCTCTCTTAGTCACGACGCCGAGCTTCTCATCTAAAAACTTATAAAA
GACCCCGACATAGCGATTGCGCTAGATAAAATTTTTCTTCCACGGAAATCCATGAGATAAAA
AAAACAGATGATATCGGATCGCAACAAGACCTCGTTGCGGATAAAAACGACAGATCTAGAA
CACGCGCCCCATAAAAAGAAAGAAAAACTCTCATAGTCTAGA ACTAGAACTAAATGACAAA
AAGAAAAAGATACCGCATCTTGGACATATTACGCAACAGAAGTTGACCTCATTCCGAGT
TTTTATGAACTACGATCACAATTTCAATCTTTATTTTACGATCTCAAATCTTTTTCTATA
ATGAAACCTCATTTCAACTGGACCATATACTTACAAGATTCTCCTATTAATCCTAATCAG
CCCTTCGCGACACCCTTTTCTGGAACAGAGTTTTCCACTTATTGTGTCAAATTTATCGAC
GTGTTTGTAGGCGTCGGATCCACAAATGACGATTCTAGCAAACAACAACAACAACTATT
TTCATAAATTTATTTATTGCCTTTTCAAGGACTTTTTCCGAAGTATTCAACGAGGCGTTGGCA
GCTTGTCAAGAAAACAATCTGGACATCATCTTCATTTACAATAATTACCTGTGCAAAAAC
ACGACCTTTTGAACACTCGAGAGAATCTTACTAAGCAAATTTCTGGCAATAGTTGATAAT
GATCACAAAAACATTATGAATGGGTCAAATCGTGGACCACACAAATGTTCCAAGGAATG
CCAAAAAACTAGATGACATAGAGAATTATCTAAAAGCCTACGTGGACCATAACCCCTGTG
AAACACTTTTCATGAATTCATTTGTCTGAATAAGGCAGAAAAATACAAAGTAGCTGTTCTG
CTAAACGAAAAACGAAAAGAGATTCAAGAGGCTATTGAACGAGAAAAGAACAGTTTTCGCA
CAACTCTCTAATTTTATTGACAAACTTGGCGAAAACCTCCCGCACTTCTATAGAATCAGAA
AACGTGCACAAAGTTCACACGAACGACATCACTGAAGGCATCGTTCCGCGTTTTCATAACC
GAATCGATTGAACTACCAAACATCTCCACGCTAAAACAACACCCAAACAAATATCCCTAGAT
AAACAACCTCAATGAAAACTCACCAACACCATTACACGCTCACAAAACAAATTCACCAAAA
ATCGTACAAGATAACTATAACAATATCGCTGCCGGTTTTCATGCCGTTACCGA ACTGAAT
TGCCATTTTGCCTACCTAGTAAACCTTTATTTTAAACATTGAAGTTTTTAAAACACAGTGGC
CTAAACATTAACACTGTGTTACTTCAAGAGGTGAAAAACTGTATGACAACACGCAATTT
CTACGCTTCGGAACTTCTCACTTCAATATAAACAACCTTTCAAACCTTTACCCTATCTATC
AGAAAAATGTTTCGTAGATTTCTATAACAGTCAGAAAACCTCGGACAGAGCTTCCGAAATC
CTAGCCGCAATCGAATCCATCCTAGCGGATCCCAGCAAAAAACAAGACAGTCGTTAATATA
GAAATGATCAAATCGCAACTTGAAGA ACTGGGGAAAAATGGAAATCTCCACCACCGAAAAAC
AAACAGACAGCTGCAATCACAAAACAAATCCTCGGGGACCAAGAGTTAACCCCATTTTAC
GATTTCTGCACCACCTCTCTGCTTATAACCTACCTAATACTACAACCGTAAAAAATTTA
CACCTCCATTTTCATCTTAGAACAACGACCAGATATCGCTATGACTCTACACGATAAAAATA
CAATCCATACTTGACATATATGTAGACGACATGCTCAATGACATAACGGTCCCTGAACAG
ACGTTTTCAACCGTTCTATTTCTTGTAGATCTCTTCCCAAACAGCACCGAAAAAACAGCA
CTGTTGCAATCCGTTCTCACGCTACGGCAATTAGCAAAGAAATGCGCTAACTTAAAAACT
CTAGATGAATTCGACGATCTGGCCCAATTCATCACGACGAATAGCGAACA ACTTCAAAAAC
ATGATGAGACAGCATTTTGGAAAAAATAACCAACGCTGATGGGACATATCAAATTTCTC
TACTCTCAAAAAATAATCACCCTGAAGAGAAAAACTGGATACAAAGAGCAAAAAACGGTC
GTCATCACATCTCCGAAGAACTGACAGCCTTCTTAGCCACTGCACCCACTAAACACGCC

TTACAAACTTGCAAACCTGACCTCGATAAAGCACTACAACGACATATGGAAGAGCAAATG
AAACAAACTGCCGAGAACGACAAAAACACATCCTCACAATCAGAAGCACCCCTCGAAAAA
AGACTGACCGACATCTTACTTATTCTAAAGGATGGTCAATTTTTCGTCTCTGGAAACAATG
CATTTGAATCTTTTAGAAACCTTTCTGAAACAACCTGCAAGATAACAATGTCATCATTAC
TTTACCCATGCACACTGCCAGTGTAAAAAGACATAGAAACGACCATCAGCAAGATCATC
TCCGACGTTATTGAAAAGATATTAATCAAACTCCTCTAAACCCCGAACAGATGTCCAAG
GAAGAACAAAAATATACCCCTTTACTCAGTTTTCTTATCCAAGTTCAAAAAACAACATTC
TGCACAGAAGATGTA AAAACAGAAATCGACCAAATGCAAAAAGAGTATCACGTTTCTCAAG
AAAAATTGCTACCTCCGCCAATAAACATACTAGACTCAGCCACTCCATCTACGGACAAGAA
TTAAACCTCTACGAAGAGCGCATTACAGAGCTGAGAAAAGAGACGAACAAAATGAAAGAA
CAGCTCTCCAAAGAATACGCATTAGCTGAAAAAAAATTTCTTCTCTCGTCACAAGATGCA
AAAACAGCAAAAATTTATCTAGTTCTAAATACGCACACGCTGAAAGAAATAAAAAACACT
CAATTCAGAGAAACAGCATTGCTAAAGCTCTCACGGTGGAAAGTCAACAATAAAGAAAAAT
CAACTCCAGGAATTGTTAAACCCTTTAACGCACACCTAAAGGCGAAAAATGGATCAGAAC
CATATAACAAAACTTTCATTCGATACGAAATGGACAGCCTTCGTTTTCGGATTCCAGGCTC
TACATCCCAGACTTCATTAACATCAAACCTGCAAGACTTTATTTCTGATCCCTTTAAAGTT
ATAAGTCAGCTAATGAACAAAGCAACCAACGAGATGCCTTACATTCAAGCAGAGATAACT
CTCAAATGGTTAAACAAAATTAATACACGACATCAACAAATTTCTGCCTGTCAGCTATAAGC
GAATTCGGTAAAGAAGCAATCCCATTCATTAACGCAGCCCTACGAGATTTGGAGTACCAG
ATCAACACCAAATACGTAGAAATGAAAAACAAAGTCATTTGCAACGAAACGGTGCAAAAAC
ACGAAAAACATTCAAAACCTGACTAAGCTCTTGAAGCAGTTAGATCCTAAACGCGTTGCC
GGTGGACAGGAACAGTATCAAACCTCATGAACAAGATTTCTCAGTTCCGAAACTTCTATG
CAACAGACCTACGAAAAAGAACAACCTCAAAAAAGAATACTTTGAGATCCTCAATAATGTC
GCCAGTTCAAGTTAGCATTCAACTTTTCCCTCAGCAACTACAAAATGTGGAACGGCTAATA
GAGAAGTTCAAATCGCTGCCAAAAGCCAGCCATTTGAAAAATTTCCACAGGAAAATGAT
CTACTTTCTGACTCACTCAACACAGAAAATTAATTAACGGTCTACGCGCACTCCTCAAC
TTCATAACGGCAGCGCAAAACCTACATCCAGAACACACTTCTGAAACAGTGGGCGGTCTTT
CAGCAACAAAATTTTCATCCCGATTGATTACTCAGTAGCGAATGTCAAACCTATTTCCGGAT
TTATATGCCCGCCTCAGGATTGAACGGGAACGGCAAGTGTCTACCAAGTAAACTCCGTC
TTTGAACTCACCTAATCGTTGACGATACAGGAGTTCCGCTGCAATTCATAACATCTTT
AACAAACGCATAGTAAAGTTTTTCTCCCTGAACTACAAGCAGATTCATGTACCAGAAGAC
ACCCCCCGCTTGGTATCTAGCCAATATAAACTCCTGTCCGTATGTAAATCTTTCATAATG
ATACTACAGCAGTTCTGGGAGAATATAATCACCTGGACCTAGGCCCTTACCTCAGAGAC
GGTACACAAAATTTTAAAAGAGAACCTCATACCGATCGTGAACCTTGAATTAATTTATATAC
TGCATAACACAAGCCTGGACAGCCTCTGAAGACAGTACTGTTTCCACAGCGTTTCAATTA
CCCATCAAGCAATTCACACTTTTGTACTATGCTCACATCCTGAATACCTCTACGGCTGC
CTCAGTCATTCACAGATCTGGTTATAAATTCATTAGCCAAAAGCATAGACAAAAACAGT
CTCTATAACACATTCGTGCTTAGCCACAACCCCTCCGAGAAGCCCATGCATTTAATGAGA
AATATATGTATCGACACCCAGCTATGGCAACCCGCCAAACTTATGAAAGATACCTTCCAA
CAAACATTTTTTACACAACCTGTGCCCAAAAACGAAAAGTTTTTTTATCTATCTCACTGCC
TTTTTAATTTTACCTACAAATTTATGAACTATATCTGGATTCAATACAAACCGGCCATT
TTTACCCAAAGATCTTATCAAAACCTAATTAAGGACTTATGCTCCGAATACGTACACCAA
AATAAGATCACACTGCTCCTCGGTAACCTCCACGAACCCAGACACAATAAAATCCGGAGAA
AGGATTACCTCAAAAATCACCGTACACAAAGCACAAAACACACCGCACTCACTCGCCTG
CAGGCTCAGGAATATGCTTTCGACTACATTTCTTACTCCTTCTTGACAGGGTATGAAATG
ACATTCGCCATGTACATCGATACCATCGAAAAACATACTGCTTTGCATGAGACACCTG
GAAAACGTACTCCACGATAAAGACTTCCAATCTGTCTCAGAGCACGCACATTCGACATC
GACTACATCTGAAACAGTCATGGACTAAAAACATAGTCAACATTCACCTTTTTTCAGTA
CAGCTCGATAAAAATCGTCTCCTACCTCAACCACACAAACAGAGCGACACCCAATATTCCT
TTGATATTTAATTTATGACAACGAAGTTGTCAACGTTTATCTACCGCCGATGTCCACC
AACCCAAAAAAGTCGCCTTTTTACATCAAAAATCCTTTCCACTTTCCCGTGAAGAATAC
GAAGCGACAAATTTAATTTCCCTTCCATCTCTATCCAAAAACAACAGATATTTTAAACCAA
CTTCCGCCAGCAATACAGTCTACTCGCCCGGGCAAAACAACCTCTGAAAGTTTAAACC
ACTAAAAACCTCTCCGAACCAAAGTTTAAAAAGCCTGCCGTTACTGGACTTATGCCCAAG
AGCCAATCGATCATCTATCCACTGATACCAACGTACCGGAAACAGTCCAGACGTAAAA
GCTAACACGGCCAGCGCAGCGATAAAAAGATGTCACTCTAGCTCGAGAAAAAATCAACGAG
TTCTCCGAATCGATAAATACAACATTTTTCAAAATTAATCTATGTATCTGTAA

>5055_U31

ATGAAAATCATTACCAGTAGCACAAAACAAAATGACTCTAAATACGGACCGAGAGCCGGA
AAACAATGCATGTGCAACAGCTTTTTCATTTCTCCATACCGTTTACTTAAACGGAATTAAC
AATTCTCTCAATGCGGGTACTATAGACGCCATAATGGAAGAAGGCTATCACTTAGACACG
GCCGGCACCTTAGCCCTAATGTTAAATAACTCAGATTTCGCAAGACTACAGACTCCCCACC
GAGATACCCAAGAGGATCCATTTCTAGATACGGTGTACACACAGCATGAACTCTCACGGCCA
TTTAAACGGAACCCCTAGACACACAAAAAATTGACAATGAGGTATATCTCGGACTCATAGAT
TTCATACTGTATGGCAAAGCAAGAATGGTCTACTTTTTGCAGTCATCACGATCGGCGTG
CTATCGCGAGCCTTTTTTTTTCTTAAACAATACTCTCTACTTATTTCGACTCTCACCCCTACT
GAACGAGAAGCCACAGCAGCCATCTACATCTGTCAAAAACATTGAAGAAGTTTATGAACTG
CTCACCACCCACGGCACCGAAGGCTTCTACTACGACGCCAGCCTCATTTTTTTTCATAGAA
ACCTCCAATTTGTGCTCTCTAGTCACGACGCCGAGCTTCTCATCCTAAAAACTTATAAA
GACCCCGACATAGCGATTGCGCTAGATAAAATTTTTCTTCCACGGAAATACATGAGATAAAA
AAAACAGATGATATCGGATCGCAACAAGACCTCGTTGCGGATAAAAACGACAGATCTAGAA
CACGCGCCCATAAAAGAAAAGAAAACTCTCATAGTCTAGAACTAGAACTAAATGACAAA
AAGAAAAAAGATACCGCATCCTTGACATATTACGCAACAGAAGTTGACCTCATTCGGAGT
TTTTATGAACTACGATCACAATTTCAATCTTTATTTTACGATCTCAAATCTTTTTCTATA
ATGAAATCTAATTTCAACTGGACCATATACTTACAAGATTCTCCATTAATCCTAATCAG
CCTTTCGCAACACCCTTTCTCTGGAACAGAGTTTTTCCACTTATTGTGTCAAATTTATCGAC
GTGTTTGTAGGCGTCGGATCCACAAATGACGATTCTAGCAAACAACAACAACAACTATT
TTCATAAAATTTATTTATTGCCTTTCAAGGACTTTTTCCGAAGTATTCAACGAGGCGTTGGCA
GCTTGTCAAGAAAACAATCTGGACATCATCTTCATTTACAATAATTACCTGTGCAAAAACC
ACGACCTTTCGAACACTCGAGAGAATCTTACTAAGTAAATTTCTGGCAATAGTTGATAAT
GATCACAAAAACATTTATGAATGGGTCAAATCGTGGACCACGCAAATGTTCCAAGGAATG
CCAAAAAACCTAGATGACATAGAAAATTATCTAAAAGCCTACGTGGACCATAACCCCTGTG
AAACACTTTCATGAATTCATTTGTCTGAATAAGGCAGAAAAATACAAAGTAGCTGTTCTG
CTAAACGAAAAACGAAAAGAGATTCAAGAGGCTATTGAACGAGAAAAGAACAGTTTCGCA
CAACTCTCTAATTTTATTGACAAACTTGCGGAAACTCCCGCACTTCTATAGAATCAGAA
AACGTGCACAAAGTTCACACGAGCGACATCACTGAAGGCATCGTTCCGCGTTTTCATAACC
GAATCGATTGAACTACCAAACATCTCCACGCTAAACAACACCCAAACAAATATCCCTAGAT
AAACAACCTCAATGAAAACTCACCAACACCATTACACGCTCACAAAACAAATTCACCAAA
ATCGTACAAGATAACTATAACAATATCGCTGCCGTTTTCATGCCGTTACCGAACTGAAT
TGCTTATTTGCCCTACCTAGTAAACCTTTATTTTAAACATTGAAGTTTTTAAAACACAGTGGC
CTAAACATTAACACTGTGTTGCTTCAAGAGGTGAAAAACTGTATGACAACACGCAATTT
CTACGCTTCGGAACCTTCTCACTTCAATATAAACCAACCTTTCAAACCTTTACCCTATCTATC
AGAAAAATGTTTCGTAGATTTCTATAACAGTCAGAAACCCTCGGACAGAGCTTCCGAAATC
CTAGCCGCAATTTGAATCCATCCTAGCGGATCCCAGCAAAAACAAAACAGTCGTAAATATA
GAAATGATCAAATCGCAACTGGAAGAACTGGGGAAAATGGAAATCTCCACCACCGAAAAC
AAACAGACAGCTGCAATCACAAAACAAATCCTTGGGGACCAAGAGTTAACCCCATTTAC
GATTTCCCTGCACCACCTCTCTGCTTATAACCTACCTAATACTACAACCGTAAAAAATTTA
CACCTCCATTTTCATCTTAGAACAACGACCAGATATCGCTATGATTCTCCACGAAAAAATA
CAATCCATACCTTGACATATATGTAGACGACATGCTCAATGACATAACGGTCCCTGAACAG
ACGTTTTTCAACCGTTCATTTCTTGTAGATCTCTTCCCAAACAGCACCGAAAAAACAGCA
CTGTTTCGAATCCGTTCTCACGCTACGGCAATTAGCAAAGAAATGCGCTAACTTAAAAACT
CTAGAAGAATTCGACGATCTGGCCCAATTCATCACGACGAATAGCGAACAACCTTCAAAC
ATGATGAGACAGCATTTTGGAAAAAATAACCAACGCTGATGGGACATATCAAATTTCTC
TACTCTCAAAAAATAATCACCACCTGAAGAGAAAAACTGGATACAGAGAGCAAAAACGGTC
GTCATCACATCTCCCGAAGAACTGACAGCCTTCTTAGCCACTGCACCCACTAAACACGCC
TTACAAACTTGCAAACCTGACCTCGATAAAGCACTACAACGACATATGGAAGAGCAAATG
AAACAACTGCCGAGAACGACAAAAAACATACTCACAATCAGAAGCACCCCTCGAAAAA
AGACTGACCGACATCTTACTTATTTCTAAAGGATGGTCAATTTTCTGTCTCTGGAAACAATG
CATTTGAATCTTTTAGAAACCTTTCTGAAACAACCTGCAAGATAACAATGTCTATCATTAC
TTTACCCATGCACCTACTGCCAGTGTAAAAGACATAGAAACGACCATCAGCAAGATCATC
TCCGACGTTATTGAAAAGATATTAATCAAACCTCCTCTAAACCCCGAACAGATGTCCAAG
GAAGAACAAAAATATACCCCTTTACTCAGTTTTCTTATCCAAGTTCAAAAAACAACATTC
TGCACAGAAGATGTAAAAACAGAAATCGACCAAATGCAAAAAGAGTATCACGTTCCCTCAAG
AAAATTGCTACCTCCGCCAATAAACATACTAGACTCAGCCACTCCATCTACGGACAAGAA
TTAAACCTCTACGAAGAACGCATTACAGAGCTGAGAAAAGAGACGAACAAAATGAAAGAA

CAGCTCTCCAAAGAATACGCATTAGCTGAAAAAAAAATTCTTCTCTCGTCACAAGATGCA
AAAAC TGACAAAATTTATCTAGTTTTAAATACGCACACGCTGAAAGAAATAAAAAACACT
CAATTCAGAGAAACAGCATTTCGCTAAAGCTCTCACGGTGGAAGTCAACAATAAAGAAAAT
CAACTCCAGGAATTGTTAAACCACTTTAAACGCACACCTAAAGGCGAAAATGGATCAGAAC
CATATAACAAAACTTTCATTCGATACGAAATGGACAGCCTTCGTTTTCGGATTCCAGGCTC
TACATCCCAGACTTCATTAACATCAAAC TGCAAGACTTTATTTCTGATCCCTTTAAAGTT
ATAAGTCAGCTAATGAACAAAGCAACCAACGAGATGCCTTACATTCAAGCAGAGATAACT
CTCAAATGGTTAAACAAAATTAATACACGACATCAACAAATTTCTGCCTGTCAGCTATAAGC
GAATTCGGTAAAGAAGCAATCCCATTCAATTACGCAGCCCTACGAGATTTGGAGTACCAG
ATCAACACCAAATACGTAGAAATTTGAAAACAAAGTCATTTGCAACGAAACGGTTCGAAAAC
ACGAAAACATTCCAAAC TGACTAAGCTCTTGAAGCAGTTAGATCCTAAACGCGTTGCC
GGTGGACAGGAACAGTATCAAAC TCTCATGAACAAGATTCTCAGTTCGAAAAC TTCTATG
CAACAGACCTACGAAAAAGAACAAC TCAAAAAAGAATACTTTGAGATCGTCGATAATGTC
GCCAGCTTCAAGTTAGCATTCAACTTTCCCTCAGCAACTACAAAATGTGGAACGGCTAATA
GAGAAGTTCAAATCGCTGCCAAAAGCCAGCCATTTGAAAAATTTCCACAGGAAAATGAT
CTACTTTCTGACTCACTCAACACAGAAAATTATATTAACGGTCTGCGCGCACTCCTCAAC
TTCATAACGGCAGCGCAAAAC TACATCCAGAACACACTTCTGAAACAGTGGGCGGTCTTT
CAGCAACAAAATTTTCATCCCATTGATTACTCAGTAGCGAATGTCAAACCTATTTCCGGAT
TTATATGCCCGCCTCAGGATTGAACGGGAACGTCAAGTGTCTACCAAGTAAACTCCGTC
TTTGAACTCACCTAATCGTTGACGATACAGGAGTTCCGCTGCAATTCATAACATCTTT
AACACGCCATAGTAAAGTTTTTCTCCCTGAACTACAAGCAGATTCATGTACCAGAAGAC
ACCCCCGCTTGGTATCTAGCCAATATAAACTCCTGTCCGTATGTAAATCTTTCATAATG
ATACTACAGCAGTTCTGGGAGAATATAATCACCCCTGGACCTAGGCCCTTACCTCAGAGAC
GGTACACAAAATTTTAAAAGAGAACTCATACCGATCGTGAAC TTGAAATTTATTTATATAC
TGCATAACACAAGCCTGGACAGCCTCTGAAGACAGTACTGTTTCCACAGCGTTTCGAATTA
CCCATCAAGCAATTCACACTTTTGTACTATGCTCACATCCTGAATACCTCTACGGCTGC
CTCAGTCATTCACAGATCTGGTTATAAATTCATTAGCCAAAAGCATAGACAAAAACAGT
CTCTATAACACATTCGTCGTTAGCCACAACCCTCCCGAGAAGCCATGCATTTAATGAGA
AATATATGTATCGACACCCAGCTATGGCAACCCGCCAAAAC TTATGAAAGATACCTTCCAA
CAAACATTTTTTACACAAC TGTGCCCAAAAACGAAAAGTTTTTTTATCTATCTCACTGCC
TTTTTAATTTCTACCTTACAAATTCATGAACTATATCTGGATTCAATACAAACCGGCCATT
TTTACCCAAAGATCTTATCAAACCTAATTAAGGACTTATGCTCCGAATACGTACACCAA
AATAAGATCACTACGTCTCGGTAAC TCCCCACGAACCAGACACAATAAAATCCGGAGAA
AGGATTACCTCAAAAATCACCGTACACAAAGCACAAAACACACCGACACTCACTCGCCTG
CAGGCTCAGGAATATGCTTCGACTACATTTCTTTACTCCTTCTTGACAGGGTATGAAATG
ACATTCGCCATGTACATCGATACCATCGAAAAACATACCTGCTTTGCATGAGACACCTG
GAAAACGTACTCCACGATAAAGACTTCCAATCTGTCTCCTCAGAGCACGCACATTCGACATC
GACTACATCCTGAAACAGTCATGGACTAAGAACATAGTCGAACATTCACTCTTTTCAGTA
CAGCTCGATAAAATCGTCTCCTACCTCAACCACACAAAACAGAGCGACACCCAATATTCCT
TTGATATTTATTTAATTTATGACAACGAAGTTGTCAACGTTTATCTACCGCCGATGTCCACC
AATCCAAAAAAGTCGCCTTTTTACATCAAATACTTTTCCACTTTCCCGTGAAGAATAC
GAAGCGACAAATTTAATTTCCCTTCATCTCTATC AAAAACAAACAGATATTTTAAACCAA
CTTCCGCCCAGCAATAC TGAGTCTACTCGCCCGGACAAAACAAACCTCTGAAAGTTTAACC
ACTGAAAACCTCTCCGAACCAAAGTTTTAAAAGCCTGCCGTTACTGGACTTATGCCCAAG
AGCCAATCGATCATCTATCCACTGATACCAACGTACCGGAAACAGTCCAGACGTAAAA
GCTAACACGGCCAGCGCAGCGATAAAAGATGTCACTCTAGCTCGAGAAAAAATCAACGAG
TTCTCCGAATCGATAAATAACAAC TATTTCAAATTTAAAATCTATGTATCTGTAA

>5814_U31

ATGAAAATCATTACCAGTAGCACAAAACCAAATGACTCTAAATACGGACCGAGAGCCGGA
AAACAATGCATGTGCAACAGCTTTTTCATTTCTCCATACCGTTTACTTAAACGGAATTAAC
AATTCCTCAATGCGGGTACTATAGACGCCATAATGGAAGAAGGCTATCACTTAGACACG
GCCGGCACCTTAGCCCTAATGTTAAATAACTCAGATTTCGCAAGACTACAGACTCCTCACC
GAGATCCCCAAGAGGATCCATTC TAGATACGGTGTACACAGCATGAACTCTCACGGCCA
TTTAAACGGAACCTTAGACACACAAAAAATTTGACAATGAGGTATATCTCGGACTCATAGAT
TTCATAC TGTATGGCAAAGCAAGAATGGTCTACTTTTCGAGTCATCACGATCGGCGTG
CTATCGCGAGCACTTTTTTTTTCTTAAACAATACTCTCTACTTATTCGACTCTCACCCACT
GAACGAGAAGCCACAGCAGCCATCTACATCTGTCAAACATTGAAGAAGTTTATGAACTG

CTCACCACCCACGGCACCGAAGGCTTCTACTACGACGCCAGCTTCATTTTTTTCATAGAA
ACCTCCAATTTGTGCGTCTCTAGTCACGACGCCGAGCTTCTCATCCTAAAACTTATAAA
GACCCCGACATAGCGATTGCGCTAGATAAAATTTTCTCCACGGAAATCCATGAGATAAAA
AAAACAGATGATATCGGATCGCAACAAGACCTCGTTGCGGATAAAAACGACAGATCTAGAA
CACGCGCCCATAAAAGAAAGAAAACTCTCATAGTCTAGAACTAGAACTAAATGACAAA
AAGAAAAAGATACCGCATCCTTGACATATTACGCAACAGAAGTTGACCTCATTCCGAGT
TTTTATGAACTACGATCACAATTTCAATCTTTATTCCACGATCTCAAATCTTTTCTATA
ATGAAACCTCATTTCAACTGGACCATATACTTACAAGATTCTCCTATTAATCCTAATCAG
CCCTTCGCGACACCCTTTCTCTGGAACAGAGTTTTCCACTTATTGTGTCAAATTATCGAC
GTGTTTGTAGGCGTCGGATCCACAAATGACGATTCTAGCAAACAACAACAACAACTATT
TTCATAAAATTTATTATTGCCTTTCAAGGACTTTTTCCGAAGTATTCAACGAGGCGTTGGCA
GCTTGTCAAGAAAACAATCTGGACATCATCTTCATTTACAATAATTACCTGTGCAAAACC
ACGACCTTTCGAACACTCGAGAGAATCTTACTAAGCAAATTTCTGGCAATAGTTGATAAT
GATCACAAAAACATTATGAATGGGTCAAATCGTGGACCACACAAATGTTCCAAGGAATG
CCAAAAAAC TAGATGACATAGAGAATTATCTAAAAGCCTACGTGGACCATAACCTGTG
AAACACTTTCATGAATTCATTTGTCTGAATAAGGCAGAAAAATACAAAGTAGCTGTTCTG
CTAAACGAAAAACGAAAAGAGATTCAAGAGGCTATTGAACGAGACAAGAACATTTTCGCA
CAACTCTCTAATTTTATTGACAAACTTGGCGAAACTCCCGCACTTCTATAGAATCAGAA
AACGTGCACAAAGTTCACACGAACGACATCACTGAAGGCATCGTTCCGCGTTTTCATAACC
GAATCGATTGAACTACCAAACATCTCCACGCTAAACAACACCCAACAAATATCCCTAGAT
AAACAAC TCAATGAAAACTCACCACACCATTACACGCTCACAAAACAAATTCACCAAA
ATCGTACAAGATAACTATAACAATATCGCTGCCGGTTTTCATGCCGTTACCGAACTGAAT
TGCTATTTGCCCTACCTAGTAAACCTTTATTTTAAACATTGAAGTTTTTAAACACAGTGGC
CTAAACATTAACACTGTGTTACTTCAAGAGGTGAAAAACTGTATGACAACACGCAATTT
CTACGCTTCGGAAC TCTCACTTCAATATAAAACAACCTTTCAAAC TTTACCTATCCATC
AGAAAAATGTTTCGTAGATTTCTATAACAGTCAGAAACCTTCGGACAGAGCTTCCGAAATC
CTAGCCGCAATCGAATCCATCCTAGCGGATCCCAGCAAAAACAAGACAGTCGTTAATATA
GAAATGATCAAATCGCAACTGGAAGAACTGGGGAAAAATGGAATCTCCACCACCGAAAAAC
AAACAGACAGCTGCAATCACAAAACAAATCCTCGGGGACCAAGAGTTAACCCCATTTAC
GATTTCCCTGCACCACCTCTCTGCTTATAACCTACCTAATACTACAACCGTAAAAAATTTA
CACCTCCATTTTCATCTTAGAACAACGACCAGATATCGCTATGACTCTACACGATAAAAATA
CAATCCATAC TTAGCATATATGTAGACGACATGCTCAATGACATAACGGTCCCTGAACAG
ACGTTTTCAACAGTTCATTTCTTGTAGATCTCTTCCAAACAGCACCGAAAAAACAGCA
CTGTTGCAATCCGTTCTCAGCTACGGCAATTAGCAAAGAAATGCGCTAACTTAAAAACT
CTAGAAGAATTCGACGATCTGGCCCAATTCATCACGACGAATAGCGAACAAC TTTCAAAC
ATGATGAGACAGCATTTTGGAAAAAATAACCAACGCTGATGGGACATATCAAATTTCTC
TACTCTCAAAAAATAATCACCAC TGAAGAGAAAAACTGGATACAAAGAGCAAAAACGGTC
GTCATCACATCTCCCGAAGAACTGACAGCCTTCTTAGCCACTGCACCCACTAAACACGCC
TTACAAACTTGCAAACCTGACCTCGATAAAGCACTACAACGACATATGGAAGAGCAAATG
AAACAACTGCCGAGAACGACAAAAAACACATCCTCACAATCAGAAACACCTTCGAAAAA
AGACTGACCGACATCTTACTTATTTCTAAAGGATGGTCAATTTTCTGCTCTTGAAACAATG
CATTTGAATCTTTTAGAAACCTTTCTGAAACAAC TGAAGATAACAATGTATCATTAC
TTTACCCATGCAC TACTGCCAGTGT TAAAAGACATAGAAACGACCATCAGCAAGATCATC
TCCGACGTTATTGAAAAGATATTAATCAAAC TCTCTAAACCCGAAACAGATGTCCAAG
GAAGAACAAAAATATACCCCTTTACTCAGTTTTCTTATCCAAGTTCAAAAAACAACATTC
TGCACAGAAGATGTAAAAACAGAAATCGACCAAATGCAAAGAGTATCACGTTCTCAAG
AAAAATTGCTACCTCCGCCAATAAACATACTAGACTCAGCCACTCCATCTACGGACAAGAA
TTAAACCTCTACGAAGAGCGCATTACAGAGCTGAGAAAAGAGACGAACAAAATGAAAGAA
CAGCTCTCCAAAGAATACGCATTAGCTGAAAAAATAATTTCTCTCTCGTCACAAGATGCA
AAAAC TGAACAAAATTTATCTAGTTCTAAATACGCACACGCTGAAAGAAATAAAAAACACT
CAATTCAGAGAAACAGCATTCGCTAAAGCTCTCACGGTGGAAAGTCAACAATAAAGAAAAT
CAACTCCAGGAATGTTTAAACCCTTTAACGCACACCTAAAGGCGAAAATGGATCAGAAC
CATATAACAAAAC TTTTATTGATACGAAATGGACAGCCTTCGTTTCGGATTCCAGGCTC
TACATCCCAGACTTCATTAACATCAAAC TGAAGACTTTATTTCTGATCCCTTTAAAGTT
ATAAGTCAGCTAATGAACAAAGCAACCAACGAGATGCCTTACATTCAAGCAGAGATAACT
CTCAAATGGTTAACAAAATTAATACACGACATCAACAAATTTCTGCCTGTCAGCTATAAGC
GAATTCGGTAAAGAAGCAATCCATTCAATTACGCAGCCCTACGAGATTTGGAGTACCAG
ATCAACACCAAATACGTCGAAAT TGAACAACAAAGTCATTTGCAACGAAACGGTCGAAAAAC

ACGAAAAACATTCCAAAAC TACTAAGCTCTTGAAGCAGTTAGATCCTAAACGCGTTGCC
GGTGGGCAGGAACAGTATCAAAC TCTCATGAACAAGATTCTCACTTCCGAAACTTCTATG
CAACAGACCTACGAAAAAGAACAAC TCAAAAAAGAATACTTTGAGATCCTCAATAATGTC
GCCAGCTTCAAGTTAGCATTCAACTTTTCCTCAGCAACTACAAAATGTGGAACGGCTAATA
GAGAAGTTCAAATCGCTGCCAAAAGCCAGCCATTTGAAAAATTTCCACAGGAAAATGAT
CTACTTTCTGACTCACTCAACACAGAAAATTATATTAACGGTCTACGCGCACTCCTCAAC
TTCATAACGGCAGCGCAAAAAC TACATCCAGAACACACTTCTGAAACAGTGGGCGGTCTTT
CAGCAACAAAATTTTCATCCCGATTGATTACTCAGTAGCGAATGTCAAACCTATTTCCGGAT
TTATATGCCCGCCTCAGGATTGAACGGGAACGGCAAGTGTCTACCAAGTAAACTCCGTC
TTTGAACTCACCTAATCGTTGACGATACAGGAGTTCCGCTGCAATTCCATAACATCTTT
AACACGCCATAGTAAAGTTTTTCTCCCTGAACTACAAGCAGATTCATGTACCAGAAGAC
ACCCCCGCTTGGTATCTAGCCAATATAAACTCCTGTCCGTATGTAAATCTTTCATAATG
ATACTACAGCAGTTCTGGGAGAATATAATCACCCCTGGACCTAGGCCCTTACCTCAGAGAC
GGTACACAAAATTTTAAAAGAGAAC TCATACCGATCGTGAAC TTGAAATTATTTATATAC
TGCATAACACAAGCCTGGACAGCCTCTGAAGACAGTACTGTTTCCACAGCGTTTGAATTA
CCCATCAAGCAATTCACACTTTTGTACTATGCTCACATCCTGAATACCTCTACGGCTGC
CTCAGTCAATCCACAGATCTGGTTATAAATTCATTAGCCAAAAGCATAGACAAAAACAGT
CTCTATAACACATTCGTCGTTAGCCACAACCCCTCCCGAGAAGCCATGCATTTAATGAGA
AATATATGTATCGACACCCAGCTATGGCAACCCGCCAAAAC TTATGAAAGATACCTTCCAA
CAAACATTTTTTACACAAC TGTGCCCAAAAACGAAAAGTTTTTTATCTATCTCACTGCC
TTTTTAATTTCTACCC TACAAATTTATGAACTATATCTGGATTCAATACAAACCGGCCATT
TTTACCCAAAGATCTTATCAAACCTAATTAAGGACTTATGCTCCGAATACGTACACCAA
AATAAGATCACTACGTCTCGGTAAC TCCCCACGAACCAGACACAATAAAAATCCGGAGAA
AGGATTACCTCAAAAATCACCGTACACAAAAGCACAAAACACACCGACACTCACTCGCCTG
CAGGCTCAGGAATATGTCTTCGACTACATTTCTTTACTCCTTCTTGACAGGGTATGAAATG
ACATTCGCCATGTACATCGATACCATCGAAAAACATACCTGCTTTGCATGAGACACCTG
GAAAACGTACTCCACGATAAAGACTTCCAATCTGTCTCAGAGCACGCACATTCGACATC
GACTACATCTGAAACAGTCAATGGACTAAGAACATAGTCGAACATTCACTCTTTTCAGTA
CAGCTCGATAAAAATCGTCTCCTACCTCAACCACACAAAACAGAGCGACACCCAATATTCTT
TTGATATTTATTTAATTTATGACAACGAAGTTGTCAACGTTTATCTACCGCCGATGTCCACC
AACCCAAAAAAGTCGCC TTTTACATCAAAAATCCTTTCCACTTTCCCGTGCAAGAATAC
GAAGCGACAAATTTAATTTCTTCCATCTCTATC AAAAACAACAGATATTTTAAACCAA
CTTCCGCCAGCAATAC TGAGTCTACTCGCCCGGGCAAACAAAACCTCTGAAAGTTTAACC
ACTAAAACCTCTCCGAACCAAAGTTTAAAAGCCTGCCGTTACTGGACTTATGCCCAAG
AGCCAATCGATCATCTATCCACTGATACCAACGTACCGGAAACCAGTCCAGACGTAAAA
GCTAACACGGCCAGCGCAGCGATAAAAAGATGTCACTCTAGCTCGAGAAAAAATCAACGAG
TTCTCCGAATCGATAAATACAAC TATTTCAAATTTAAAATCTATGTATCTGTAA

U32

>2284_U32

ATGACGACCATACGTAGTGATGATTTGTGCGAATCAGATCACTCAGATTTTCAAGTTCTTCA
AAGAAAGAAGAAGAAAAGAAGAAGCAGCAGATGCTTACCGGTGTTTTGGGGCTGCAGCCT
AATATGGCCAGTCAATCCCGTTCTTGGGGTTTTTTTTGCCAAAATATGCTAAGCAGAACGGC
GGTAACGTGGATAAAAACGGCGTTCCGACTCGATTTGATAAGAATGTGGCGCTGCATAGA
CTGAACACAAAACAGGGAGTGATTGA

>5055_U32

ATGACGACCATACGTAGTGATGATTTGTGCGAATCAGATCACTCAGATTTTCAAGTTCTTCA
AAGAAAGAAGAAGAAAAGAAGAAGCAGCAGATGCTTACCGGTGTTTTGGGGCTGCAGCCT
AATATGGCCAGTCAATCCCGTTCTTGGGGTTTTTTTTGCCAAAATATGCTAAGCAGAACGGC
GGTAACGTGGATAAAAACGGCGTTCCGACTCGATTTGATAAGAATGTGGCGCTGCATAGA
CTGAACACAAAACAGGGAGTGATTGA

>5814_U32

ATGACGACCATACGTAGTGATGATTTGTGCGAATCAGATCACTCAGATTTTCAAGTTCTTCA
AAGAAAGAAGAAGAAAAGAAGAAGCAGCAGATGCTTACCGGTGTTTTGGGGCTGCAGCCT

AATATGGCCAGTCATCCCGTTCCTGGGGTTTTTTTTGCCAAAATATGCTAAGCAGAACGGC
GGTAAAGTGGATAAAACGGCGTTCGACTCGATTTGATAAGAATGCTGGCGCTGCATAGA
CTGAACACAAAAACAGGGAGTGATTGA

U34

>2284_U34

ATGGCTAACGTACTCAAAGAAAAGATGTACGATGAACTACTGAGTGCGACCTGTCGTATC
TTAAAGTTGGGGTCTCATGACTATAGGATAACTGAGCGGAATCTATTGTGAAGAACCCG
AAGTTTCCTTTGTGTGATATCATTCTCAAGTTAGACTACGCTTATAATCTAGAATATCTG
CTAAGTCTTTGGGAACATGTAACATAAGCAGGAACCCAGATTTGTGTTCAAGAACACAGGC
GGAGCTGTGTCGATGCTTTGCTATTTGCATGCCCCGGTTAAAGTAGAAGGTCATCATGCT
GTGAGAGAATGCAACATCCTGCGTGTTAATGAATGTTTGACTGTCCGCATGAGCGACATA
GTGGCTATGAAGCCTTCAACTTTTGCAGTGTTACGAAATGTATAATCAGGCGTAATCGA
GACGACACTTACGTTGTTGAGTTTGTGGCGTTCGGACCTGAAAACGAATCTGAATACATT
TCGCTTTTGAAGGCAATTTTCTTAAAGAAGTGTTCTATGGGGAAACAGCATCTCGAATCA
AATAGGTTTTGTCAAGTTTTGAGAAGAAGGTTTACATGTTTTGGAAAAGGGGCGATTC
GGAAGTAGCGGAAAGTTGCAAATAAAGCTTCTGCGTTGTGACATCACAGGAGAGTATA
AAGCAGTCTATGAAAAAGAGAAGTCATTACTGTCCGGTGTGAAGTTTTGGAGACTAAGT
GAAAGGCATTGTGCGTTTGCCTTGTGCGCATCTGCTTTCTTTTAGCCTTGTATTTTTGT
TATGTTCTACTTAAAAAACTCCTACGCCAGCTTCTGGCTCCGTTGTGTAA

>5055_U34

ATGGCTAACGTACTCAAAGAAAAGATGTACGATGAACTACTGAGTGCGACCTGTCGTATC
TTAAAGTTGGGGTCTCATGACTATAGGATAACTGAGCGGAATCTATTGTGAAGAACCCG
AAGTTTCCTTTGTGTGATATCATTCTCAAGTTAGACTACGCTTATAATCTAGAATATCTG
CTAAGTCTTTGGGAACATGTAACATAAGCAGGAACCCAGATTTGTGTTCAAGAACACAGGC
GGAGCTGTGTCGATGCTTTGCTATTTGCATGCCCCGGTTAAAGTAGAAGGTCATCATGCT
GTGAGAGAATGCAACATCCTGCGTGTTAATGAATGTTTGACTGTCCGCATGAGCGACATA
GTGGCTATGAAGCCTTCAACTTTTGCAGTGTTACGAAATGTATAATCAGGCGTAATCGA
GACGACACTTACGTTGTTGAGTTTGTGGCGTTCGGACCTGAAAACGAATCTGAATACATT
TCGCTTTTGAAGGCAATTTTCTTAAAGAAGTGTTCTATGGGGAAACAGCATCTCGAATCA
AATAGGTTTTGTCAAGTTTTGAGAAGAAGGTTTACATGTTTTGGAAAAGGGGCGATTC
GGAAGTAGCGGAAAGTTGCAAATAAAGCTTCTGCGTTGTGACATCACAGGAGAGTATA
AAGCAGTCTATGAAAAAGAGAAGTCATTACTGTCCGGTGTGAAGTTTTGGAGACTAAGT
GAAAGGCATTGTGCGTTTGCCTTGTGCGCATCTGCTTTCTTTTAGCCTTGTATTTTTGT
TATGTTGTACTTAAAAAACTCCTACGCCAGCTTCTGGCTCCGTTGTGTAA

>5814_U34

ATGGCTAACGTACTCAAAGAAAAGATGTACGATGAACTACTGAGTGCGACCTGTCGTATC
TTAAAGTTGGGGTCTCATGACTATAGGATAACTGAGCGGAATCTATTGTGAAGAACCCG
AAGTTTCCTTTGTGTGATATCATTCTCAAGTTAGACTACGCTTATAATCTAGAATATCTG
CTAAGTCTTTGGGAACATGTAACATAAGCAGGAACCCAGATTTGTGTTCAAGAACACAGGC
GGAGCTGTGTCGATGCTTTGCTATTTGCATGCCCCGGTTAAAGTAGAAGGTCATCATGCT
GTGAGAGAATGCAACATCCTGCGTGTTAATGAATGTTTGACTGTCCGCATGAGCGACATA
GTGGCTATGAAGCCTTCAACTTTTGCAGTGTTACGAAATGTATAATCAGGCGTAATCGA
GACGACACTTACGTTGTTGAGTTTGTGGCGTTCGGACCTGAAAACGAATCTGAATACATT
TCGCTTTTGAAGGCAATTTTCTTAAAGAAGTGTTCTATGGGGAAACAGCATCTCGAATCA
AATAGGTTTTGTCAAGTTTTGAGAAGAAGGTTTACATGTTTTGGAAAAGGGGCGATTC
GGAAGTAGCGGAAAGTTGCAAATAAAGCTTCTGCGTTGTGACATCACAGGAGAGTATA
AAGCAGTCTATGAAAAAGAGAAGTCATTACTGTCCGGTGTGAAGTTTTGGAGACTAAGT
GAAAGGCATTGTGCGTTTGCCTTGTGCGCATCTGCTTTCTTTTAGCCTTGTATTTTTGT
TATGTTCTACTTAAAAAACTCCTACGCCAGCTTCTGGCTCCGTTGTGTAA

U35

>2284_U35

ATGATGGATTCTTCACCTGAAAAGACGAATCTTGAACTTTTGTATGAGCGAGTCTGTGAG
CAGGGAAGGGAGTTTGAGGTTGTGTTTTATCCAATGTTGCCACGATTATATGAAATGATG
CTTCCTTCTTTGGAGGCGCGACTGAATTTTTCTTAGTGTTGGATATAGACACGTTGCTTTT
GCTAGATACGTTACGGGGATGTCGATTGCGTGCATAGAGAGGTCATGGCTCAAAAAATG
GTAATTTTAAATACGGAATAG

>5055_U35

ATGATGGATTCTTCACCTGAAAAGACGAATCTTGAACTTTTGTATGAGCGAGTCTGTGAG
CAGGGAAGGGAATTTGAGGTTGTGTTTTATCCAATGTTGCCACGATTATATGAAATGATG
CTTCCTTCTTTGGAGGCGCGACTGAATTTTTCTTAGTGTTGGATATAGACACGTTGCTTTT
GCTAGATACGTTACGGGGATGTCGATTGCGTGCATAGAGAGGTCATGGCTCAAAAAATG
GTAATTTTAAATACGGAATAG

>5814_U35

ATGATGGATTCTTCACCTGAAAAGACGAATCTTGAACTTTTGTATGAGCGAGTCTGTGAG
CAGGGAAGGGAGTTTGAGGTTGTGTTTTATCCAATGTTGCCACGATTATATGAAATGATG
CTTCCTTCTTTGGAGGCGCGACTGAATTTTTCTTAGTGTTGGATATAGACACGTTGCTTTT
GCTAGATACGTTACGGGGATGTCGATTGCGTGCATAGAGAGGTCATGGCTCAAAAAATG
GTAATTTTAAATACGGAATAG

U36

>2284_U36

ATGTATAAAGGATGGGACTCAGAAGCGTTGCCTATACAATCTGAGATACTAAACGAGATT
CTGTTATATTGCTACCTCACTCCTCAATCACCGATCCCTTCCGAGACGACAACAGCCACC
AGCCCCACAAACATAGAAAACGAAATATCCAATTTAGACAACCTTGAAAACCTAGAGGAA
TTAAAGCAACTATCCGTGCGATTAAATATAGACAGACGATGCAACATCTGCTCCATCGTA
AATCTGTGCCCTCAAACAAAACAAATCATGGATCTACGACTACAGTCTACTGTGCTACAAA
TGCAACTACGCGCCAAAACCTCCACTGTCTCTTCTGATCGTCTCCGAGAATTTATAATG
TTAATACGGGAACGATTTCCAAACATCAATTTGACGGCCTTTTTCTAAACAACATTGTC
TCCATCTTTGACTTTTACGTTTCAATTTTTTATTACAGATGCTTCGCAAATACAGTCAAC
GATCACATCCAAAGCGAAAACATAACCCTGAATCACATGGCAATAATCCGATCGACCCTG
CTAAAAGAAGACACTATACCGCACATCAAAAATAAAAAAATTTCTAACCAAAAAAATGAAC
CCGAAAAAACACAAAGCTCCGAACCTGAATAAAAAAATCACAGTGCCCATGAAGACTCGC
TTCACGACACTCTTATTCTATATGTGGTCTGGCAGAACGTTTCGACCGAGTTCCCTTT
ACGGACCTGACTATCCGCAAACATCGCTTCATCAAAAACCTGTATTCCAATAAAACCGAC
ATCGAACTCACAGCTGGTCCCATCTTTTTGGCACAGATTCCATTTTCTATCACCAAAAAAC
AAAACAACCAGCGTCTGTCTCCATGTGAACCTGATGGCAGCTTCCAAACAAGATTACCTA
TTTTTAAAATATCTACATCAGAGCATCATGGACTACTGTCAAAAACAACCTAAAGATGATA
GATAGAGTTCAGTTCGTCATCGCCGATATCTTCGAGAAGACAAAAATACACACGCACGTC
AAAAATCTTCCGACTACAGCAAGGCGATATTCGATAACGAATTCTCATTCTCAGATGAT
AACTTCACTCTCGATACCCACATTTACCTGATACTCCGCCAGACTGGAACAGTGGGAGTC
TACAAGCACTTCTTTTGTGATCCCTGTGTCTAGCAAACCTGTAACCATCAACCCGGAG
GTTCTCTTCAACACTACAGACGCCGAGAAATCCAAGACTTGAAAGTCACCATATGCTAC
CATAACGAATATTTAAGCATCGTCGAAAAACATGTTTGGCTAGCGATACATCTTTTCAA
GCTTTTCAGATTATTAACCAAACATAAAAAATAAACTCAAATTACAGAGTTTCTCAA
GACTTCACTAACCTATTAGCACTTCAACACTTCGATATCATCGATCCATTTTACCGTC
AACTACTACGTTTGA

>5055_U36

ATGTATAAAGGATGGGACTCAGAAGCGTTGCCTATACAATCTGAGATACTAAACGAAATT
CTGTTATATTGCTACCTCACTCCTCAACCACCGATCCCTTCCGAGACGACAACAGCCACC
AGCCCCACAAACATAGAAAACGAAATATCCAATTTAGACAACCTTGAAAACCTAGAGGAA
TTAAAGCAACTATCCGTGCGATTAAATATAGACAGACGATGCAACATCTGCTCCATCGTA

AATCTGTGCCCTCAAACAAAACAAATCATGGATCTACGACTACAGTCTACTGTGCTACAAA
TGCAACTACGCGCCAAAACTCCACTGTCTCTTCTGATCGTCTCCGCAGAATTTATAATG
TTAATACGGGAACGATTTCCAAACATCAATTTTCGACGGCCTTTTTCTAAACAACATTGTC
TCCATCTTTGACTTTTACGTTTCAATTTTTTTTATTACAGATGCTTCGCAAATACAGTCAAC
GATCACATCCAAAGCGAAAACATAACCCTGAATCACATGGCAATAATCCGATCGACCCTG
CTAAAAGAAGACACTATACCGCACATCAAAATAAAAAAATTTCTAACCAAAAAAATGAAC
CCGAAAAAACACAAAGCTCCGAACCTGAATAAAAAAATCACAGTGCCCATGAAGACTCGC
TTCACGACACTCTTATTCTATATGTGGTCTGGCACGAACGTCTTCGACCGAGTTCCCTTC
ACGGACCTAACTATCCGCAAACATCGCTTCATCAAAAACCTGTATTCCAATAAAAACCGAC
ATCGAACTCACAGCTGGTCCCATTCTTTTGGCACAGATTCCATTTTCTATCACAAAAAAC
AAAACAACCAGCGTCTGTCTCCTATGTGAACCTGATGGCAGCTTCCAAACAAGATTACCTA
TTTTTAAAAATTTCTACATCAGAGCATCATGGACTACTGTCAAAAACAACCTAAAGATGATA
GATAGAGTTCAGTTCGTTCATAGCCGATATATTCGAGAAGACAAAAATACACACGCACGTC
AAAAATCTTTCCGACTACAGCAAGGCGATATTCGATAACGAATTCTCATTCTCAGATGAT
AACTTCACTCTCGATACCCACGTTTACCTGATACTCCGCCAGACTGGAACAGTGGGAGTC
TACAAGCACTTCTTTTGTGATCCCCGTGTGTCTAGCAAACCTGTAACCATCAACCCGGAG
GTTCTCTTCAACACTACAGACGCCGGAGAAATCCAAGACTTGAAAGTCACCATATGCTAC
CGTAACGAATATTTAAGCATCGTCGAAAAACATGTTTGGCTAGCGATACATCTTTTCAA
GCTTTTTCAGATTATTAACCAAACCATAAAAAATAAACTCAAATTACAGAGTTTCTCAA
GACTTCACTAACCTATTAGCACTTACCACCTTCGATATCGTCGATCCCATTTTACCCTG
AACTACTACGTTTGA

>5814_U36

ATGTATAAAGGATGGGACTCAGAAGCGTTGCCCTATACAATCTGAGATACTAAACGAGATT
CTGTTATATTGCTACCTCACTCCTCAATCACCGATCCCTTCCGAGACGACAACAGCCACC
AGCCCCACAAACATAGAAAACGAAATATCCAATTTAGACAACCTTGAAAACCTAGAGGAA
TTAAAGCAACTATCCGTGCGATTAATATAGACAGACGATGCAACATCTGCTCCATCGTA
AATCTGTGCCCTCAAACAAAACAAATCATGGATCTACGACTACAGTCTACTGTGCTACAAA
TGCAACTACGCGCCAAAACTCCACTGTCTCTTCTGATCGTCTCCGCAGAATTTATAATG
TTAATACGGGAACGATTTCCAAACATCAATTTTCGACGGCCTTTTTCTAAACAACATTGTC
TCCATCTTTGACTTTTACGTTTCAATTTTTTTTATTACAGATGCTTCGCAAATACAGTCAAC
GATCACATCCAAAGCGAAAACATAACCCTGAATCACATGGCAATAATCCGATCGACCCTG
CTAAAAGAAGACACTATACCGCACATCAAAATAAAAAAATTTCTAACCAAAAAAATGAAC
CCGAAAAAACACAAAGCTCCGAACCTGAATAAAAAAATCACAGTGCCCATGAAGACTCGC
TTCACGACACTCTTATTCTATATGTGGTCTGGCACGAACGTCTTCGACCGAGTTCCCTTT
ACGGACCTGACTATCCGCAAACATCGCTTCATCAAAAACCTGTATTCCAATAAAAACCGAC
ATCGAACTCACAGCTGGTCCCATTCTTTTGGCACAGATTCCATTTTCTATCACAAAAAAC
AAAACAACCAGCGTCTGTCTCCTATGTGAACCTGATGGCAGCTTCCAAACAAGATTACCTA
TTTTTAAAAATATCTACATCAGAGCATCATGGACTACTGTCAAAAACAACCTAAAGATGATA
GATAGAGTTCAGTTCGTTCATAGCCGATATCTTCGAGAAGACAAAAATACACACGTACGTC
AAAAATCTTTCCGACTACAGCAAGGCGATATTCGATAACGAATTCTCATTCTCAGATGAT
AACTTCACTCTCGATACCCACATTTACCTGATACTCCGCCAGACTGGAACAGTGGGAGTC
TACAAGCACTTCTTTTGTGATCCCCGTGTGTCTAGCAAACCTGTAACCATCAACCCGGAG
GTTCTCTTCAACACTACAGACGCCGGAGAAATCCAAGACTTGAAAGTCACCATATGCTAC
CATAACGAATATTTAAGCATCGTCGAAAAACATGTTTGGCTAGCGATACATCTTTTCAA
GCTTTTTCAGATTATTAACCAAACCATAAAAAATAAACTCAAATTACAGAATTTCTCAA
GACTTCACTAACCTATTAGCACTTACCACCTTCGATATCGTCGATCCCATTTTACCCTG
AACTACTACGTTTGA

U37

>2284_U37

ATGACTGTACACAAGAGCCGTATTCGACGATCGCGAAGCCTGTCCGTGACACACAGAATA
CAAAAACGACCCGATCACAGAGAAAAACAAAGCTTTACTTACAACCTGAAACTTCACGAT
CTACACACCGTCTTCAACCTGTTTCCAGAATACGAACAGAAATTTCTAGCGATCATAAAA
CTGCCAATTACCGGTAAAGAACCCATCGACGTCCATTTCAGTCTCAGCAATCACCAACAA
CACACCTGCCCTAGAATTTCCCCCTACGCTAACGAACAAATATCCAAAAGCGCCTGCCTA

CACTGCGAATCCGTCTCCGTTCCAACCTCCTCAGATGCAATGGTCGCCCACTTGAACCAA
GTCAACAACGTCATGCAAAACAGACTTTACTTTCTACGGTTTTTCGAAAGGACATGGAGCTC
ATACGAATGTCCGCCAAACAACCCACTATCTTTCAAATCTTTTATATTGTCCACAACACG
ATCAACAACATATTCCTTATCATGTTTCGAAAAAAAACAAAAGCTCGGAATGCACATAGTT
TTCCAGTCTCGCACTCTGCACATTCCTGTGAATGCATTAACAAAATAGTAGCGGTATCC
TCGGGATACAACGTTTATCTAGATATACTACAAGAAAGTGAATCCTAACTGTCTCTGC
GAAACACTAGATACCAATACTAATATCCACATCGACATAGGGATGCTTCAAAAAAACTC
GAAGAAATGGACATCCCAAACGAAATAAGCGACAGGCTCGAAAAATACAAAGGACACCTA
ATAGGTTTCCATTAA

>5055_U37

ATGACTGTACACAAGAGCCGTATTCGACGATCGCGAAGCCTGTCCGTGACACACAGAATA
CAAAAACGACCTGATCACAGAGAAAAACAAAGCTTTACTTTACAACCTGAACTTCAGGAT
CTACACACCGTCTTCAACCTGTTTCCAGAATACGAACAGAAATTTCTAGCGATCATAAAA
CTGCCAATTACCGGTAAAGAACCCATCGACGTCCATTTCAGTCTCAGCAATCACCACCAA
CACACCTGCCTAGAATTCCTCCCTACGCTAACGAACAAAATATCCAAAAGCGCCTGCCTA
CACTGCGAATCCGTCTCCGTTCCAACCTCCTCAGATGCAATGGTCGCCCACTTGAACCAA
GTCAACAACGTCATGCAAAACAGACTTTACTTTCTACGGTTTTTCGAAAGGACATGGAGCTC
ATACGAATGTCCGCCAAACAACCCACTATCTTTCAAATCTTTTATATTGTCCACAACACG
ATCAACAACATATTCCTTATCATGTTTCGAAAAAAAACAAAAGCTCGGAATGCACATAGTT
TTCCAGTCTCGCACTCTGCACATTCCTGTGAATGCATTAACAAAATAGTAACGGTATCC
TCGGGCTACAACGTTTATCTAGATATACTACAAGAAAGTGAATCCTAACTGTCTCTGC
GAAACACTAGATACCAATACTAATATCCACATCGACATAGGGATGCTTCAAAAAAACTC
GAAGAAATGGACATCCCAAACGAAATAAGCGACAGACTCGAAAAATACAAAGGATACCTA
ATAGGTTTCCATTAA

>5814_U37

ATGACTGTACACAAGAGCCGTATTCGACGATCGCGAAGCCTGTCCGTGACACACAGAATA
CAAAAACGACCCGATCACAGAGAAAAACAAAGCTTTACTTTACAACCTGAACTTCACGAT
CTACACACCGTCTTCAACCTGTTTCCAGAATACGAACAGAAATTTCTAGCGATCATAAAA
CTGCCAATTACCGGTAAAGAACCCATCGACGTCCATTTCAGTCTCAGCAATCACCACCAA
CACACCTGCCTAGAATTCCTCCCTACGCTAACGAACAAAATATCCAAAAGCGCCTGCCTA
CACTGCGAATCCGTCTCCGTTCCAACCTCCTCAGATGCAATGGTCGCCCACTTGAACCAA
GTCAACAACGTCATGCAAAACAGACTTTACTTTCTACGGTTTTTCGAAAGGACATGGAGCTC
ATACGAATGTCCGCCAAACAACCCACTATCTTTCAAATCTTTTATATTGTCCACAACACG
ATCAACAACATATTCCTTATCATGTTTCGAAAAAAAACAAAAGCTCGGAATGCACATAGTT
TTCCAGTCTCGCACTCTGCACATTCCTGTGAATGCATTAACAAAATAGTAGCGGTATCC
TCGGGATACAACGTTTATCTAGATATACTACAAGAAAGTGAATCCTAACTGTCTCTGC
GAAACACTAGATACCAATACTAATATCCACATCGACATAGGGATGCTTCAAAAAAACTC
GAAGAAATGGACATCCCAAACGAAATAAGCGACAGGCTCGAAAAATACAAAGGACACCTA
ATAGGTTTCCATTAA

U38

>2284_U38

ATGGATTCGGTGTGCTTTTTTAATCCATATTTGGAAGCGAATCGCTTAAAGAAAAAAGC
AGATCGAGTTACATTCGTATACTTCTCGCGGTATAATGCATGATGGTGCGGCGGGCTTA
ATAAAGGATGTTTGTGACTCTGAACCGCGTATGTTTTATCGAGACCGACAGTATTTACTG
AGCAAAGAAATGACCTGGCCGAGTTTGGACATAGCTCGGTCCAAGGATTATGATCATATG
AGGATGAAGTTTCACATATATGATGCTGTAGAAACGTTAATGTTTACGGATTTCGATCGAG
AATCTTCTTTTCAGTATAGACATTTTGTGATTCTTTCGGGGACAGTGATTAGAATGTTT
GGGAGAAGTGAAGACGGTGAGAAGATCTGCGTGAACGTGTTTGGACAGGAGCAATATTTTC
TACTGCGAATGCGTCGACGGAAGAAAGCCTGAAGGCTACTATAAACAATTTGATGTTAACC
GGCGAGGTTAAAATGTCGTGTTCTTTTGTGATTGAGCCGGCTGATAAGTTGTCGTTGTAT
GGGTACAATGCCAACACTGTCGTTAATCTGTTTAAAGTGAGTTTTGGAAATTTTTATGTA
TCTCAACGTATTGGAAAGATTCTGCAGAATGAGGGATTTCGTAGTTTATGAAATCGACGTA
GATGTTTTGACTCGTTTTCTTCGTCGATAATGGTTTTTTGAGTTTCGGATGGTATAATGTA

AAAAAATATATTCCTCAAGATATGGGAAAAGGGAGTAATCTTGAGGTGGAAATTAATTGT
CATGTCTCTGATTTAGTTTCTTTGGAAGACGTTAATTGGCCCTTATATGGATGCTGGTCT
TTCGACATAGAGTGTGGGCAAAAATGGGAATTTCCCGGATGCCGAAAATTTAGGTGAT
ATAGTTATTCAGATTTCTGTAATTAGTTTTCGATACGGAAGGTGACCGTGATGAGCGACAT
CTGTTTACTCTGGGAACATGTGAAAAAATTGACGGCGTGCATATATATGAATTTGCGTCA
GAGTTTGAATTACTTTTTGGGTTTTTTTCATATTTTTAAGGATTGAATCTCCGGAGTTTATT
ACCGGTTATAATATTAATAATTTTTGATTTAAAAATATTTGTGTATAAGGATGGATAAGATT
TACCATTATGATATTGGTTGTTTTTCTAAACTGAAGAATGGAAAGATTGGAATCTCTGTC
CCTCACGAACAGTACAGGAAGGGATTCCCTTCAGGCGCAAACCAAGGTGTTTACTTCCGGA
GTGTTGTATCTGGATATGTATCCCGTCTATTCTAGTAAAATAACGGCTCAGAATTACAAA
CTGGATACTATTGCTAAGATCTGTCTCCAGCAAGAAAAGGAGCAGTTATCGTACAAGGAA
ATACCAAAGAAAATTTATTAGCGGACCCAGTGGCAGGGCTGTTGTTGGCAAATATTGTCTG
CAGGACTCTGTCTTAGTTGTGCGTCTCTTTAAACAGATTAATTATCATTATGAGGTTGTT
GAGGTCGCCAGATTGGCACACGTCACGGCTAGATGTGTGGTGTTCGAGGGTCAGCAGAAG
AAGATATTTCCCTGCATTCCTACGGAAGCAAACGCCGTAATATGATTCTTCCGAGTATG
GTGCTTTCGCACAATAGACAAGGGATAGGTTACAAAGGGGCTACCGTTTTGGAGCCTAAG
ACGGGTTATTATGCTGTGCCACCGTGGTGTGTTGATTTCAAAGTTTGTATCCGAGCATT
ATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATAGCTGGA
TTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGGATGAGACTCATCGTTTTGTGAAG
CCTTGTATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAGAGACGA
GAAGTGAAGGCGGAGATGCAGAACTGTTCCGGATCCGATGATGAAACTTCTTCTGGATAAAA
AAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACAGGAGCGGGCAC
GGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGAAGGGAGATGCTT
TGTTCCACGGTGGATTATGTTAATCCAAGATGCAGTCCGAGCAATTTTTTTGCGAGGAA
TTTGGTTTAAACGTATCAGATTTTACTGGTGATTTGAAAGTGGAGGTAATTTATGGTGAT
ACGGATAGCATTTTTATGTCTGTGCAAAATATGGTTAATCAGTCTCTGCGAAGGATTGCG
CCGATGATCGCAAACATATCACAGATCGTCTGTTCAAGTCGCCTATCAAGCTCGAGTTT
GAAAAGATTTTATGTCCGCTAATTTTTGATTTGAAAAAAGATACATTGGTAGACAGGAT
GATTCGCTTTTTAATTTTTAAGGGGGTAGATCTGGTGAGAAAGACTTCTTTCGATTTTTGTG
AAGGGTGTGGTGAAGATATCGTGGACTTGTGTTCTTTGATGAAGAGGTTTCAGACTGCT
GCTGTGGAGTTTTCTCACATGACACAAACACAGTTGCGTGAGCAAGGAGTGCCTGTGGGT
ATTCATAAAAATTTTGCCTGTCTGTGCGAAGCGCGGGAGGAGCTTTTTTCAGAATCGGGCA
GACGTGAGACATTTAATGTTGTCTCTGTGCTTTCAAAGAAATGGCTGCATATAAGCAA
CCGAATCTGGCTCACCTTAGCGTCATTAGAAGTTGGCGCAGAGAAAGGAAGAAATCCA
AATGTAGGTGACCGAATTATGTACGTGTTAATAGCACCATCTATTGGCAATAAACAGACG
CATAACTATGAATTAGCAGAAGATCCAACTATGTGATAGAACACAAGATTCCATACAT
GCGGAGAAGTATTTGATCAGATTATCAAGGCTGTGACTAATGCGATCTCACCCATTTTT
CCGAAAACCGATATAAAAAAAGAGAAGTTACTGTTGATTTACTTCTATGAAAGTGTAT
TTGGATGAAACATTTTCTGCTATTGCAGAGGTAATGTGA

>5055_U38

ATGGATTCCGGTGTGCTTTTTTAATCCATATTTGGAAGCGAATCGCTTAAAGAAAAAAGC
AGATCGAGTTACATTCGTATACTTCCCTCGCGGTATAATGCATGATGGTGCGGCGGGCTTA
ATAAAGGATGTTTGTGACTCTGAACCGCGTATGTTTTATCGAGACCGACAGTATTTACTG
AGCAAAGAAATGACCTGGCCGAGTTTGGACATAGCTCGGTCCAAGGATTATGATCATATG
AGGATGAAGTTTCACATATATGATGCTGTAGAAACGTTAATGTTTACGGATTTCGATCGAG
AATCTTCTTTTTCAGTATAGACATTTTGTGATTCTTCCGGGACAGTGATTAGAATGTTT
GGGAGAACTGAGGACGGTGAGAAGATCTGCGTGAACGTGTTTGGACAGGAGCAATATTTT
TACTGCGAATGCGTCGACGGA AAAAGCCTGAAGGCTACTATAAAACAATTTGATGTTAACC
GGCGAGGTTAAAATGTGCTGTTCTTTTTGTCAATTGAGCCGGCTGATAAGTTGTCGTTGTAT
GGGTACAATGCCAACACTGTGCTTAATCTGTTTAAAGTGAGTTTTGGAAATTTTTATGTA
TCTCAACGTATTGGAAAGATTCTGCAGAATGAGGGATTCTGATGTTTATGAAATCGACGTA
GATGTTTTGACTCGTTTTCTTCGTCGATAATGGTTTTTTGAGTTTTCGGATGGTATAATGTA
AAAAAATATATTCCTCAAGATATGGGAAAAGGGAGTAATCTTGAGGTGGAAATTAATTGT
CATGTCTCTGATTTAGTTTCTTTGGAAGACGTTAATTGGCCCTTATATGGATGCTGGTCT
TTCGACATAGAGTGTGGGCAAAAATGGGAATTTCCCGGATGCCGAAAATTTAGGTGAT
ATAGTTATTCAGATTTCTGTAATTAGTTTTCGATACGGAAGGTGACCGTGATGAGCGACAT
CTGTTTACTCTGGGAACATGTGAAAAAATTGACGGCGTGCATATATATGAATTTGCGTCA

GAGTTTGAATTACTTTTGGGTTTTTTCATATTTTTAAGGATTGAATCTCCGGAGTTTATT
ACCGTTTATAATATTAATAATTTTTGATTTAAAATATTTGTGTATAAGGATGGATAAGATT
TACCATTATGATATTGGTTGTTTTTCTAAACTGAAGAATGGAAAGATTGGAATCTCTGTC
CCTCACGATCAGTACAGGAAGGGATTCCCTTCAGGCGCAGACCAAGGTGTTTACTTCCGGA
GTGTTGTATCTGGATATGTATCCCGTCTATTCTAGTAAAATAACGGCTCAGAATTACAAA
CTGGATACTATTGCTAAGATCTGTCTCCAGCAAGAAAAGGAGCAGTTATCGTACAAGGAA
ATACCAAAGAAATTTATTAGCGGACCCAGTGGCAGGGCTGTTGTTGGCAAATATTGTCTG
CAGGACTCTGTCTTAGTTGTGCGTCTCTTTAAACAGATTAATTATCATTATGAGGTTGTT
GAGGTCGCCAGATTGGCACACGTCACGGCTAGATGTGTGGTGTTCGAGGGTCAGCAGAAG
AAGATATTTCCCTGCATTCCTTACGGAAGCAAAACGCCGTAATATGATTCTTCCGAGTATG
GTGCTTTCGCACAATAGACAAGGGATAGGTTACAAAGGGGCTACCGTTTTGGAACCTAAG
ACGGGTTATTATGCTTTGCCACCGTGGTGTGTTGATTTTTCAAAGTTTGTATCCGAGCATT
ATGATGGCGCATAATCTGTGTTATAGTACTTTAGTTTTGGATGAACGACAGATAGCTGGA
TTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGGATGAGACTCATCGGTTTTGTGAAG
CCTTGTGTCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAGAGACGA
GAAGTGAAGGCGGAGATGCAGAATGTTCCGGATCCGATGATGAAACTTCTTCTGGATAAAA
AAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACGGGAGCGGGCAC
GGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAAGAGAGATGCTT
TGTTCCACGGTGGATTATGTTAATTCGAAGATGCAGTCCGAGCAATTTTTTTCGAGGAA
TTTGGTTTAAACGTACATCAGATTTTACTGGTGATTTGGAAGTGGAGGTAATTTATGGTGAT
ACGGATAGCATTTTTTATGTCTGTGCAAAATATGGTTAATCAGTCTCTGCGAAGGATTGCG
CCGATGATCGCCAAACATATCACAGATCGTCTGTTCAAGTCGCCTATCAAGCTCGAGTTT
GAAAAGATTTTATGTCCGCTAATTTTTGATTTGTAAGAAAGATACATTGGTAGACAGGAT
GATTCGCTTTTTAATTTTTAAGGGGGTAGATCTGGTGAGAAAGACTTCTTTCGATTTTTGTG
AAGGGTGTGGTGAAGATATCGTGGACTTGTGTTCTTTGATGAAGAGGTTTCAGACTGCT
GCTGTGGAGTTTTCTCACATGACACAGACACAGTTGCGTGAACAAGGAGTGCCTGTGGGT
ATTCATAAAATTTTTGCGTCGTCTGTGCGAAGCGCGGGAGGAGCTTTTTTCAGAATCGGGCA
GACGTGAGACATTTAATGTTGCTCTGTGCTTTCCAAAGAAATGGCTGCATATAAGCAA
CCGAATCTGGCTCACCTTAGCGTCATTAGAAGGTTGGCGCAGAGAAAGGAAGAAATCCCA
AATGTAGGTGACCGAATTTATGTACGTGTTAATAGCACCATCTATTGGCAATAAACAGACG
CATAACTATGAATTAGCAGAAGATCCAAACTATGTGATAGAACAAGATTCCATATACAC
GCGGAGAAGTATTTGATCAGATTATCAAGGCTGTGACTAATGCGATCTCACCCATTTTT
CCGAAAACCGATATAAAAAAAGAGAAGTTACTATTGTATTTACTTCTATGAAAGTGTAT
TTGGATGAAACATTTTTCTGCTATTGCAGAGGTAATGTGA

>5814_U38

ATGGATTCCGGTGTGCTTTTTTAAATCCATATTTGGAAGCGAATCGCTTAAAGAAAAAAGC
AGATCGAGTTACATTCGTATACTTCCCTCGCGGTATAATGCATGATGGTGCGGCGGGCTTA
ATAAAGGATGTTTGTGACTCTGAACCGCGTATGTTTTATCGAGACCGACAGTATTTACTG
AGCAAAGAAATGACCTGGCCGAGTTTGGACATAGCTCGGTCCAAGGATTATGATCATATG
AGGATGAAGTTTCACATATATGATGCTGTAGAAACGTTAATGTTTACGGATTTCGATCGAG
AATCTTCCTTTTTCAATATAGACATTTTGTGATTCTTTCGGGGACAGTGATTAGAATGTTT
GGGAGAACTGAGGACGGTGAGAAGATCTGCGTGAACGTGTTTGGACAGGAGCAATATTTTC
TACTGCGAATGCGTCGACGGAAGAAAGCCTGAAGGCTACTATAAACAATTTGATGTTAACC
GGCGAGGTTAAAATGTCGTGTTCTTTTTGTGATTGAGCCGGCTGATAAGTTGTCGTTGTAT
GGGTACAATGCCAACACTGTGCTTAATCTGTTTAAAGTGAGTTTTGGAAATTTTTATGTA
TCTCAACGTATTGGAAAGATTCGCAAGATGAGGGATTCTAGTTTTATGAAATCGACGTA
GATGTTTTGACTCGTTTTCTTCGTGATAATGGTTTTTTGAGTTTCGGATGGTATAATGTA
AAAAAATATATTCCTCAAGATATGGGAAAAGGGAGTAATCTTGGAGTGGAAATTAATTGT
CATGTCTCTGATTTAGTTTTCTTTGGAAGACGTTAATTGGCCCTTATATGGATGCTGGTCT
TTCGACATAGAGTGTGTTGGGCCAAAATGGGAATTTCCCGGATGCCGAAAATTTAGGTGAT
ATAGTTATTCAGATTTCTGTAATTAGTTTTCGATACGGAAGGTGACCGTGATGAGCGACAT
CTGTTTACTCTGGGAACATGTGAAAAAATTGACGGCGTGCATATATATGAATTTGCGTCA
GAGTTTGAATTACTTTTTGGGTTTTTTCATATTTTTAAGGATTGAATCTCCGGAGTTTATT
ACCGTTTATAATATTAATAATTTTTGATTTAAAATATTTGTGTATAAGGATGGATAAGATT
TACCATTATGATATTGGTTGTTTTTCTAAACTGAAGAATGGAAAGATTGGAATCTCTGTC
CCTCACGAACAGTACAGGAAGGGATTCCCTTCAGGCGCAAACCAAGGTGTTTACTTCCGGA
GTGTTGTATCTGGATATGTATCCCGTCTATTCTAGTAAAATAACGGCTCAGAATTACAAA

CTGGATACTATTGCTAAGATCTGTCTCCAGCAAGAAAAGGAGCAGTTATCGTACAAGGAA
ATACCAAAGAAATTTATTAGCGGACCCAGTGGCAGGGCTGTTGTTGGCAAATATTGTCTG
CAGGACTCTGTCTTAGTTGTGCGTCTCTTTAAACAGATTAATTATCATTATGAGGTTGTT
GAGGTCGCCAGATTGGCACACGTCACGGCTAGATGTGTGGTGTTCGAGGGTCAGCAGAAG
AAGATATTTCCCTGCATTCCTTACGGAAGCAAAACGCCGTAATATGATTCTTCCGAGTATG
GTGTCTTTCGCACAATAGACAAGGGATAGGTTACAAAGGGGCTACCGTTTTTGGAGCCTAAG
ACGGGTTATTATGCTGTGCCCTACCGTGGTGTGTTGATTTTTCAAAGTTTGTATCCGAGCATT
ATGATGGCGCATAATCTGTGTTATAGTACTTTTAGTTTTTGGATGAACGACAGATAGCTGGA
TTGTCAGAGAGTGACATCTTAACCGTGAAGTTGGGGGATGAGACTCATCGGTTTTGTGAAG
CCTTGTATCCGTGAGTCTGTGCTTGGGAGTCTACTAAAGGACTGGCTGGCCAAGAGACGA
GAAGTGAAGGCGGAGATGCAGAAGTTCGGATCCGATGATGAAACTTCTTCTGGATAAAA
AAGCAGCTCGCTCTGAAAACAACATGTAACCTCGGTGTACGGTGTACAGGAGCGGGCAC
GGTTATTGCCGTGTGTTGCGATTGCTGCTTCTGTAACGTGTCTTGGAAAGGGAGATGCTT
TGTTCACGGTGGATTATGTTAATTCCAAGATGCAGTCCGAGCAATTTTTTTTGCAGGAA
TTTGGTTTAAACGTATCAGATTTTACTGGTGATTTGAAAGTGGAGGTAATTTATGGTGAT
ACGGATAGCATTTTTTATGTCTGTGCAAAATATGGTTAATCAGTCTCTGCGAAGGATTGCG
CCGATGATCGCCAAACATATCACAGATCGTCTGTTCAAGTCGCCTATCAAGCTCGAGTTT
GAAAAGATTTTTATGTCCGCTAATTTTTGATTTGTAAGAAAGATACATTGGTAGACAGGAT
GATTCGCTTTTTAATTTTTAAGGGGGTAGATCTGGTGAGAAAGACTTCTTTCGATTTTTGTG
AAGGGTGTGGTGAAGATATCGTGGACTTGTGTTCTTTGATGAAGAGGTTTCAGACTGCT
GCTGTGGAGTTTTCTCACATGACACAGACACAGTTGCGTGAACAAGGAGTGCCTGTGGGT
ATTCATAAAATTTTTGCGTCTGTGCGAAGCGCGGGAGGAGCTTTTTTCAGAATCGGGCA
GACGTGAGACATTTAATGTTGTCTCTGTGCTTTTCAAAGAAATGGCTGCATATAAGCAA
CCGAATCTGGCTCACCTTAGCGTCATTAGAAGGTTGGCGCAGAGAAAGGAAGAAATCCCA
AATGTAGGTGACCGAATTATGTACGTGTTAATAGCACCATCTATTGGCAATAAACAGACG
CATAACTATGAATTAGCAGAAGATCCAACTATGTGATAGAACAAGATTCCCTATACAC
GCGGAGAAGTATTTTCGATCAGATTATCAAGGCTGTGACTAATGCGATCTCACCCATTTTT
CCGAAAACCGATATAAAAAAAGAGAAGTTACTATTGATTTACTTCTGTGAAAGTGTAT
TTGGATGAAACATTTTTCTGCTATTGCAGAGGTAATGTGA

U39

>2284_U39

ATGAGCAAGATGGCAGTATTATTCCTGGTTGTCTTTTTGATGAATAGTGTTTAATGATA
TATTGTGATCCGGATCATTATATCAGAGCGGGCTATAATCACAAATATCCTTTTTCGGATT
TGTTTCGATTGCCAAAGGCACGGATTTAATGCGGTTTCGACAGAGATATTTTCGTGCTCGCCG
TATAAGTCTAATGCAAAGATGTCGGAGGGTTTTTTTCATCATTTACAAAACAAATATCGAG
ACCTACACTTTTTCCAGTGAGAACGTATAAAAAAGAGTTGACGTTCCAAAGTAGTTACCGT
GATGTGGGTGTGGTTTTATTTCTGGATCGGACGGTGTGGGTTTGGCCATGCCGGTGTAC
GAAGCAAATTTAGTTAATCTCATGCGCAGTGTATTCCGCCGTAGCGATGAAACGACCC
GATGGTACGGTGTTTAGTGCCTTTTCATGAGGATAATAATAAAAAACAATACTCTAAATTTA
TTTCTCTGAATTTCAAATCTATAACTAATAAAAAGATTTATCACTACGAAAGAACCCTAC
TTTGCAAGGGTCTTTTGTGGCTATATTCTACATCAACGTCTCTCAATTGTATTGTGACG
GAGGCTACGGCTAAGGCGAAATACCCGTTTAGTTACTTTGCTTTGACGACTGGCGAAATC
GTGGAAGGGTCTCCGTTCTTCAACGGTTCAAACGGTAAACATTTTGCAGAGCCGTTAGAA
AAATTGACAATCTTGGAAAACTATACTATGATAGAAGATCTAATGAATGGTATGAATGGG
GCCACTACGTTAGTGAGGAAAATCGCTTTTCTCGAGAAAGCGGATACTTTGTTTTCTTGG
GAAATCAAGGAAGAGAATGAATCGGTGTGTATGCTAAAGCACTGGACCACGGTACTCAC
GGGCTTCGAGCGGAGACGAATGAGACCTATCACTTTATCTCTAAGGAGTTGACAGCCGCT
TTCGTCGCCCCCAAGGAGTCTTAAATCTTACCGATCCGAAACAAACGTGTATTAAGGAT
GAATTTGAAAAATAATTAATGAAGTCTATATGTCAGATTATAATGATACATATAGCATG
AATGGTAGTTATCAAATTTTTAAGACTACGGGAGATTTGATTTTGATTTGGCAGCCTCTT
GTGCAAAAATCTCTTATGTTTCTTGGAGGGTTCGGAAAAAATACGTAGGAGGGCAGAT
GTGGTGGATGTTAAGTCTAGACATGACATTTCTTTATGTACAATTACAGTATCTTTATGAT
ACTTTGAAAGATTATATCAATGATGCGTTGGGGAATTTGGCAGAATCTTGGTGTCTCGAT
CAAAGCGAACGATAACGATGTTGCACGAACTTAGTAAGATTAGTCCGTCGAGTATCGTG
TCTGAGGTTTACGGTCTCGATATCTGCACAGTTGCATGGTGACGTGTTAGCTATCTCG

AAATGCATAGAGGTTAACCAGTCATCCGTTTCAGCTTCATAAGAGTATGCGGGTCGTTGAT
GCAAAGGGAGTAAGGAGTGAAACGATGTGTTATAATCGGCCTTTGGTGACGTTTAGCTTT
GTGAACTCGACGCCGTGAGGTTGTCCCTGGCCAGCTAGGGTTAGATAATGAAATTTCTGTTG
GGTGATCATAGGACAGAGGAATGCGAAATACCTAGTACAAAGATCTTTTTATCTGGAAAT
CATGCGCACGTGTATACCGATTATACGCATACGAATTCGACGCCCATAGAAGACATTGAG
GTATTGGATGCTTTTTATTAGACTAAAAATCGATCCTCTTGAAAATGCCGATTTTAAAGTA
CTCGATTTATATTCCCGGACGAATTGAGTAGAGCAAACGTTTTTCGATCTAGAGAATATT
CTTCGTGAATATAACTCATATAAGAGCGCACTATATACTATAGAAGCTAAAAATTGCTACG
AATACACCTTCGTATGTTAATGGGATTAATTCTTTTTTACAAGGGCTTGGGGCTATAGGC
ACTGGATTGGGCTCGGTTATAAGTGTTACGGCAGGAGCGCTTGGGGATATTGTGGGTGGA
GTGGTGTCTTTTCTAAAAATCCATTCGGGGGTGGTCTCATGTTGATTTTAGCGATAGTA
GTTGTTGTGATAATAATTGTGGTTTTTCGTTAGACAAAGACATGTGCTTAGTAAGCCTATT
GACATGATGTTTCTTATGCCACCAATCCGGTGACTACTGTGTCCAGTGTTACGGGGACC
ACTGTGTCGAAGACACCTAGTGTTAAGGATGTTGACGGGGGCACATCTGTTGCGGTTTCG
GAAAAGGAGGAGGTATGGCTGACGTCAGTGGGCAAGTAAGTGACGATGAATATTCACAA
GAAGATGCTTTAAAAATGCTCAAGGCCATAAAGTCTTAGACGAGTTCTACAGAAGAAAA
CCTTCGTCTTCTGAGTCTCATGCCTCAAACCTAGTTTTGATAGACAGGATCAGGTATAGA
GTTATAAGAGTGTAATGTAGAAGAAGCGTGA

>5055_U39

ATGAGCAAGATGGCAGTATTATTCCTGGCTGTCTTTTTGATGAATAGTGTTTTAATGATA
TATTGTGATCCGGATCATTATATCAGAGCGGGCTATAATCACAAATATCCTTTTCGGATT
TGTTGATTCGCAAAGGTACGGATTTAATGCGGTTTCGACAGAGATATTTTCGTGCTCGCCG
TATAAGTCTAATGCAAAGATGTCGGAGGGTTTTTTCATCATTTACAAAACAAATATCGAG
ACCTACACTTTTCCAGTGAGAACGTATAAAAAAGAGTTGACGTTCCAAAGTAGTTACCGT
GATGTGGGTGTGGTTTTATTTCTGGATCGGACGGTGATGGGTTTTGGCCATGCCGGTGTAC
GAAGCAAATTTAGTTAATTCATGCGCAGTGTTATTCCGCCGTAGCGATGAAACGACCC
GATGGTACGGTGTTTAGTGCCTTTCATGAGGATAATAATAAAAACAATACTCTAAATTTA
TTTCTCTGAATTTCAAATCTATAACTAATAAAAAGATTTATCACTACGAAAGAACCCTAC
TTTGCAAGGGTCTTTTGTGGCTATATTCTACATCAACGTCTCTCAATTGTATTGTGACG
GAGGCTACGGCTAAGGCGAAATACCCGTTTAGTTACTTTGCTTTGACGACTGGCGAAATC
GTGGAAGGGTCTCCGTTCTTCAACGGTTCAAACGGTAAACATTTTGCAGAGCCGTTAGAA
AAATTGACGATCTTGGAACCTATACTATGATAGAAGATCTAATGAATGGTATGAATGGG
GCCACTACGTTAGTGAAGAAAATCGCTTTTTCTCGAGAAAGCGGATACTTTGTTTTCTTGG
GAAATTAAGGAAGAGAATGAATCGGTGTGTATGCTAAAGCACTGGACCACGGTGACTCAC
GGGCTTCGAGCGGAGACGAATGAGACCTATCACTTTATCTCTAAGGAGTTGACAGCCGCT
TTCGTGCCCCCAAGGAGTCCTTAAATCTTACCGATCCGAAACAAACGTGTATTAAGGAT
GAATTTGAAAAATAATTAATGAAGTCTATATGTGATGATTATAATGATACATATAGCATG
AATGGTAGTTATCAAATTTTTAAGACTACGGGAGATTTGATTTTGATTTGGCAGCCTCTT
GTGCAAAAATCTCTTATGTTTCTTGGAGGTTTCGAAAAAATACGTAGGAGGCGAGAT
GTGGTGGATGTTAAGTCTAGGCATGACATTTTATGTGCAATTACAGTATCTTTATGAT
ACTTTGAAAGATTATATCAATGATGCATTGGGGAATTTGGCAGAATCTTGGTGTCTCGAT
CAAAAGCGGACGATAACGATGTTGCACGAACCTTAGTAAGATTAGTCCGTCGAGTATCGTG
TCTGAGGTTTACGGTCTCCGATATCTGCACAGTTGCATGGTGACGTGTTAGCTATCTCG
AAATGCATAGAGGTTAATCAGTCATCCGTTTCAGCTTCATAAGAGTATGCGGGTCGTTGAT
GCAAAGGGAGTAAGGAGTGAAACGATGTGTTATAATCGGCCTTTGGTGACGTTTAGCTTT
GTGAACTCGACGCCGTGAGGTTGTCCCTGGCCAGCTAGGGTTAGATAATGAAATTTCTATTG
GGTGATCATAGGACAGAGGAATGTGAAATACCTAGTACAAAGATCTTTTTATCTGGAAAT
CATGCGCACGTGTATACCGATTATACGCATACGAATTCGACGCCCATAGAAGACATTGAG
GTATTGGATGCTTTTTATTAGACTAAAAATCGATCCTCTTGAAAATGCCGATTTTAAAGTA
CTCGATTTATATTCCCGGACGAATTGAGTAGAGCAAACGTTTTTCGATCTAGAGAATATT
CTTCGTGAATATAACTCATATAAGAGCGCACTATATACTATAGAAGCTAAAAATTGCTACG
AATACACCTTCGTATGTTAATGGGATTAATTCTTTTTTACAAGGGCTTGGGGCTATAGGC
ACTGGATTGGGCTCGGTTATAAGTGTTACGGCAGGAGCGCTTGGGGATATTGTGGGTGGA
GTGGTGTCTTTTCTAAAAATCCATTCGGGGGTGGTCTCATGTTGATTTTAGCGATAGTA
GTTGTTGTGATAATAATTGTGGTTTTTCGTTAGACAAAGACATGTGCTTAGTAAGCCTATT
GACATGATGTTTCTTATGCCACCAATCCGGTGACTACTGTGTCCAGTGTTACGGGGACC
ACTGTGTCGAAGACACCTAGTGTTAAGGATGTTGACGGGGGCACATCTGTTGCGGTTTCG

GAAAAGGAGGAGGGTATGGCTGACGTCAGTGGGCAAGTAAGTGACGATGAATATTCACAA
GAAGATGCTTTAAAAATGCTCAAGGCCATAAAGTCCTTAGACGAGTCTACAGAAGAAAA
CCTTCGTCTTCTGAGTCTCATGCCCAAAACCTAGTTTTGATAGACAGGATCAGGTATAGA
GTTATAAGAGTGTAATGTAGAAGAAGCGTGA

>5814_U39

ATGAGCAAGATGGCAGTATTATTCCTGGTTGTCTTTTTGATGAATAGTGTTTTAATGATA
TATTGTGATCCGGATCATTATATCAGAGCGGGCTATAATCACAAATATCCTTTTCGGATT
TGTTTCGATTGCCAAAGGCACGGATTTAATGCGGTTTCGACAGAGATATTTTCGTGCTCGCCG
TATAAGTCTAATGCAAAGATGTCGGAGGGTTTTTTTCATCATTTACAAAACAAATATCGAG
ACCTACACTTTTCCAGTGAGAACGTATAAAAAAGAGTTGACGTTCCAAAGTAGTTACCGT
GATGTGGGTGTGGTTTATTTTCTGGATCGGACGGTGATGGGTTTGGCCATGCCGGTGTAC
GAAGCAAATTTAGTTAATTCTCATGCGCAGTGTATTCCGCCGTAGCGATGAAACGACCC
GATGGTACGGTGTTTAGTGCCTTTCATGAGGATAATAATAAAAAACAATACTCTAAATTTA
TTTCTCTGAATTTCAAATCTATAACTAATAAAAAGATTTATCACTACGAAAGAACCCTAC
TTTGCAAGGGTCCTTTGTGGCTATATTCTACATCAACGTCTCTCAATTGTATTGTGACG
GAGGCTACGGCTAAGGCGAAATACCCGTTTAGTTACTTTGCTTTGACGACTGGCGAAATC
GTGGAAGGGTCTCCGTTCTTCAACGGTTCAAACGGTAAACATTTTGCAGAGCCGTTAGAA
AAATTGACAATCTTGGAAAACATACTATGATAGAAGATCTAATGAATGGTATGAATGGG
GCCACTACGTTAGTGAGGAAAATCGCTTTTCTCGAGAAAAGCGGATACTTTGTTTTCTTGG
GAAATCAAGGAAGAGAATGAATCGGTGTGTATGCTAAAGCACTGGACCACGGTGACTCAC
GGCTTCGAGCGGAGACGAATGAGACCTATCACTTTATCTCTAAGGAGTTGACAGCCGCT
TTCGTGCCCCCAAGGAGTCCTTAAATCTTACCGATCCGAAACAAACGTGTATTAAGGAT
GAATTTGAAAAATAATTAATGAAGTCTATATGTCAGATTATAATGATACATATAGCATG
AATGGTAGTTATCAAATTTTTAAGACTACGGGAGATTTGATTTTGATTTGGCAGCCTCTT
GTGCAAAAATCTCTTATGTTTCTTGGAGCAGGGTTTCGAAAAAATACGTAGGAGGCGAGAT
GTGGTGGATGTTAAGTCTAGACATGACATTTTATGTACAATTACAGTATCTTTATGAT
ACTTTGAAAGATTATATCAATGATGCGTTGGGGAATTTGGCAGAATCTTGGTGTCTCGAT
CAAAGCGAACGATAACGATGTTGCACGAACTTAGTAAGATTAGTCCGTCGAGTATCGTG
TCTGAGGTTTACGGTTCGTCGGATATCTGCACAGTTGCATGGTGACGTGTTAGCTATCTCG
AAATGCATAGAGGTTAACCAGTCATCCGTTTCAGCTTCATAAGAGTATGCGGGTTCGTTGAT
GCAAAGGGAGTAAGGAGTAAACGATGTGTTATAATCGGCCTTTGGTGACGTTTAGCTTT
GTGAACTCGACGCCGAGGTTGTCCCTGGCCAGCTAGGGTTAGATAATGAAATTTCTGTTG
GGTGATCATAGGACAGAGGAATGCGAAATACCTAGTACAAAGATCTTTTTATCTGGAAAT
CATGCGCACGTGTATACCGATTATACGCATACGAATTCGACGCCCATAGAAGACATTGAG
GTATTGGATGCTTTTATTAGACTAAAAATCGATCCTCTTGGAAATGCCGATTTTAAAGTA
CTCGATTTATATTCGCCGGACGAATGAGTAGAGCAAACGTTTTTCGATCTAGAGAATATT
CTTCGTGAATATAACTCATATAAGAGCGCACTATATACTATAGAAGCTAAAAATTGCTACG
AATACACCTTCGTATGTTAATGGGATTAATTTCTTTTTTACAAGGGCTTGGGGCTATAGGC
ACTGGATTGGGCTCGGTTATAAGTGTACGGCAGGAGCGCTTGGGGATATTGTGGGTGGA
GTGGTGTCTTTTCTAAAAATCCATTCGGGGGTGGTCTCATGTTGATTTTAGCGATAGTA
GTTGTTGTGATAATAATTTGTGGTTTTTCGTTAGACAAAGACATGTGCTTAGTAAGCCTATT
GACATGATGTTTCCTTATGCCACCAATCCGGTGACTACTGTGTCCAGTGTACGGGGACC
ACTGTGCTCAAGACACCTAGTGTTAAGGATGTTGACGGGGGCACATCTGTTGCGGTTTCG
GAAAAGGAGGAGGGTATGGCTGACGTCAGTGGGCAAGTAAGTGACGATGAATATTCACAA
GAAGATGCTTTAAAAATGCTCAAGGCCATAAAGTCCTTAGACGAGTTCTACAGAAGAAAA
CCTTCGTCTTCTGAGTCTCATGCCCAAAACCTAGTTTTGATAGACAGGATCAGGTATAGA
GTTATAAGAGTGTAATGTAGAAGAAGCGTGA

U40

>2284_U40

ATGAATTCGTTACAATCTCTGTGTGTGTTGTGCGCTAGGCTTAATGAATGTGCTTTGGAC
CTTGAATGTTTAAAGTTTTGTGATCCTGTAATAGTATTGTCTGATATGGCAAATTTCAAA
AAGAATGGGATTGTGATTTCTGCATTTATATCAAACTTTTTTTTGAAGGTATCAAAGAACAA
AATCTTCTTTGTGCGTCAGCTTTGACGGTTTATATGCAGGTAAGTCTAAAAGCTATGTAT
GAACAGGTGCTTCTTTTGGATGCAGCGTTGGAAAGCTTTATGGTAGACCAGGATAGAAAA

AAATATTTTAAAAAGGTGTTGTGTTTGTGATGTTGTGCTGAGCACTTGTCAATAAACATT
TCATTGAATAACGGTGTGAGTTTATCGTGCAATTATCTACGCTAAATGATATAGAACAA
TTAATCAGTAAGATCAACTCCGTATATGCTTTACTTTTTGCCCAAGAAGGGTTGCAGATT
TGCGGAAAAATTATAGACTTGTAACTATCATGTGTGGTGCATGTATGGTTGCGAAACCG
GAGAGTTATCTCGAAACCAAGACTTGCATGAAGTGTATGAAGAATTGACTTTGACTCCA
AACCAAGGAAAAATCACTTCGAAGAAGATTACATGGGAAGTTTTGTAATCATTTGACCGAA
CAAAAAGCTTTTTTTTAAATATTGAAAAAACATTGAGACGATAGAGAAAGACTTAGGGGAG
GCTATTTTGAACATATGGGACTGTTTCAGTCTGTTGCGACTGAAATCAAAAAGATTTTTAAG
CAGCAGCGTTCAGCCGAGTCTGTTGCATGTTAGTGATGCTGAGAAAACCTCTGAAAAAGTAT
AATATTTTTTAGCAAGGTACCTGACGTGATATATTCTTTAAGTGAGTTTACGTATTGGTCT
AAAATATCGGAAACGATAGTGAGAAATGTGCGTATTACGCTTCAGCAGTTAAATAGTTGT
CATACGTTGTATAAGCAATTACAGAATGATGTTAGTCTTTATCTTTACGGTGAAGTGTCT
GAAGACTTCTTGGCTTTGAGTGAAAATTTGTTAACGCACGACGAGCGCTTGTACGTTGGC
TCCATTTATGTGTCTCCTTCGAGGCTGATAGATCTTGTCACTGGCCTGAGCATTAAAAAT
CTCGAGGAAAGTCCAATATTTAAAAGATTGGCTGAGGAAGATGAGGTTCAACATAAAATT
AAGAGTCTCTTGCATGACATTCGGGATCCACAAACGACAGAAACGCCGGGAAGGCTGAAC
ACGATCAATTGCATGCTTCAAACCTATAATCTTCAGCAGGAAGTCTTGGCCCGGAAAAAG
GCGTATTTTCAAAGGTTTCTGAATCTGGTTATAATAGGGTGATGGCATGTATTCGAGAG
CAGGAATCGCTCATCAATAAGGTAGTGAGTGTGAATGTGTATGGGAATTTTTATTTTTGAG
GCCTTATCGAAGATTATGAACGGTTTCGTTCTCAGAAAAATGTACTTGGATGGCTCTCTC
CGTGTAGATAGCTGTACATACGATGAACATCTTTATATTAAAAACAATCTGATGCCGAAG
AAGCTCCCTTTGGAGCTCTTGC CGGATTTAAGTGAGATTATGTACACATTGTTAACGGGT
CCGTTATCGGATTTTCATAAATCTGCCTATCCACTTCCGGCGAATATCTCAATGGCGTAT
GGTTGCGACCACGGGAGATGTTACCGCATATGAAGGAAGATTTAGCTAGATGCATAGAG
GGCACTATTCATCCTAGTGTGTGGATGGTGTGCGAATACAATGAATTTTTTAACTTTTCC
GGTGTAACTGATGTAATGACATGCAGAAAAAATGTGGAATTTTATCAGAGAGTTAACA
CTTCTGTTGCACTATATAATGATGTTTTTGGAAAGCGATTAAAAATTGTGCGTATAGAT
GAGGAGGGGATTTGAGTGGAATGTGTTTTGACATTTAATCACGAGTCTCCTTTATTA
TTTCACACTGGTGGTGGCATGACAAAGTTTTAAAGACGTGTATTTCGCTGTTATATTGTGAT
TTGCAAGCTCAGTTGTGCGAGAGAGACTTTAGACGTGCCGGAAGGAGTAAGTTACTCTGTT
CGAACGCCAATCTACTAGATCTCGTTAGGGAAAAATGAGCAAGATGGCAGTATTATTCCT
GGTTGTCTTTTTTGATGAATAG

>5055_U40

ATGAATTCGTTACAATCTCTGTGTGTGTTGTGCGCTAGGCTTAATGAATGTGCTTTGGAC
CTTGAATGTTTTAAAGTTTTGTGATCCTGTAATAGTATTGTCTGATATGGCAAATTTTAAA
AAGAATGGGATTTGTGATTCCTGCATTTATATCAAACTTTTTTTTGAAGGTATCAAAGAACAA
AATCTTCTTTGTGCGTACGCTTTGACGGTTTATATGCAGGTAAGTCTAAAAGCTATGTAT
GAACAGGTGCTTCTTTTGGATGCAGCGTTGGAAAGCTTTATGGTAGACCAGGATAGAAAA
AAATATTTTAAAAAGGTGTTGTGTTTGTGAGGCGTTGTGCTGAGCACTTGTCAATAAACATT
TCATTGAATAACGGTGTGAGTTTATCGTGCAATTATCTACGCTAAATGATATAGAACAA
TTAATCAGTAAGATCAACTCCGTATATGCTTTGCTTTTTGCCCAAGAAGGGTTGCAGATT
TGCGGAAAAATTATAGACTTGTAACTATCATGTGTGGTGCATGTATGGTTGCGAAACCG
GAGAGTTATCTCGAAACCAAGACTTGCATGAAGTGTACGAAGAATTGACTTTGACTCCA
AACCAAGGAAAAATCACTTCGAAGAAGATTACATGGGAAGTTTTGTAATCATTTGACCGAA
CAAAAAGCTTTTTTTTAAATATTGAAAAAACATTGAGACGATAGAGAAAGACTTAGGGAGAG
GCCATTTTGAACATATGGGACTGTTTCAGTCTGTTGCGGCTGAAATCAAAAAGATTTTTAAG
CAGCAGCGTTCAGCCGAGTCTGTTGCATGTTAGTGATGCTGAGAAAACCTCTGAAAAAGTAT
AATATTTTTTAGCAAGGTACCTGACGTGATATATTCTTTAAGTGAGTTTACGTATTGGTCT
AAAATATCGGAAACGATAGTAAGAAATGTGCGTATTACGCTTCAGCAGTTAAATAGTTGC
CATACGTTGTATAAGCAATTACAGAATGATGTGAGTCTTTATCTTTACGGTGAAGTGTCT
GAAGACTTCTTGGCTTTGAGTGAAAATTTGTTAACGCATGACGAGCGCTTGTACGTTGGC
TCCATTTATGTGTCTCCTTCGAGGCTGATAGATCTTGTCACTGGCTTGGCATTAAAAAT
CTCGAGGAAAGTCCAATATTTAAAAGATTGGCTGAGGAAGATGAGGTTCCAGCATAAAATT
AAGAGTCTCTTGCATGACATTCGGGATCCACAGACGACAGAAACACCGGGAAGGCTGAAC
ACGATCAATTGCATGCTTCAAACCTATAATCTTCAGCAGGAAGTATTGGCCCGGAAAAAG
GCGTATTTTCAAAGGTTTCTGAATCTGGTTATAATAGGGTGATGGCATGTATTCGAGAG
CAGGAATCGCTCATCAATAAGGTAGTGAGTGTGAATGTGTATGGGAATTTTTATTTTTGAG

GCCTTATCGAAGATTATGAACGGTTTTCGTTCTCAGAAAAATGTAAGTTGGATGGCTCTCTC
CGTGTAGATAGCTGTACATACGATGAACATCTTTATATTTAAAAACAATCTGATGCCGAAG
AAGCTCCCTTTGGAGCTCTTGCCGGATTTAAGTGAGATTATGTACACATTGTTAACGGGT
CCGTTATCGGATTTTCATAAATCTGCCTATCCACTTCCGGCGAATATCTCAATGGCGTAT
GGTTGCGACCACGCGGAGATGTTACCGCATATGAAGGAAGATTTAGCTAGATGCATAGAG
GGTACTATTTCATCCTAGTGTGTGGATGGTGTGCGAATACAATGAGTTTTTTAACTTTTCC
GGTGTAAGTATGTAATGACATGCAGAAAAAATGTGGAATTTTATCAGAGAGTTAACA
CTTCTGTTGCACTATATAATGATGTTTTTGGAAAGCGATTAAAAATTTGTGCGTATAGAT
GAGGAGGGGATTTGAGTGGAAATGTCGTTTTGACATTTAATCACGAGTCTCCTTTATTA
TTTCACACTGGTGGTGGCATGACAAAGTTTTAAAGACGTGTATTTCGCTGTTATATTGTGAT
TTGCAAGCTCAGTTGTGAGAGAGACTTTAGACGTGCCGGAAGGAGTAAGTTACTCTGTT
CGAGCGCTAATCTACTAGATCTCGTTAGGGAAAAATGAGCAAGATGGCAGTATTATTCCT
GGCTGTCTTTTTGATGAATAG

>5814_U40

ATGAATTCGTTACAATCTCTGTGTGTGTTGTGCGCTAGGCTTAATGAATGTGCTTTGGAC
CTTGAATGTTTAAAGTTTTGTGATCCTGTAATAGTATTGTCTGATATGGCAAATTTCAAAA
AAGAATGGGATTGTGATTCCTGCATTTATATCAAACTTTTTTTTGAAGGTATCAAAGAACAA
AATCTTCTTTGTGCGTCAGCTTTGACGGTTTTATATGCAGGTAAGTCTGTAAGCTATGTAT
GAACAGGTGCTTCTTTTGGATGCAGCGTTGGAAAGCTTTATGGTAGACCAGGATAGAAAA
AAATATTTTGAAGGTTGTTGTGTTTGGATGTTGTGCTGAGCACTTGTCAATAAACATT
TCATTGAATAACGGTGTGAGTTTATCGTGCAATTATCTACGCTAAATGATATAGAACAA
TTAATCAGTAAGATCAACTCCGTATATGCTTTACTTTTTGCCCAAGAAGGGTTGAGATT
TGCGGAAAAATTATAGACTTGTAACTATCATGTGTGGTGCATGTATGGTTGCGAAACCG
GAGAGTTATCTCGAAACCAAGACTTGCATGAAGTGTATGAAGAATTGACTTTGACTCCA
AACCAAGGAAAAATCACTTCGAAGAAGATTACATGGGAAGTTTTGTAATCATTTGACCGAA
CAAAAAGCTTTTTTTAATATTGAAAAAACATTGAGACGATAGAGAAAGACTTAGGGGAG
GCTATTTTGAACATATGGGACTGTTTCAGTCTGTTGCGACTGAAATCAAAAAGATTTTTAAG
CAGCAGCGTTCAGCCGAGTCTGTTGCATGTTAGTGATGCTGAGAAAACTCTGAAAAAGTAT
AATATTTTTTAGCAAGGTACCTGACGTGATATATTCTTTAAGTGAGTTTACGTATTGGTCT
AAAATATCGGAAACGATAGTGAGAAATGTGCGTATTACGCTTCAGCAGTTAAATAGTTGT
CATACGTTGTATAAGCAATTACAGAATGATGTTAGTCTTTATCTTTACGGTGAAGTGTCT
GAAGACTTCTTTGGCTTTGAGTGAATAATTTGTTAACGCACGACGAGCGCTTGTACGTTGGC
TCCATTTATGTGTCTCCTTCGAGGCTGATAGATCTTGTCACTGGCCTGAGCATTAAAAAT
CTCGAGGAAAGTCCAATATTTAAAAGATTGGCTGAGGAAGATGAGGTTCAACATAAAAAT
AAGAGTCTCTTGCATGACATTCGGGATCCACAAACGACAGAAACGCCGGAAGGCTGAAC
ACGATCAATTGCATGCTTCAAACCTATAATCTTCAGCAGGAAGTCTTGGCCCCGAAAAAG
GCGTATTTTCAAAGGTTTCTGAATCTGGTTATAATAGGGTGATGGCATGTATTCGAGAG
CAGGAATCGCTCATCAATAAGGTAGTGAGTGTGAATGTGTATGGGAATTTTATTTTTGAG
GCCTTATCGAAGATTATGAACGGTTTTCGTTCTCAGAAAAATGTAAGTTGGATGGCTCTCTC
CGTGTAGATAGCTGTACATACGATGAACATCTTTATATTTAAAAACAATCTGATGCCGAAG
AAGCTCCCTTTGGAGCTCTTGCCGGATTTAAGTGAGATTATGTACACATTGTTAACGGGT
CCGTTATCGGATTTTCATAAATCTGCCTATCCACTTCCGGCGAATATCTCAATGGCGTAT
GGTTGCGACCACGCGGAGATGTTACCGCATATGAAGGAAGATTTAGCTAGATGCATAGAG
GGCACTATTTCATCCTAGTGTGTGGATGGTGTGCGAATACAATGAATTTTTTAACTTTTCC
GGTGTAAGTATGTAATGACATGCAGAAAAAATGTGGAATTTTATCAGAGAGTTAACA
CTTCTGTTGCACTATATAATGATGTTTTTGGAAAGCGATTAAAAATTTGTGCGTATAGAT
GAGGAGGGGATTTGAGTGGAAATGTCGTTTTGACATTTAATCACGAGTCTCCTTTATTA
TTTCACACTGGTGGTGGCATGACAAAGTTTTAAAGACGTGTATTTCGCTGTTATATTGTGAT
TTGCAAGCTCAGTTGTGAGAGAGACTTTAGACGTGCCGGAAGGAGTAAGTTACTCTGTT
CGAACGCCAATCTACTAGATCTCGTTAGGGAAAAATGAGCAAGATGGCAGTATTATTCCT
GGTTGTCTTTTTGATGAATAG

U41

>2284_U41

ATGGCTGATGAAAACGAAACGGTAGTTTTCTGCTCCGTTTTCTACTGCTGCTTGGATCTAT

GTTTTCCAAAAGATAAAGAACTTTTGGATGTTTTATCTGTTTTATCGTTAATGGAGAGA
AATTCCTCCTGTAGTTATTTTCGCCGTTGTTAATGAATTTGACCGTAGAAAACGATTTCTCG
ACCACGGTTAAAACACCTATCACTAATTTTTGGAGGCACAATTCTAACTAAAATTACGTCT
TTTATGCCCGTGTGTTTTTTTTTTTCATGGAAGTGGAGCAGTTAGTTGGTATGGCCGAAGAT
CATGGTGATCTTATTTCGATTATGTGAGCAAACCAGACAAAAATTTCACTTGCAATCGTTT
GAGGTGCCCTACAGCACGCAAAGTATTCGATATCAAAGCTCTTTGTTCTGCTGTGCGGCAAA
GATGCAGATTCTGTTATTTGTTCATGTTGCGTGTGGAACGGTTTTCAAAGAGTTGTTGTTT
GCTGGTCTGTTAATTCCTTGTGTTGAAGAACAATTTCAAGTACAAGTTGGGGAGTATTCG
TGTGTGAAAATTCCTCTATATTCGCAACTTTATTTGAAACTGAAGAAACGATTTCTTTA
TCTTCCTGTACAGAATTTATTCAGGAACGAGGATTTTTTCTGCCTGCGTTAAGTGAAACG
CTTTTTTATTATGTTTTTACATCATGGGGTACAACATTACGATTTTCTAATACTAAGGAA
TTAATCGATGCTGGCCTGAAACAATTTACACAGGATGGGGAGCAAACGTCAAATTAGCA
CCCCATAAAACTTATCTGGGTATCTCCGGACAGAAAATTTAGCGGTGAGAAAAGATTTT
TTGATGTTGGTTGATTCCGTTGTGACAGAATTATCGTTTAGTCATGTAGCCGAATATCTG
GATTCGGTATATGACCCTAGTCAGATAATGAATTTAATGACTGGCCAATCATAAGAAAC
TCAGAGACACACGCTGAACGCATGGCGCAGCTCACGAATTTAAAGTTACATTTGTCTAGT
CATCTTGCTGTTTTGATCTTTGCCCAAATTCATCTTATATTGCTCAAATTTGGCATT
ATACCAAATGTGAAACAAGCGTTAATTCGGTGATGACTCAGGAATTAAGTGTGAGATCT
TTGCTTTTTTGTAAATGCGTTGTGCTCTTTGAGTGAGGATGTGTATAACGATAATAGAAA
ATTATCAAATGTGATTCGACATCAGGCAAAGATGACAAATTTTCACTAATCACTTTGGCT
TATGCCCTGTGCGACTTCTCCACAGCTGCTTTCTTATGTAGTGTGGAATTTAAATCGTATG
AGCGTTTATAACGCGGGTAAATGCGCATACTGAGATTTACAATCAATTTGGTAAATTTGTCG
GCGAATTTATGTGAATTCGTGATGGGAAATGTTGTCAATCGTGTATTGGAACGTCCATG
GTGCGGGTCCGTTACTCGTTTTGCCGGCAATACCAAAGAATGTTAAAAAGGAACCACTTGT
ATGAGCATGTTTTCTCGGTATTATGCCGAGGTTGATATATTAGGCTCTTTTGGCAGGAAG
CCTGTCAGCGAGCTCAAAGAGATTGAAAAGATCAACAAAACACGCTCTCACTGGACAGG
GGTAAATTCGTATCTCAGATCTTCGATTACTGTAAAAAAATTTCTCTGATTGATCCTGTA
ACAGGCGAAGATACTTTCAATGTTGAAAGTAAAAAAGATTTTGTCTATAATTCACGGA
TTAACACAGTGTATCGAGGAATGTGTTTCTAGATGCATTGTTGAAATGAGAAGGACACAA
ACGCCTAGAGAACAGATAGAAAATTTGTCTACAGTCATTTAATGTAGATACGACCCCTTAT
GCAACTGCCTTTTCTCCTTTTTTACGTTTTTTCATATTATAAGGTCATTTTAAACGGTTTTG
CAGAACCCTCGATTGATAGTGGCTTCAGGTCATGTTGTGATAGACCTTGTACCGGTAAT
TCAATTTCTAAATGGCTCGTTACAGCAGTACCAGTCACTATACGGAACGTTTACAGCAGC
TATCTCAAAAAGGTTTTCTGAATACGAGAACGGTTAAAGTGCCTTCAAATGTTGATATG
GAGCAAATCTTAGATTGCGATCTATATAAATCGGGTAAATATGTGAAAACCTACGATTGAG
GCGAAACTTTGCCGTTTGTCTATGCAGTGTCTTCGAGATTTTAGAATAAAAAATAGACCG
TTCAACAAATCCAGTAAGACAGCACACAACAATCCTTATTTTTAAAAAAGCTGAAACAT
AAGAAGAATCCCTTGTCTGGCTGTATCTCCTTTCTTTTATTTAAATATCATGATAAATTTG
TTTCCGAATGTGAAGATTTTCGTGCTTAGAGCTATGGCAACGTTTTCTGCTCAACAATGTT
CCGAAAACTTTAGATATTGGTAACCCGAAGAAGTTAAACTTTTTATTAATTTTGCATTC
AGTATTACGAATACCTATGATGAGATAGATATTATCGACATACAGCCAGAGTGTCTCTCA
ACTTTTATAGATTGTTATTTTTCATAACAAATTTTTGTGCGCTTTAGGTTTTTCATGATTAC
TTGACTAGTTTACATGGTCTTACATCCAAGCTGGTTACGCAGAATCCTGTGTTGTTTCCC
GTGGTGTAGACAAGCAGCCTAAATTTTTCTTCTATTTCAGGAATATTTAGTTTTATGTTAAA
AAGTTAGTTTTGGATGGCGTTCCGAATCCTGTGATTGCGTCACTGTCAAAGAACCCAAT
TTTGGTACAATTTTACAGAGTGCATCTTTGGTTACATTTGGTTTTGACATTAGAGAAATTT
GTGAGTTTGGCTAATAGGGAATATTTTTCAGTTTGGCCAGTTAGGTTGGATAGGCGGAAGT
GGCGTGGATAGGAACTTAAACCCCTACCTCTTCTGCATTGCAGGATTTTAGATTTATGAGA
CAGAAAACGATTATTGCTACAAAGTTTTTTCAGAAGTTATAGTAAAAAAGTTTCGGCGTGAG
GCAATTATGTTTGTACGGAAGTAGTTAAGGGCAAGGTTCTCAGCATTGTTGAGAATTTA
ACTAATGATATTGATCCAGAATTTAATAAATAGCTGAGGTGATGAGGGATAGGGAAGAC
AAGCCGACTATGGATGATATGCTTTTTTTTTGTGGATGGTAGGGAAGCCTTAGCGGCATCC
ATAATGCTTAAATTTGAATCACCTTAGTTGATATGAATGTGAGAGACTTTTCAATAACAAAT
CTGCAGTCCGTTATTCGAGGCGGTGTCTTCAAACGATGCTCCGGTCTATGACTTTTCTGAA
ATTCTAGCGGAGGAGGACGATCAAGGAAACGGTGTATTAATTTGATGAAACGGAACCC
GAAACGGATGAGCCGATGACTAAGAAAAACCGATTATAA

>5055_U41

ATGGCTGATGAAAACGAAACGGTAGTTTCTGCTCCCGTTTCTACTGCTGCTTGGATCTAT
GTTTTCCAAAAGATAAAGAACTTTTGGATGTTTTATCTGTTTTATCGTTAATGGAGAGA
AATTCCTCTGTAGTTATTTTCGCCGTTGTTAATGAATTTGACTGTAGAAAACGATTTCTCG
ACCACGGTTAAAACACCTATCACTAATTTTTGGAGGCACAATTTCTAACTAAAATTACGTCT
TTTATGCCCGTGTGTTTTTTTTTTTTCATGGAACCTGAGCAGTTAGTTGGTATGGCCGAAGAT
CATGGTGATCTTATTTCGATTATGTGAGCAAACCAGACAGAAATTTCACTTGCAGTCTGTTT
GAGGTGCCCTACAGCACGCAAAGTAATCGATATCAAAGCTCTTTGTTCTGCTGTCTGGCAAAA
GATGCAGATTCGTATTATTTGTTCATGTTGCGTGTGGAAACGGTTTTCAAAGAGTTGTTGTTT
GCTGGTCTGTTAATTCCTTGTGTTGAAGAACAATTTCAAGTACAAGTTGGGGAGTATTTCG
TGTGTGAAAATTCCTCTATATTCGCCAACTTTTATTTGAAACTGAAGAAACGATTTCTTTA
TCTTCTCTGTACAGAATTTATTCAGGAACGAGCATTTTTTTCTGCCTGCGTTAAGTGAGACG
CTTTTTTATTATGTTTTTACATCGTGGGGTACAACATTACGATTTTTCTAATACTAAGGAA
TTAATCGATGCTGGCCTGAAACAATTTACACAGGATGGGGAGCAAACCTGTCAAATTAGCA
CCCCATAAAACTTATCTGAGTATCTCCGGACAGAAAATTTTCAGCGGTTCGAGAAAGATTTT
TTGATGTTGGTTGATTCCGTTGTGACAGAATTATCGTTTAGTCATGTAGCCGAATATCTG
GATTCGGTGTATGACCCTAGTCAGATAATGAATTTAATGAGTGGCCAATCATAAGAAAC
TCAGAGACACACGCTGAACGCATGGCGCAGCTTACGAATTTAAAGTTACATCTGTCTAGT
CATCTTGCTGTTTTGATCTTTGCCCAAATTCATCTTATATTGCTCAAATTTGGCATT
ATACCAAACGTGAAACAAGCGTTTAATTCGGTGTACTCAGGAGTTACTGCTGAGATCT
TTGTCTTTTTGTAAATGCGTTGTCTGCTTTTACTGAGGATGTGTATAACGATAATAGAAAA
ATTATCAAATGTGATTCGACATCAGGCAAAGATGACAAATTTTCAGCTAATCACTTGGCT
TATGCCCTGTGCGACTTCTCCACAGCTGCTTTCTTATGTAGTGTGGAATTTAAATCGTATG
AGCGTTTATAACGCGGGTAATGCGCATACTGAGATTTACAATCATTTGGTAAATTTGTTTCG
GCGAATTTATGTGAATTCGTGATGGGAAATGTTGTCAATCGTGTATTGGAACCTGCCATG
GTGCGGGTCCGTTACTCGTTTGC CGCAATACCAAAGAATGTTAAAAAGGAACCACTTGTT
ATGAGCATGTTTTCTCGGTATTATGCCGAGGTTGATATATTAGGCTCTTTTGGCAGGAAG
CCTGTCAGCGAGCTCAAAGAGATTGGAAAAGATCAACAAAACACGCTCTCACTGGACAGG
GGTAAATTCGTATCTCAGATCTTCGATTACTGTAAAAAAATTTCTCTGATTGATCTCTGTA
ACAGGCGAAGATACTTTCAATGTTCGAAGTAAAAAAGATTTTTGTTTCTATAATTCACGGA
TTAACACAGTGTATCGAGGAATGTGTTTCTAGATGCATTGTTGAAATGAGAAGGACACAA
ACGCCTAGAGAACAGATAGAAAATTTGTCTACAGTCATTTAATGTAGATACGACCCCTTAT
GCAACTGCCTTTTCTCCTTTTTTGTACGTTTTTCATATTATAAGGTCATTTTAAACGGTTTTG
CAGAACCCTCGATTGATAGTGGCTTACAGGTCATGTTGTCTGATAGACCTTGTACCGGTAAT
TCAATTTCTAAATGGCTCGTTCAGCAGTACCAGTCACTATACGGAACGTTTTCACAGCAGC
TATCTCAAAAAGGTTTTCTGAATACGAGAACGGTTAAAGTTCGCTTCAAATGTTGATATG
GAGCAAATCTTAGATTGCGATCTATATAAATCGGGTAAATATGTGAAAACCTACGATTGAG
GCGAAACTTTGCCGTTTTGTCTATGCAATGTCTTCGAGATTTTAGAATAAAAAATAGACCG
TTCAACAAGTCTAGTAAGACAGCACACAACAATCCTTATTTTTAAAAAAGCTGAAACAT
AAGAAGAATCCCTTGTCTGGCTGTATCTCCTTTCTTTTATTTAAATATCATGATAAATTTG
TTTCCGAATGTGAAGATTTTCGTGCTTAGAGCTATGGCAACGTTTTCTGCTCAACAATGTT
CCGAAAACTTTAGATATTGGTAACCCCGAAGAAGTTAAACTTTTTATTAATTTGCATTC
AGTATTACGAATACCTATGATGAGATAGATATTATCGACATACAGCCAGAGTGTCTCTCA
ACTTTTATAGATTGTTATTTTTCATAACAAATTTTTGTCTGGCTTTAGGTTTTTCATGACTAC
TTGACTAGTTTACATGGTCTTACATCTAAGCTGGTTACGCAGAATCCTGTGTTGTTTCCC
GTGGTGTTAGACAAGCATCCTAAATTTTTCTTCTATTTCAGGAATATTTAGTTTTATGTTAAA
AAGTTAGTTTTGGATGGCGTTCGAATCCTGTGATTGCGTCACTGTCCAAAGAACCCAAT
TTTGGTACAATTTTTACAAGTCGATCTTTGGTTACATTTGGTTTTGACATTAGAGAAATTT
GTGAGTTTGGCTAATAGGGAATATTTTCAGTTTGGCCAGTTAGGTTGGATAGGCGGAAGT
GGCGTGGATAGGAACTTAAACCCCTACCTCTTCTGCATTGCAGGATTTTAGATTTATGAGA
CAGAAAACGATTATTGCTACAAAGTTTTTCAGAAGTTATAGTAAAAAAGTTTCGGCGTGAG
GCAATTATGTTTTGATACGGAAGTAGTTAAGGGCAAGGTTCTCAGCATTGTTGAGAATTTA
ACTAACGATATTGATCCAGAATTTAATAAATAGCTGAGGTGATGAGGGATAGGGAAGAT
AAGCCGACTATGGATGATATGCTTTTTTTTTGTGGATGGTAGGGAAGCCTTAGCGGCATCC
ATAATGCTTAAATTTGAATCACCTTAGTTGATATGAATGTGAAAGACTTTTCAATAACAAAT
CTGCAGTCCGTTATTCGAGACGGTGTCTTCAAACGATGCTCCGGTCTATGACTTTTCTGAA
ATTCTAGCGGAGGAGGACGATCAAGGAAACGGTGTATTAATGTTGATGAAACGGAACCC
GAAACGGATGAGCCGATGACTAAGAAAAACCGATTATAA

>5814_U41

ATGGCTGATGAAAACGAAACGGTAGTTTCTGCTCCCGTTTCTACTGCTGCTTGGATCTAT
GTTTTCCAAAAGATAAAGAACTTTTGGATGTTTTATCTGTTTTATCGTTAATGGAGAGA
AATTCCTCCTGTAGTTATTTCCGCCGTTGTTAATGAATTTGACCGTAGAAAACGATTTCTCG
ACCACGGTTAAAACACCTATCACTAATTTTTGGAGGCACAATTTCTAACTAAAATTACGTCT
TTTATGCCCGTGTGTTTTTTTTTTTCATGGAACCTGAGCAGTTAGTTGGTATGGCCGAAGAT
CATGGTGATCTTATTTCGATTATGTGAGCAAACCAGACAAAAATTTCACTTGCAATCGTTT
GAGGTGCCCTACAGCACGCAAAGTATTCGATATCAAAGCTCTTTGTTCTGCTGTGCGGCAAA
GATGCAGATTCCTGTTATTTGTTCATGTTGCGTGTGGAAACGGTTTCAAAGAGTTGTTGTTT
GCTGGTCTGTTAATTCCTTGTGTTGAAGAACAATTTCAAGTACAAGTTGGGGAGTATTTCG
TGTGTGAAAATTCCTCTATATTCGCCAACTTTTATTTGAAACTGAAGAAACGATTTCTTTA
TCTTCTGTACAGAATTTATTCAGGAACGAGGATTTTTTCTGCCTGCGTTAAGTGAAACG
CTTTTTTATTATGTTTTTACATCATGGGGTACAACATTACGATTTTTCTAATACTAAGGAA
TTAATCGATGCTGGCCTGAAACAATTTACACAGGATGGGGAGCAAACCTGTCAAATTAGCA
CCCATAAAAACTTATCTGGGTATCTCCGGACAGAAAATTTAGCGGTCGAGAAAGATTTT
TTGATGTTGGTTGATTCCGTTGTGACAGAATTATCGTTTAGTCATGTAGCCGAATATCTG
GATTCGGTATATGACCCTAGTCAGATAATGAATTTTAATGACTGGCCAATCATAAGAAAC
TCAGAGACACACGCTGAACGCATGGCGCAGCTCACGAATTTAAAGTTACATTTGTCTAGT
CATCTTGCTGTTTTGATCTTTGCCCAAATTCATCTTATATTGCTCAAATTTGGCATT
ATACCAAATGTGAAACAAGCGTTTAATTCGGTGATGACTCAGGAATTAAGTCTGAGATCT
TTGTCTTTTTGTAAATGCGTTGTGCTCTTTGAGTGAGGATGTGTATAACGATAATAGAAAA
ATTATCAAATGTGATTCGACATCAGGCAAAGATGACAAATTTTTCAGCTAATCACTTGGCT
TATGCCCTGTGCGACTTCTCCACAGCTGCTTTCTTATGTAGTGTGGAATTTAAATCGTATG
AGCGTTTATAACGCGGGTAATGCGCATACTGAGATTTACAATCATTTGGTAAATTTGTTTCG
GCGAATTTATGTGAATTCGTGATGGGAAATGTTGTCAATCGTGTATTGGAACCTGCCATG
GTGCGGGTCGGTACTCGTTTCCCGCAATACCAAAGAATGTTAAAAAGGAACCACTTGTT
ATGAGCATGTTTTCTCGGTATTATGCCGAGGTTGATATATTAGGCTCTTTTGGCAGGAAG
CCTGTCAGCGAGCTCAAAGAGATTGGAAGGATCAACAAAACACGCTCTCACTGGACAGG
GGTAAATTCGTATCTCAGATCTTCGATTACTGTAAAAAAATTTCTCTGATTGATCCTGTA
ACAGGCGAAGATACTTTCAATGTTTCAAGTAAAAAAGATTTTTGTTTCTATAATTCACGGA
TTAACACAGTGTATCGAGGAATGTGTTTCTAGATGCATTGTTGAAATGAGAAGGACACAA
ACGCCTAGAGAACAGATAGAAAATTTGTCTACAGTCATTTAATGTAGATACGACCCCTTAT
GCAACTGCCTTTTCTCCTTTTTTACGTTTTTTCATATTATAAGGTCATTTTAAACGGTTTTG
CAGAACCCTCGATTGATAGTGGCTTCAGGTCATGTTGTGATAGACCTTGTACCGGTAAT
TCAATTTCTAAATGGCTCGTTCAGCAGTACCAGTCACTATACGGAACGTTTTCACAGCAGC
TATCTCAAAAAGGTTTTCTGAATACGAGAACGGTTAAAGTTCGCTTCAAATGTTGATATG
GAGCAAATCTTAGATTGCGATCTATATAAATCGGGTAAATATGTGAAAACCTACGATTACAG
GCGAAACTTTGCCGTTTGTCTATGCAGTGTCTTCGAGATTTTAGAATAAAAAATAGACCG
TTCAACAAATCCAGTAAGACAGCACACAACAATCCTTATTTTTAAAAAAACGTGAAACAT
AAGAAGAATCCCTTGTCTGGCTGTATCTCCTTTCTTTTATTTAAATATCATGATAAATTTG
TTTCCGAATGTGAAGATTTCTGCTTAGAGCTATGGCAACGTTTTCTGCTCAACAATGTT
CCGAAAACTTTAGATATTGGTAACCCCGAAGAAGTTAAACTTTTTATTAATTTGCATTC
AGTATTACGAATACCTATGATGAGATAGATATTATCGACATACAGCCAGAGTGTCTCTCA
ACTTTTATAGATTGTTATTTTTCATAACAAATTTTTGTGCGCTTTAGGTTTTTCATGATTAC
TTGACTAGTTTACATGGTCTTACATCCAAGCTGGTTACGCAGAATCCTGTGTTGTTCCC
GTGGTGTAGACAAGCAGCCTAAATTTTTCTTCTATTTCAGGAATATTTAGTTTTATGTTAAA
AAGTTAGTTTTGGATGGCGTTCCGAATCCTGTGATTGCGTCACTGTCAAAGAACCCAAT
TTTGGTACAATTTTTCAGAGTGCATCTTTGGTTACATTTGGTTTTGACATTAGAGAAATTT
GTGAGTTTGGCTAATAGGGAATATTTTTCAGTTTGGCCAGTTAGGTTGGATAGGCGGAAGT
GGCGTGGATAGGAACTTAAACCCCTACCTCTTCTGCATTGCAGGATTTTATGATTTATGAGA
CAGAAAACGATTATTGCTACAAAGTTTTTTCAGAAGTTATAGTAAAAAGGTTTCGGCGTGAG
GCAATTATGTTTTGATACGGAAGTAGTTAAGGGCAAGGTTCTCAGCATTGTTGAGAATTTA
ACTAATGATATTGATCCAGAATTTAATAAATAGCTGAGGTGATGAGGGATAGGGAAGAC
AAGCCGACTATGGATGATATGCTTTTTTTTTGTGGATGGTAGGGAAGCCTTAGCGGCATCC
ATAATGCTTAAATTTGAATCACCTTAGTTGATATGAATGTGAGAGACTTTTCAATAACAAAT
CTGCAGTCCGTTATTCGAGGCGGTGTCTTCAAACGATGCTCCGGTCTATGACTTTTCTGAA
ATTCTAGCGGAGGAGGACGATCAAGGAAACGGTGTATTAATGTTGATGAAACGGAACCC
GAAACGGATGAGCCGATGACTAAGAAAAACCGATTATAA

U42

>2284_U42

ATGTATCCTCGTGGAGTAAAACGCTCACATCATGATTATCATAGACAAACCGCCTTTAGA
ACGATAAAAACGTTCCACGCATAGACAAACCAGCAAGTTCATCAGCCATTTTCGCCAAAAAT
TTTCGAGGAAAACTCGCACCTCTGAAACAACCTGGATGAATCCAGACTAGACGCACTAAGT
CTGACAGAGTTAGAACAGCTGAAGACCATCATCGAAGAGAAAACAACAAGAAAAACGCGCT
CAAAACAATGCGATCACCTTTCTGCCTAATCTTCCAACAGTCCCCTTTGCCGACACAAAAC
TTTTTCGCTAAAATCTCTCGGCCCTGAGACCTTACAACGGAGACGCCCGTGATCCAAAACAA
CGCATCCGAGACAGATTCCCACAGACTCACGAACGAATTTGTCTCCTGACTAATGATATA
CTGGAAACCGACCTACTGCTTCGCTACCGACAATGTCTAGACTCGCTCACACGAGAAGAA
AATCAGCAACTGATGGGCGACCCGGATATTCTCCCTGACCAACTCACCTTGTCTTGCTTTT
ACAGTTGCTACAGTAGAAGAAGCCTGCAGTTATTTTTAAATTTTCAGATCTTCACAACCTG
CCGGTCAACCCACAAGACTTATTCATGTATAACCATCACCGTCATGAAATTCGAATTTTTTC
AACAACTCAACATGGCAAAATTAACCTGTGTTTTTAACGACAACGGACACGGTGACATC
GAATACCGAAAACTCAGGCAACTCTGCGGTAAACCTGTACTAGATCGAGAGATGCCAAAC
TCCGAGCTTGAAGTACAACAACAACTCCAGACTCGTTTAGACACCCAATTCAACAAGCC
ATGTCATCGTTGTTACGTTTGCTCGAATACTACGACAAATTAAGAGCAAATCATACAA
ACAAAGAAACCGCAATTCATCCGCGATTTTCGACACAGGACGGGTAGCCGAGCGATACGAA
TGCGGTCTGATGAGCAGACTCATCGGAAAACAATTCAGTAACCATAAATGCGATGATGTT
AGTTGTCAGAATAGAATCGAACGAATCATGGCTCCTTGAAACCGAGTCTATTTTTTTTGC
ACTTATTTTCGCCAAAGACGCACCAAAGTTTAAGCTATTTCCAAATTTTCTGAAGAATAC
CGAAATCTTTCTTTTACTTGTCCAGAGGTCGACACGGAACTTCATGTAGCTATTCAACA
AATCAGGACCTACCTCAGACCTCTCATCGTTTACATAAAAACCATGGTACGCCTAAAGTC
AAGTCCAAAGTTTGTGTCGAAAAACCGGACACTTCAATCCTAACAACGACAAAAACAAC
ACCGAAATTCATAGAAGAATCAATGGAAACCGACAACAAAATACCAAATCCTAGAGAA
CTAAACTTCAATCAAGCAAAAACAAGAAGAAATTGTGATTATTAACATTAATGAAAACGTC
AACTCTAAACATGAAAGTGAATCGAGTGTGAAATGGACTTAGATTTAGACTATGAAGCT
GACACCTGTGAAACTAACTTAAACGCATGTTCTTCTGACTCAGAATAA

>5055_U42

ATGTATCCTCGTGGAGTAAAACGCTCACATCATGATTATCATAGACAAACCGCCTTTAGA
ACGATAAAAACGTTCCACGCATAGACAAACCAGCAAGTTCATCAGCCATTTTCGCCAAAAAT
TTTCGAGGAAAACTCGCACCTCTGAAACAACCTGGATGAATCCAGACTAGACGCACTAAGT
CTGACAGAGTTAGAACAGCTGAAGACCATCATCGAAGAGAAAACAACAAGAAAAACGCGCT
CAAAACAATGCGATCACCTTTCTGCCTAATCTTCCAACAGTCCCCTTTGCCGACACAAAAC
TTTTTCGCTAAAATCTCTCGGCCCTGAGACCTTACAACGGAGACGCCCGTGATCCAAAACAA
CGCATCCGAGACAGATTCCCACAGACTCACGAACGAATTTGTCTCCTGACTAATGATATA
CTGGAAACCGACCTACTGCTTCGCTACCGACAATGTCTAGACTCGCTCACACGAGAAGAA
AATCAGCAACTGATGGGCGACCCGGATATTCTCCCTGACCAACTCACCTTGTCTTGCTTTT
ACAGTTGCTACAGTAGAAGAAGCCTGCAGTTATTTTTAAATTTTCAGATCTTCACAACCTG
CCGGTCAACCCACAAGACTTATTCATGTATAACCATCACCGTCATGAAATTCGAATTTTTTC
AACAACTCAACATGGCAAAATTAACCTGTGTTTTTAACGACAACGGACACGGTGACATC
GAATACCGAAAACTCAGGCAACTCTGCGGTAAACCTGTACTAGATCGAGAGATGCCAAAC
TCCGAGCTTGAAGTACAACAACAACTCCAGACTCGTTTAGACACCCAATTCAACAAGCC
ATGTCATCGTTGTTACGTTTGCTCGAATACTACGACAAATTAAGAGCAAATCATACAA
ACAAAGAAACCGCAATTCATCCGCGATTTTCGACACAGGACGGGTAGCCGAGCGATACGAA
TGCGGTCTGATGAGCAGACTCATCGGAAAACAATTCAGTAACCATAAATGCGATGATGTC
AGTTGTCAGAATAGAATCGAACGAATCATGGCTCCTTGAAACCGAGTCTATTTTTTTTGC
ACTTATTTTCGCCAAAGACGCACCAAAGTTTAAGCTATTTCCAAATTTTCTGAAGAATAC
CAAAATCTTTCTTTTACTTGTCCAAAGGTCGACACGGAACTTCATGTAGCTATTCAACA
AATCAGGACCTACCTCAGACCTCTCATCGTTTACACAAAAACCATGGTACGCCTAAAGTC
AAGTCCAAAGTTTGTGTCGAAAAACCGGACACTTCAATCCTAACAACGACAAAAACAAC
ACCGAAATTCATAGAAGAATCAATGGAAACCGACAACAAAATACCAAATCCTAGAGAA
CTAAACTTCAATCAAGCAAAAACAAGAAGAAATTGTGATTATTAACATTAATGAAAACGTC
AACTCTAAACATGAAAGTGAATCGAGTGTGAAATGGACTTAGATTTAGACTATGAAGCT

GACACCTGTGAAACTAACTTAAACGCATGTTCTTCTGACTCAGAATAA

>5814_U42

ATGTATCCTCGTGGAGTAAAACGCTCACATCATGATTATCATAGACAAAACCGCCTTTAGA
ACGATAAAAACGTTCCACGCATAGACAAAACAGCAAGTTCATCAGCCATTTTCGCCAAAAAT
TTTCGAGGAAAACTCGCACCTCTGAAACAACCTGGATGAATCCAGACTAGACGCACTAAGT
CTGACAGAGTTAGAACAGCTGAAGACCATCATCGAAGAGAAAACAACAAGAAAAACGCGCT
CAAAACAATGCGATCACCTTTCTGCCTAATCTTCCAACAGTCCCCTTTGCCGACACAAAAC
TTTTTCGCTAAAATCTCTCGGCCCTGAGACCTTACAACGGAGACGCCCGTGATCCAAAACAA
CGCATCCGAGACAGATTCCCACAGACTCACGAACGAATTTGTCTCCTGACTAATGATATA
CTGGAAAACCGACCTACTGCTTCGCTACCGACAATGTCTAGACTCGCTCACACGAGAAGAA
AATCAGCAACTGATGGGCGACCGGATATTCTCCCTGACCAACTCACCTTGTCTTGCTTTT
ACAGTTGCTACAGTAGAAGAAGCCTGCAGTTATTTTTAAATTTTACGATCTTTCACAACCTG
CCGGTCAACCCACAAGACTTATTCATGTATACCATCACCGTCATGAAATTCGAATTTTTTC
AACAAACTCAACATGGCAAAATTAACCTGTGTTTTTAACGACAACGGACACGGTGACATC
GAATACCGAAAACCTCAGGCAACTCTGCGGTAACCTGTACTAGATCGAGAGATGCCAAAC
TCCGAGCTTGAAGTACAACAACAACCTCCAGACTCGTTTAGACACCCAATTAACAAGCC
ATGTCCATCGTTGTACGTTTGTCTCGAATACTACGACAAATTAAGAGCAAATCATACAA
ACAAAGAAAACCGCAATTCATCCGCGATTTTCGACACAGGACGGGTAGCCGAGCGATACGAA
TGCGGTCTGATGAGCAGACTCATCGGAAAACAATTCAGTAACCATAAATGCGATGATGTT
AGTTGTCAGAATAGAATCGAACGAATCATGGCTCCTTGGAAAACCGAGTCTATTTTTTTGC
ACTTATTTTCGCCAAAGACGCACCAAAGTTTTAAGCTATTTCCAAATTTTCTGAAGAATAC
CGAAATCTTTCTTTTACTTGTCCAGAGGTGACACGGAACTTCATGTAGCTATTCAACA
AATCAGGACCTACCTCAGACCTCTCATCGTTTACATAAAAACCATGGTACGCCTAAAGTC
AAGTCCAAAGTTTGTGTGCAAAAACCGGACACTTCAATCCTAACAACGACAAAAACAAC
ACCGAAATTCATAGAAGAATCAATGGAACCGACAACAATAACCAAATCCTAGAGAA
CTAAACTTCAATCAAGCAAAAACAAGAAGAAATTTGTGATTATTAACATTAATGAAAACGTC
AACTCTAAACATGAAAGTGAATCGAGTGTGAAATGGACTTAGATTTAGACTATGAAGCT
GACACCTGTGAAACTAACTTAAACGCATGTTCTTCTGACTCAGAATAA

U43

>2284_U43

ATGACCATCGTAGTGTGTTGCGACTGAATACGATGCCGCCAACGTCATTTTTAGTATCCTA
TGTTAGAAGTCCCAGTGAACATTTAATCTTTCCGATTATTGTTAAATATAAACCATCAAAT
AACATTAGTTTCTGCCCTTCAAACCTCAAAAATGCAAAAATTCAAAGCGAATCGACACGGTA
TTTGTCTGCCATGCCGAAAAATTAATCTCAGTCACTACATACAAAACGCAAGTCCGATA
AAGGCTGAAGATGTCGCCAACAGCCTCAACGACAAAAGAAACCGAATCGTTATATGTCGAC
ATGATATTATCTCAAACCTGAAAAGAAAGAGAGGACGTCGAATTTAAATACATGGCCTAT
TTCCACAAATCGCTCATTATCAAATATCTCACCGGTAATTTCTATTGCCCACTTCCCCC
TTTTGGTTTTTTATCGACCTACGGGCAACAGAAGGTATGCTTTTTATTAACAATGTTTTAC
TATTTGTTTCGAGGAACAAAAGAGTACCATCACGACTACCAAAAATTTATGTTCAATGCTTC
ACGGAAAACACCGGAAACATGGTGTTCACGTATTCTTCCATGTCCGAATTTATAAACATC
ACTTTAAAATCAAAAATTTAGAAAACATTTTGCAGACTTCTCAACGTACGCCAGACAAAAA
AATCTTCGCGATAAGGAAGAATTTAAACATCTAGACACTCAAATCAATCTATTTAGAAAAG
AGTTCACATCTAACCAACACATTCGCGTCCATTACATATATATCGCTTACAACACGGCC
CTAGAGACAACAAAATTCGTCAATTACTGCAATCTCACTTCTTACGACTCTAATTTACCA
ATAGGGCAACAATGTCAGAGAAATCTCCACATTTCTGGAAATTCCTGCACGAAAATTTA
CTGTGTATCATGAAGCAATACTTTAATGCGGATTGTTACTTTAAAACATACATAGACATA
AAAAGATTAAAAAATCCAAATCTAAACGTTACAGAATACGAGTATGCTCTGGTCTCTAAA
AAGAAGACCTCCAAGCATTAACCTCCGAACAAAATAACGCGTGCCATAGCGAAATGTAAT
AAAAACGGAGAAGGTCTTTTTTCTCCAGTAAAACAGGATTACAAGGATTATTAGAAAATC
TCCGCTTCCGATAGATACGTTTCCAGATACAAGATAAGCGCATCTACAGAAGACAACACCTA
CACAAAGATTATCACCGTCCCTTTCTGTATTTCAGGGTGCAGCTTTTACATAAAAAATATT
TTTTGCTTCGAAAACCTGAGGATTGGTACGAAAACATGGGATTTAACCGCATCTTACAG
TATCTCCCCGACGAATATATCTCTGATGAAGCTCTAACTCGCGCCATCTGGCTGCAAGAC
ACACATTTTCTCTGCGACGACGTCGAAAACAATTTCTATACCACACGACACGAAATATTC

AACGAACGCATACCAGTCACAAATTACATAGGAGACTTGGATCTTCCGCTGCAAGATACA
GCAAATATTACGGAGGAAACATTTTTCTCCATGTGTGCGCTCATAAGACTTACCCTCATC
AACGCCTGGAAAAAATTTTTCCAAGCATCGACACAGACACACATCCCATCTTCTTCTTC
AAAACCCAGTGCACACAAC TAACGACACACTAGACTATACTGAAGATCCAACCGAAATC
AAACAATTCGTGTTTTGTGCGAAAAAGATCGGTCTTAGAATATCAATCCCATTACCTAAT
GGCACAGCGATTGCCGGTGGAGAACCATTAAAAACAATTATCCAAGATTTTAAATCACGTT
ATGTGCTTAGATCAGGAGCTTAGTCAGATCCTCAATTCCATAACATTTCCCGGGGAATGT
TTCGACATCGGCATTTACCACACAGGTCACTGCATCAGAATAGGTTATATGTACAAGACA
GACATGGACAAAGGGAAAAATGCTACATGGACGTCTAACTCCCATATTTATCGTCCCAGAA
GGTTACCGAAATTC TTGCAAGACGTTTATTTCAGATGCAAATGGATTTAAACAACCTTCTA
CATCACGGTACTAAGAAGGCTCCCATTGAAGAATTAATCTATAACATAACGGATAAAGGC
TGCCCTAAAGAAAACCTAAGTTTCATGGATTTAAATCCAGACAACCTGTGGAACAAAGTT
AACATAGCTACAGAGACTCTAATCACCAAGTACCTAAACACGCATGGTTTTAAACAACAAT
GCAACATCAGCGGATGACTCCCTTTTTATCGTTTTATCAGATTAATAGGATGGCCAATCATT
AAAACGCAACTAATTACTCATTACGAGACTAGGATAGCGCAACAGTTTTTACAAGTAACT
TTTATTTAAATAGATTCAAAGAACCCTCAGATCAAAAAACACAATTTGGAAGAGTTTTCG
GATTTTAGCTGCC TGAACAGACAACATAGAGGAAATCGAGATAACGTCCTAGTTTACATA
CAACTCAAAGCCGACGGAAATCGGTTAATTTCTCATCTTATGGAGCACGTGCTTTGCTACT
AAATGCCAGTCCAACAGCAAACAAGTACATTGTTCCATCGCCTTGGAGCAATTAAAAAAC
TAA

>5055_U43

ATGACCATCGTAGTGTGTTGCGACTGAATACGATGCCGCCAACGTCATTTTTAGTATCCTA
TG TAGAAGTCCCAGTGAACATTTAATCTTTCCGATTATTGTTAAATATAAACCATCAAAT
AACATTAGTTTTCTGCC TTTCAAACCTCAAAAATGCAAAAATTCAAAGCGAATCGACACGGTA
TTTGTCTGCCATGCCGAAAAATTAATCTCAGCCACTACATACAAAACCTGCAAGTCCGATA
AAGGCTGAAGATGTCGCCAACAGCCTCAACGACAAAGAAACCGAATCGTTATATGTGCGAC
ATGATATTATCTCAAAC TGGAAAAGAAAGAGAGGACGTCGAATTTAAATACATGGCCTAT
TTCCACAAATCGCTCATTTGTCAAATATCTCACCGGTAAATTCCTATTGCCCACTTCCCCC
TTTTGGTTTTTTATCGACCTACGGGCAAACAGAAGGTATGCTTTTTATTAACAATGTTTTAC
TATTTGTTTCGAGGAACAAAAGAGTACCATCACGACTACCAAAAATTATGTTCAATGCTTC
ACGGAACACCCGAAACATGGTGTTCACGTATTTCTCCATGTCCGAATTTATAAACATC
ACTTTAAATCAAATTTAGAAAAC TATTTGCAGACTTCTCAACGTACGCCAGACAAAAA
AATCTTCGCGATAAGGAAGAATTTAAACATCTAGACACTCAAATCAATCTATTTAGAAAG
AGTTCACATCTAACCAACACATTCGCGTCCATTACATATATATCGCTTACAACACGGCC
CTAGAGACAACAAAATTCGTCAATTACTGCAATCTCACTTCTTACGACTCTAATTTACCA
ATAGGGCAACAATGT CAGAGAAATCTCCACATTTCTGGAAATTCCTGCACGAAAAATTTA
CTGTGTATCATGAAGCAATAC TTTAATGCGGATTGTTACTTTAAACATACATAGACATA
AAAAGATTAAAAAATCCAAATCTAAACGTTACAGAATACGAGTATGCTCTGGTCTCTAAA
AAGAAGCCATCCAAGCATTAAC TCCGAACAAATAACGCGTGCCATAGCGAAATGTAAT
AAAAACGGAGAAGGTCTTTTTCTCCAGTAAAAC CAGGATTACAAGGATTATTAGAAATC
TCCGCTTCCGATAGATACGTT CAGATACAAGATAAGCGCATCTACAGAAGACAACACCTA
CACAAAGATTATCACCGTCCC TTTCTGTATTTCAGGGTGCAGCTTTTACATAAAAAATATT
TTTTGCTTCGAAACCTCTGAGGATTGGTACGAAAACATGGGATTTAACCGCATCTTACAG
TATCTCCCCGACGAATATATCTCTGATGAAGCTCTAACTCGCGCCATCTGGCTGCAAGAC
ACACATTTTCTCTGCGACGACGTGCAAAAACAATTTCTATACCACACGACACGAAATATTC
AACGAACGCATACCAGTCACAAATTACATAGGAGACTTGGATCTTCCGCTGCAAGATACA
GCAACTATTACGGAGGAAACATTTTTCTCCATGTGTGCGCTCATAAGACTTACCCTCATC
AACGCCTGGAAAAAATTTTTCCAAGCATCGACACAGACACACATCCCATCTTCTTCTTC
AAAACCCAGTGCACACAAC TAACGACACACTAGACTATACTGAAGATCCAACCGAAATC
AAACAATTCGTGTTTTGTGCGAAAAAGATCGGTCTTAGAATATCAATCCCATTACCTAAT
GGCACAGCGATTGCCGGTGGAGAACCATTAAAAACAATTATCCAAGATTTTAAATCACGTT
ATGTGCTTAGATCAGGAGCTTAGTCAGATCCTCAATTCCATAACATTTCCCGGGGAATGT
TTCGACATCGGCATTTACCACACAGGTCACTGTATCAGAATAGGTTATATGTACAAGACA
GACATGGACAAAGGGAAAAATGCTACATGGACGTCTAACTCCCATATTTATCGTCCCAGAA
GGTTACCGAAATTC TTGCAAGACGTTTATTTCAGATGCAAATGGATTTAAACAACCTTCTA
CATCACGGTACTAAGAAGGCTCCCATCGAAGAATTAATCTATAACATAACGGATAAAGGC
TGCCCTAAAGAAAACCTAAGTTTCATGGATTTAAATCCAGACAACCTGTGGAACAAAGTT

AACATAGCTACAGAGACTCTAATCACCAAGTACCTAAACACGCATGGTTTTAACAACAAT
GCAACATCAGCGGATGACTCCCTTTTATCGTTTATCAGATTAATAGGATGGCCAATCATT
AAAACGCAACTAATTACTCATTACGAGACTAGGATAGCGCAACAGTTTTTACAAGTAACT
TTTATTTAAAATAGATTCAAAGAACCCTCCAGATCAAAAAACACAATTTGGAAGAGTTTCG
GATTTTAGCTGCCCTGAACAGACAACATAGAGGAAATCGAGATAACGTCCTAGTTTACATA
CAACTCAAAGCCGACGGAAATCGGTTAATTCTCATCTTATGGAGCACGTGCTTTGCTACT
AAATGCCAGTCCAACAGCAAACAAGTACATTGTTCCATCGCCTTGGAGCAATTAAAAAAC
TAA

>5814_U43

ATGACCATCGTAGTGTTCGCGACTGAATACGATGCCGCCAACGTCATTTTTAGTATCCTA
TGTAAGTCCCAGTGAACATTTAATCTTTCCGATTATTGTTAAATATAAACCATCAAAT
AACATTAGTTTCTGCCCTCAAACCTCAAAAATGCAAAAATTCAAAGCGAATCGACACGGTA
TTTGTCTGCCATGCCGAAAAATTAATCTCAGTCACTACATACAAACTGCAAGTCCGATA
AAGGCTGAAGATGTCGCCAACAGCCTCAACGACAAAAGAAACCGAATCGTTATATGTCGAC
ATGATATTATCTCAAACCTGGAAGAAAGAGAGGACGTCGAATTTAAATACATGGCCTAT
TTCCACAAATCGCTCATTATCAAATATCTCACCGGTAAATTCCTATTGCCCACTTCCCC
TTTTGGTTTTTATCGACCTACGGGCAAAACAGAAGGTATGCTTTTATTAACAATGTTTTAC
TATTTGTTTCGAGGAACAAAAGAGTACCATCACGACTACCAAAAATTATGTTCAATGCTTC
ACGGAACACCCGAAACATGGTGTTCACGATTTCTTCATGTCCGAATTTATAAACATC
ACTTTAAAATCAAATTTAGAAAACATTTTGCAGACTTCTCAACGTACGCCAGACAAAAA
AATCTTCGCGATAAGGAAGAATTTAAACATCTAGACACTCAAATCAATCTATTTAGAAAAG
AGTTCACATCTAACCAACACATTCGCGTCCATTACATATATATCGCTTACAACACGGCC
CTAGAGACAACAAAATTCGTCAATTACTGCAATCTCACTTCTTACGACTCTAATTTACCA
ATAGGGCAACAATGTCAGAGAAATCTCCACATTTCTGGAAATTCCTGCACGAAAATTTA
CTGTGTATCATGAAGCAATACTTAATGCGGATTGTTACTTTAAAACATACATAGACATA
AAAAGATTAAAAAATCCAAATCTAAACGTTACAGAATACGAGTATGCTCTGGTCTCTAAA
AAGAAGACCATCCAAGCATTAACCTCCGAACAAAATAACGCGTGCCATAGCGAAATGTAAT
AAAAACGGAGAAGGTCTTTTTTCTCCAGTAAAAC CAGGATTACAAGGATTATTAGAAATC
TCCGCTTCCGATAGATACGTTT CAGATACAAGATAAGCGCATCTACAGAAGACAACACCTA
CACAAAGATTATCACCGTCCCTTTCTGTATTTCAGGGTGCAGCTTTTACATAAAAAATATT
TTTTGCTTCGGAACCTCTGAGGATTGGTACGAAAACATGGGATTTAACCGCATCTTACAG
TATCTCCCCGACGAATATATCTCTGATGAAGCTCTAACTCGCGCCATCTGGCTGCAAGAC
ACACATTTTCTCTGCGACGACGTCGAAAAACAATTTCTATACCACACGACACGAAATATTC
AACGAACGCATACCAGTACAAAATTACATAGGAGACTTGGATCTTCCGCTGCAAGATACA
GCAAATATTACGGAGGAAACATTTTTCTCCATGTGTGCGCTCATAAGACTTACCCTCATC
AACGCTTGAAAAAAATTTTTTCCAAGCATCGACACAGACACACATCCCATCTTCTTCTTC
AAAACCCAGTGCACACAAC TAACGACACACTAGACTATACTGAAGATCCAACCGAAATC
AAACAATCTGTGTTTTGTTCGAAAAAGATCGGTCTTAGAATATCAATCCCATTACCTAAT
GGCACAGCGATTGCCGTTGGAGAACCATTAACAATTATCCAAGATTTTAAATCACGTT
ATGTGCTTAGATCAGGAGCTTAGTCAGATCCTCAATTCATAACATTTCCCGGGGAATGT
TTCGACATCGGCATTTACCACACAGGTCACTGCATCAGAATAGGTTATATGTACAAGACA
GACATGGACAAAGGAAAAATGCTACATGGACGTCTAACTCCCATATTTATCGTCCCAGAA
GGTTACCGAAATTTTGAAGACGTTTATTTCAGATGCAAATGGATTTAAACAACCTTCTA
CATCACGGTACTAAGAAGGCTCCCATTTGAAGAATTAATCTATAACATAACGGATAAAGGC
TGCCCTAAAGAAAACCTAAGTTTTCATGGATTTAAAATCCAGACAACTGTGGAACAAAGTT
AACATAGCTACAGAGACTCTAATCACCAAGTACCTAAACACGCATGGTTTTAACAACAAT
GCAACATCAGCGGATGACTCCCTTTTATCGTTTATCAGATTAATAGGATGGCCAATCATT
AAAACGCAACTAATTACTCATTACGAGACTAGGATAGCGCAACAGTTTTTACAAGTAACT
TTTATTTAAAATAGATTCAAAGAACCCTCCAGATCAAAAAACACAATTTGGAAGAGTTTCG
GATTTTAGCTGCCCTGAACAGACAACATAGAGGAAATCGAGATAACGTCCTAGTTTACATA
CAACTCAAAGCCGACGGAAATCGGTTAATTCTCATCTTATGGAGCACGTGCTTTGCTACT
AAATGCCAGTCCAACAGCAAACAAGTACATTGTTCCATCGCCTTGGAGCAATTAAAAAAC
TAA

U44

>2284_U44

ATGGGAAATTGTTTTCGACGATTATCTTTGTTTGGACAGCCGACTCAGGCAAAGTATGAA
ATGTTGTGCTCTTCCGACGATTTAGAGACCGAAGAGCTGACGATTTTTCTGGATATGACG
TATAAAGATTTTGGAGTTTATCAGAATGATATTATAAGTCATCAAAAAGACACGGAAACT
ATGAAAACCTCTGTTAGGCTTATTACCAATGTACAAAAAACTAAGCTCAGGCATACGGTT
ATGGAGCGTTGTTTATCTAATTGTCTAACCACGTTAAAGATGCGCTCTGTGTAGAGCTT
ATGAAAGCTGAAAAAATATTACAAACGATGGATGTCTGTTTTTATGAAAACTTTAAATTGGC
GAATTTTCCATGTGCACCGATAATCTTAATCAGTTGCTTAACAAATTCGCTACAGATCAG
TCAACACTTAGCGATGTGAGAAAATAAACAGTCTCATAGAGATTGATGGCGAAAACAGC
AAACGTCTTTTAGTTCGAGTTGGATCCGATTCTCCATGAGGAGACGGGATTATATCAAGCG
CTGCCAAATGTTGTGACGGAAGCTCCAAGTGAAAAGGTTAAATCGATACGCGTTGAGTCT
GAAGGTGAGAGTGTATAG

>5055_U44

ATGGGAAATTGTTTTCGACGATTATCTTTGTTTGGACAGCCGACTCAGGCAAAGTATGAA
ATGTTGTGCTCTTCCGACGATTTAGAGACCGAAGAGCTGACGATTTTTCTGGATATGACG
TATAAAGATTTTGGAGTTTATCAGAATGATATTATAAGTCATCAAAAAGACACGGAAACT
ATGAAAACCTCTGTTAGGCTTATTACCAATGTACAAAAAACTAAGCTCAGGCATACGGTT
ATGGAGCGTTGTTTATCTAATTGTCTAACCACGTTAAAGATGCGCTCTGTGTAGAGCTT
ATGAAAGCTGAAAAAATATTACAAACGATGGATGTCTGTTTTTATGAAAACTTTAAATTGGC
GAATTTTCCATGTGCACCGATAATCTTAATCAGTTGCTTAACAAATTCGCTACAGATCAG
TCAACACTTAGCGATGTGAGAAAATAAACAGTCTCATAGAGATTGATGGCGAAAACAGC
AAACGTCTTTTAGTTCGAGTTGGATCCGATTCTCCATGAGGAGACGGGATTATATCAAGCG
CTGCCAAATGTTGTGACGGAAGCTCCAAGTGAAAAGGTTAAATCGATACGCGTTGAGTCT
GAAGGTGAGAGTGTATAG

>5814_U44

ATGGGAAATTGTTTTCGACGATTATCTTTGTTTGGACAGCCGACTCAGGCAAAGTATGAA
ATGTTGTGCTCTTCCGACGATTTAGAGACCGAAGAGCTGACGATTTTTCTGGATATGACG
TATAAAGATTTTGGAGTTTATCAGAATGATATTATAAGTCATCAAAAAGACACGGAAACT
ATGAAAACCTCTGTTAGGCTTATTACCAATGTACAAAAAACTAAGCTCAGGCATACGGTT
ATGGAGCGTTGTTTATCTAATTGTCTAACCACGTTAAAGATGCGCTCTGTGTAGAGCTT
ATGAAAGCTGAAAAAATATTACAAACGATGGATGTCTGTTTTTATGAAAACTTTAAATTGGC
GAATTTTCCATGTGCACCGATAATCTTAATCAGTTGCTTAACAAATTCGCTACAGATCAG
TCAACACTTAGCGATGTGAGAAAATAAACAGTCTCATAGAGATTGATGGCGAAAACAGC
AAACGTCTTTTAGTTCGAGTTGGATCCGATTCTCCATGAGGAGACGGGATTATATCAAGCG
CTGCCAAATGTTGTGACGGAAGCTCCAAGTGAAAAGGTTAAATCGATACGCGTTGAGTCT
GAAGGTGAGAGTGTATAG

U45

>2284_U45

ATGTACAGCGCAATTTTCAGAAAAGATCTCCGAAACGATTACCCTACAGAGACAAACGAGC
TCGAGATACATTTGAATTTTTTCGTCTTTTCGAAATGTGGACATCAATGAACTTTGGACTACA
GACATATCTGAAGACAAAACGCACGACGTGTGGCCGGCTGTAAATGAGAAATCATTTAAA
AAATTTCTAGAAAACGAACTAACCTCTTATCAACGCCCAATCCCGCTTCTAGGTATTCCG
CAAAATGGAACAGTGTGCAAGACCTGCAAAAAAGAGAAGCAAAGAGAAAACCGATTGTGTT
AACTATGAAAGGAAACACGGCAACCCAGTAACTTTCTACCCAAGGCACCGCGCAAAAAGA
AATGCAAATACGGACACTTGTATTTTCGGAAGAACCATCTATTTTAGTAAGCCATCACAGG
AATAGTAAAATGGACGTTTTTATGGACACCAACAAAATCACACTCGTAAATAGAGAACTC
ATATGGGTGCCGCACGATCAAGTTCGAATCGTAAAACCTAGATATTTCCCTGTACATCCCA
GATGGTTTCTTCGGAGTTATTACTGGCCACAGCAACGACGATTTCTGCGAATGCGTCAGT
GAAATAATAACCGACGAAACCGACATTTTCAGTTTTTCTTAATGAACTTATCCGAGCACTCT
CTAATGCTTCTCCCAGGAGATGTGGAATTCTCAATCAACTTTCTCCCTTGCTACATCCCG
GAACCATGGGAGATGATCAATCTGTCTCCCCCGAATCTGCAATTTTCCATCTAAAAGCA
TGCAGGGAATTTATAGCTAAACCGAACAGCTATACGATACAATACTTCGACGCCATGTAC
GTTTGCGCCGATGAACTGAAGGCTCTCATGATTCCCTCAAAAAGAAATTGCCAAACTCGGC

TTATTGATCGAAACCTACATCTGGAACAAAGACACAATTCCCTCCATAAAGATCTTCAAC
TCCACACGCAAACTATCTACATCCCTACAGGCATCTGCATCGCCAGGATAATTTTTACT
TGCGGACATTTTTGCCTCAGTTTAATGCCTGAACGTGCAATAAACCGGCTACAGGTACTA
GACGCCAATTCCTTTCTTTGTTTTATTATGCCGCCTCCAATAACGCTTAA

>5055_U45

ATGTACAGCGCAATTTTCAGAAAAGATCTCCGAAACGATTACCCTACAGAGACAAACGAGC
TCGAGATACATTGAATTTTTTCGTCTTTTCGAAATGTGGACATCAATGAACTTTGGACTACA
GACATATCTGAAGACAAAACGCACGATGTGTGGCCGGCTGTAAATAAGAAATCATTTAAA
AAATTTCTAGAAAACGAACTAACCTCTTATCAACGCCCAATCCCGCTTCTAGGTATTCCG
CAAAATGGAACAGTGTGCAAGACCTGCAAAAAAGAGAAGCAAAGAGAAACCGATTGTGTT
AACTATGAAAGGAAACACGGCAACCCAGTAACTTTCTACCCAAGGCACCGCGCAAAAAGA
AATGCAAAATACGGACACTTGTATTTTCGGAAGAACCATCTATTTTAGTAAGCCATCACAGG
AATAGTAAAATGGACGTTTTTATGGACACCAACAAAATCACACTCGTAAATAGAGAACTC
ATATGGGTGCCGCACGATCAAGTTCGAATCGTAAAACCTAGATATTTCCCTGTACATCCCA
GATGGTTTTCTTCGGAGTTATTACTGGCCACAGCAACGACGTATTCTGCGAATGCGTCACT
GAAATAATAACCGACGAAACCGACATTTTCAGTTTTTCTAATGAACTTATCCGAGCACTCT
CTAATGCTTCTCCCAGGAGATGTGGAATTTCTCAATCAACTTTCTCCCTTGCTACATCCCG
GAACCATGGGAGATGATCAATCTGTCTCCCCCGAATCTGCAATTTTCCATCTAAAAGCA
TGCAGGGAATTTATAGCTAAACCGAACAGCTATACGATACAATACTTTCGACGCCATGTAC
GTTTGCGCCGATGAACTGAAGGCTCTCATGATTCCCTCAAAAAGAAATTGCCAAACTCGGC
TTATTGATCGAAACCTACATCTGGAACAAAGACACAATTCCCTCCATAAAGATCTTCAAC
TCCACACGCAAACTATCTACATCCCTACAGGCATCTGCATCGCCAGGATAATTTTTACT
TGCGGACATTTTTGCCTCAGTTTAATGCCTGAACGTGCAATAAACCGGCTACAGGTACTA
GACGCCAATTCCTTTCTTTGTTTTATTATGCCGCCTCCAATAACGCTTAA

>5814_U45

ATGTACAGCGCAATTTTCAGAAAAGATCTCCGAAACGATTACCCTACAGAGACAAACGAGC
TCGAGATACATTGAATTTTTTCGTCTTTTCGAAATGTGGACATCAATGAACTTTGGACTACA
GACATATCTGAAGACAAAACGCACGACGTGTGGCCGGCTGTAAATGAGAAATCATTTAAA
AAATTTCTAGAAAACGAACTAACCTCTTATCAACGCCCAATCCCGCTTCTAGGTATTCCG
CAAAATGGAACAGTGTGCAAGACCTGCAAAAAAGAGAAGCAAAGAGAAACCGATTGTGTT
AACTATGAAAGGAAACACGGCAACCCAGTAACTTTCTACCCAAGGCACCGCGCAAAAAGA
AATGCAAAATACGGACACTTGTATTTTCGGAAGAACCATCTATTTTAGTAAGCCATCACAGG
AATAGTAAAATGGACGTTTTTATGGACACCAACAAAATCACACTCGTAAATAGAGAACTC
ATATGGGTGCCGCACGATCAAGTTCGAATCGTAAAACCTAGATATTTCCCTGTACATCCCA
GATGGTTTTCTTCGGAGTTATTACTGGCCACAGCAACGACGTATTCTGCGAATGCGTCACT
GAAATAATAACCGACGAAACCGACATTTTCAGTTTTTCTAATGAACTTATCCGAGCACTCT
CTAATGCTTCTCCCAGGAGATGTGGAATTTCTCAATCAACTTTCTCCCTTGCTACATCCCG
GAACCATGGGAGATGATCAATCTGTCTCCCCCGAATCTGCAATTTTCCATCTAAAAGCA
TGCAGGGAATTTATAGCTAAACCGAACAGCTATACGATACAATACTTTCGACGCCATGTAC
GTTTGCGCCGATGAACTGAAGGCTCTCATGATTCCCTCAAAAAGAAATTGCCAAACTCGGC
TTATTGATCGAAACCTACATCTGGAACAAAGACACAATTCCCTCCATAAAGATCTTCAAC
TCCACACGCAAACTATCTACATCCCTACAGGCATCTGCATCGCCAGGATAATTTTTACT
TGCGGACATTTTTGCCTCAGTTTAATGCCTGAACGTGCAATAAACCGGCTACAGGTACTA
GACGCCAATTCCTTTCTTTGTTTTATTATGCCGCCTCCAATAACGCTTAA

U46

>2284_U46

ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTACATT
CTTGTTTTTGCGCCGCTACGGAAGTGGACTTTTTATTCTCCAGAGTGCCATTTCGCATACC
TATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATTG
TTATGTTCTTTTCGGATTTTCTTGAATACTGGTGTATAAGACTTTTGCTTCAGAGACG
GTGAAGGGTTATTAG

>5055_U46

ATGAGTTGTAAGAAAAGCGAAAGGCAACGCTTATATGTCAGTCTCTGGCTCTTTTCCATT
CTTGTTTTTTCGCGCCGCTACGGAAGTGGACTTTTTATTCTCCAGAGTGCCATTTCGCATACC
TATGAGATCGTCTTGAATTCGTTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATTG
TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAAGACTTTTGCTTCAGAGACG
GTGAAGGGTTATTAG

>5814_U46

ATGAGTTGTAAGAAAAGCGCAAGGCAAAGCTTATATGTCAGTCTCTGCCTCTTTTACATT
CTTGTTTTTTCGCGCCGCTACGGAAGTGGACTTTTTATTCTCCAGAGTGCCATTTCGCATACC
TATGAGATCGTCTTGAATTCATTTTCTTCGATCTGGCTTTTGATAAATCTTTTTTTATTG
TTATGTTCTTTTCGCGATTTTCTTGAATACTGGTGTATAAAGACTTTTGCTTCAGAGACG
GTGAAGGGTTATTAG

U51

>2284_U51

ATGGGGAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
TTTATTTTTTTCGTCATACAGTTGTGTACAATGGTTTTGCTGACGGTTAGGTTTAATAGT
TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTTAATTGTTTT
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
GCAATTCCTTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGTTT
ATGGCGATCGAGCGCATCAACCACCTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTCCTC
GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTTATTTCGGTCATGATTCCA
GTTATGCTGGTGTGTTTATTTACCTATACAGCAAAAATTTTATGCTTTTGAAAGATGAACTT
TCGACCGTGACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG
AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTTTCGCGGATATTTTT
ATGGCGCTTCTGTTTTTAAAAGTTTATTATATACCCGTTATGGCCTATTGCATGGCATGT
GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
TGTTTTTTCGTGTACGCGGCGGAAAAGGAGTCCGATTTAGAGGTTGGAATAAAAATGTTA
AAATGA

>5055_U51

ATGGGGAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
TTTATTTTTTTCGTCATACAGTTGTGTACAATGGTTTTGCTGACGGTTAGGTTTAATAGT
TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTTAATTGTTTT
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
GCAATTCCTTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGTTT
ATGGCGATCGAGCGCATCAACCACCTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTGGATCATATCGGGAGTGCAAGCTCTT
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTCCTC
GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTTATTTCGGTCATGATTCCA
GTTATGCTGGTGTGTTTATTTACCTATACAGCAAAAATTTTATGCTTTTGAAAGATGAACTT
TCGACCGTGACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG
AAAGCTGGCTTAAGTGAATCGAGAGTGACAAGATTTTTTATGGTTTTCGCGGATATTTTT
ATGGCGCTTCTGTTTTTAAAAGTTTATTATATACCCGTTATGGCCTATTGCATGGCATGT
GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
TGTTTTTTCGTGTACGCGGCGGAAAAGGAGTCCGATTTAGAGGTTGGAATAAAAATGTTA
AAATGA

>5814_U51

ATGGAGAAAGAAACGAAGTCTTTGGCTTGGCCAGCCACTGCGGAGTTTTATGGTTGGGTA
TTTATTTTTTTCGTCATACAGTTGTGTACAATGGTTTTACTGACGGTTAGGTTTAATAGT
TTTAAAGTCGGTCGAGAATACGCTGTGTTTACATTTGCCGGGATGAGCTTTAATTGTTTT
TTACTTCCGATAAAGATGGGATTGTTGAGTGGACATTGGTCTTTGCCGCGAGATTTTTGC
GCAATTCCTTCTACATCGATGATTTTTCTATCTATTTCTCCTCGTGGTCATTGGTGTTT

ATGGCGATCGAGCGCATCAACCACTTCTGTTACAGCACACCGTTGTTGAACGAGAATTCC
AAGGCTCTGGCAAAGGTGTGTTTCCCGATAGTTTTGGATCATATCGGGAGTGCAAGCTCTT
CAGATGCTGAATAATTACAAAGCAACAGCTTTGCAGAACGAGACCTCACAGTGTTCCTC
GCGTTCTTAAGGAGCGGACACGACATGTGGTTAATGTTGGTTTATTCGGTCATGATTCCA
GTTATGCTGGTGTTCATTTACATATACAGCAAAAATTTTCATGCTTTTGAAAGATGAACTT
TCGACCGTGACTACGTATCTCTGTATTTATTTGCTGCTAGGTACAATAGCGCATCTGCCG
AAAGCTGGCTTAAGTAAAATCGAGAGTGACAAGATTTTTTATGGTTTGCGCGATATTTTT
ATGGCGTTCTGTTTTAAAAGTTTTATTATATACCCGTTATGGCCTATTGCATGGCATGT
GATGATCACACTGTGCCGTTTCGTTTGTGTAGTATCTGGCTGGTCAATTTGTGTAAGAAA
TGTTTTTCGTGTACGCGGCGGAAAAGGAGTCCGATTTAGAGGTTGGAATAAAAATGTTA
AAATGA

U53

>2284_U53

ATGTCCTAAAGTTTGGGTAGGTGGATTCCTTTTGTGTTTATGGTGAAGAACCCTCGGAAGAA
TGTTTAGCTCTGCCAGGGATACGGTTCAAAAAGAAGTGGGGTCCGGAATATTCCTTTA
CCGTTGAATATTAATCACAACGAAAAGGCCACTATTGGAATGGTACGCGGTCTTTTCGAT
TTAGAGCATGGGCTTTTCTGCGTTGCGCAGATACAGTCTCAAACGTTTCATGGACATCATA
AGAAATATTGCTGGCAAGTCAAAGTTAATCACAGCAGGCTCGGTAATTGAACCTTTACCA
CCGGACCCGGAGATTGAGTGTAAAGTTCAAGTTTTCTGGTCTATCGTTGTCTAGCAAG
GTTTTGCAGGATGAAAATTTAGACGGTAAACCCTTTTTTTCATCATGTGTCTGTATGTGGA
GTCGGTCCGAGACCCGGAACATAGCGATTTTTCGGGCGAGAAATAAGCTGGATTCTAGAT
AGGTTTTTCATGCATTAGTGAGAGTGAGAAGAGACAGTTTTTGGAGGGGGTCAATGTTTAT
TCTCAGGGTTTCGATGAAAATTTATTTTTCGGCGGATTTATATGACTTGCTCGCGGATAGT
TTGGATACATCTTATATTAGAAAACGGTTTTCCAAAGTTGCAGTTGGATAAACAGCTATGT
GGTTGTCTAAATGTACGTATATTAAGCTAGTGAACCACCTGTGGAGATTATTGTGCGCA
GCGGCTAAAGTTGCCGGTGATCAAGTTTCAGTTAACTACGGAACCTGGGTCTGAATTAGCG
GTTGAAACGTGTGACGTGTCTGTTGTGCACGGGAATTACGACGCCGTTGAGTCTGCGACA
GCTACATCGGCTATGAGTAATCAAAATCCGCCAAATACTACTCTTTGCTGTGAGTCCA
CCGTTTTTCGGATTGTGTTTTTTTTTACCAGAAAGATGCTTTTTTTTTCTCTTTTAAATGTTACA
ACCGGACAGCAATCGAAAGTAGTTCCGCCTGTTTCTGTTTCATCCGCCCGTCACTGAGCAG
TATCAAATGCTACCGTATTCTGAGTCCGCTGCTAAGATTGCAGAACAGGAGTCAATCGA
TATCACAGTCTTGTGAGGCAATGTATCCCTATTGGCAATATTCCCCTGTTCCCCAGTAT
CCGGCCGCTTACACGGTTATCGTCAATCGAAAACGCTTAAAAAGCGGCATTTCCAGAGT
GATTCGGAAGATGAATTAAGTTTTCCGGGAGATCCGGAATACACAAAAAAGGAGGCGC
CATAGAGTTGACAACGACGACGATAAGGAGATGGCTCGAGAAAAGAACGATTTAAGAGAA
TTGGTGGATATGATAGGAATGTTAAGACAAGAGATTAGTGCCTGGAAGCAGTTTCGCGCT
CAATCGCCGACAGACATGTGCTTCCGATGGAGACTCTGCCTACGATCGAGGAGAAAAGGC
GCCGCTCCCGAAGCCATCTATTTTAAACGCTTCTTTGGCGCCTGAAACCGTAAATAGG
AGCCTTGCTGGTCAAGCAATCCACGGATCTGCTGAAACTCAACAAGAAATTGTTTGT
GACGCGTTAAATAAAATGGATAGTTAA

>5055_U53

ATGTCCTAAAGTTTGGGTAGGTGGATTCCTTTTGTGTTTATGGTGAAGAACCCTCGGAAGAA
TGTTTAGCTCTGCCAGGGATACGGTTCAAAAAGAAGTGGGGTCCGGAATATTCCTTTA
CCGTTGAATATTAATCACAACGAAAAGGCCACTATTGGAATGGTACGCGGTCTTTTCGAT
TTAGAGCATGGGCTTTTCTGCGTTGCGCAGATACAGTCTCAAACGTTTCATGGACATCATA
AGAAATATTGCTGGCAAGTCAAAGTTAATCACAGCAGGCTCGGTAATTGAACCTTTACCA
CCGGACCCGGAGATTGAGTGTAAAGTTCAAGTTTTCTGGTCTATCGTTGTCTAGCAAG
GTTTTGCAGGATGAAAATTTAGACGGTAAACCCTTTTTTTCATCATGTGTCTGTATGTGGA
GTCGGTCCGAGACCCGGAACATAGCGATTTTTCGGGCGAGAAATAAGCTGGATTCTAGAT
AGGTTTTTCATGCATTAGTGAGAGTGAGAAGAGACAGTTTTTGGAGGGGGTCAATGTTTAT
TCTCAGGGTTTCGATGAAAATTTATTTTTCGGCGGATTTATATGACTTGCTCGCGGATAGT
TTGGATACATCTTATATTAGAAAACGGTTTTCCAAAGTTGCAGTTGGATAAACAGCTATGT
GGTTGTCTAAATGTACGTATATTAAGCTAGTGAACCACCTGTGGAGATTATTGTGCGCA
GCGGCTAAAGTTGCCGGTGATCAAGTTTCAGTTAACTACGGAACCTGGGTCTGAATTAGCG

GTTGAAACGTGTGACGTGTCTGTTGTGCACGGGAATTACGACGCCGTTGAGTCTGCGACA
GCTACATCGGCTATGAGTAATCAAAATCCGCCAAATACTACTCCTTTGCTGTGCGAGTCCA
CCGTTTTTCGGATTGTGTTTTTTTTACCGAAAGATGCTTTTTTTTTCTCTTTTAAATGTTACA
ACCGGACAGCAATCGAAAGTAGTTCCGCCTGTTTTCTGTTTCATCCGCCCGTCACTGAGCAG
TATCAAATGCTACCGTATTCTGAGTCCGGCTGCTAAGATTGCAGAACAGGAGTCAATCGA
TATCACAGTCCCTGTGACGGCAATGTATCCCTATTGGCAATATTCCCCTGTTCCCCAGTAT
CCGGCCGCTTACACGGTTATCGTCAATCGAAAACGCTTAAAAAGCGGCATTTCCAGAGT
GATTCGGAAGATGAATTAAGTTTTCCGGGAGATCCGGAATACACAAAAAAAAGGAGGCGC
CATAGAGTTGACAACGACGACGATAAGGAGATGGCTCGAGAAAAGAACGATTTAAGAGAA
TTGGTGGATATGATAGGAATGTTAAGACAAGAGATTAGTGCCTTGAAGCACGTTCCGCGCT
CAATCGCCGACAGAGACATGTCGTTCCGATGGAGACTCTGCCTACGATCGAGGAGAAAAGGC
GCCGCTCCCCGAAGCCATCTATTTTAAACGCTTCTTTGGCGCCTGAAACCGTAAATAGG
AGCCTTGCTGGTCAGAACGAATCCACGGATCTGCTGAAACTCAACAAGAAATTGTTTGT
GACGCGTTAAATAAAATGGATAGTTAA

>5814_U53

ATGTCATAAAGTTTGGGTAGGTGGATTCCCTTTGTGTTTATGGTGAAGAACCCTCGGAAGAA
TGTTTAGCTCTGCCAGGGATACGGTTCAAAAAGAAGCTGGGGTCCGGAATATTCCCTTTA
CCGTTGAATATTAATCACAAACGAAAAGGCCACTATTGGAATGGTACGCGGTCTTTTCGAT
TTAGAGCATGGGCTTTTCTGCGTTGCGCAGATACAGTCTCAAACGTTTCATGGACATCATA
AGAAATATTGCTGGCAAGTCAAAGTTAATCACAGCAGGCTCGGTAATTGAACCTTTACCA
CCGGACCCGGAGATTGAGTGTAAAGTTCAAGTTTTCTGGTCTATCGTTGTCTAGCAAG
GTTTTGCAGGATGAAAATTTAGACGGTAAACCCTTTTTTTCATCATGTGTCTGTATGTGGA
GTCGGTCCGAGACCGGGAACATAGCGATTTTCGGGCGAGAAATAAGCTGGATTCTAGAT
AGGTTTTTCATGCATTAGTGAGAGTGAGAAGAGACAGTTTTTGGAGGGGGTCAATGTTTAT
TCTCAGGTTTTCGATGAAAATTTATTTTCGGCGGATTTATATGACTTGCTCGCGGATAGT
TTGGATACATCTTATATTAGAAAACGGTTTTCCAAAGTTGCAGTTGGATAAACAGCTATGT
GGTGTGCTAAATGTACGTATATTAAGCTAGTGAACCACCTGTGGAGATTATTGTGCGCA
GCGGCTAAAGTTGCCGGTGATCAAGTTCAAGTTAACTACGGAACCTGGGTCTGAATTAGCG
GTTGAAACGTGTGACGTGTCTGTTGTGCACGGGAATTACGACGCCGTTGAGTCTGCGACA
GCTACATCGGCTATGAGTAATCAAAATCCGCCAAATACTACTCCTTTGCTGTGCGAGTCCA
CCGTTTTTCGGATTGTGTTTTTTTTACCGAAAGATGCTTTTTTTTTCTCTTTTAAATGTTACA
ACCGGACAGCAATCGAAAGTAGTTCCGCCTGTTTTCTGTTTCATCCGCCCGTCACTGAGCAG
TATCAAATGCTACCGTATTCTGAGTCCGGCTGCTAAGATTGCAGAACAGGAGTCAATCGA
TATCACAGTCCCTGTGACGGCAATGTATCCCTATTGGCAATATTCCCCTGTTCCCCAGTAT
CCGGCCGCTTACACGGTTATCGTCAATCGAAAACGCTTAAAAAGCGGCATTTCCAGAGT
GATTCGGAAGATGAATTAAGTTTTCCGGGAGATCCGGAATACACAAAAAAAAGGAGGCGC
CATAGAGTTGACAACGACGACGATAAGGAGATGGCTCGAGAAAAGAACGATTTAAGAGAA
TTGGTGGATATGATAGGAATGTTAAGACAAGAGATTAGTGCCTTGAAGCACGTTCCGCGCT
CAATCGCCGACAGAGACATGTCGTTCCGATGGAGACTCTGCCTACGATCGAGGAGAAAAGGC
GCCGCTCCCCGAAGCCATCTATTTTAAACGCTTCTTTGGCGCCTGAAACCGTAAATAGG
AGCCTTGCTGGTCAGAACGAATCCACGGATCTGCTGAAACTCAACAAGAAATTGTTTGT
GACGCGTTAAATAAAATGGATAGTTAA

U56

>2284_U56

ATGGAAACCGTCTACTGCACCTTTTCGATCACAACTGTCACCTTTCCGATATCAGCACCCCTA
TGCAAGCTCATGAACATCGTCATACCGATCCCAGCTCACCACCATCTAATAGGTAGCGGC
AATTTAGGTCTTTATCCCATCGTCTCCTCCAACAAAGATTACGTCCACGTCCGAAACGTA
TTACGGACCATGGTCGTAACCATCTGCAAAAGGTGGAAGGCAATCAACTGGTTTTGAGA
AAGCCGGTGACCGGCCAGCAGTATGCCATAAAGAACACAGGACCCTTCCCCTGGGAAAAG
GGTGACACCTTGACATTGATTCCACCTTTTATCAACGCACCTCGGAAGAAAAGCTATTGAAA
CTAGGTGATTGGGAGCTCACTGTACCTTTAGTCGTTCCGACAGCCATTGCGGCCGAGATC
AACATCAGATTGCTATGCATTGGTCTTATAGCAGTTTACAGAGAATACAATGAGATGCAA
ACAATCATCGATGAGTTATGCTCTATCCAATACAGAGACGTAATTAATTAATACCAGAC
ATAGTCAACGACAAACAATCAATGTACTCCATGAAAACAGCTTGCATCTCGCTGTCCATG

ATCACGGCCATGGCCCCAGACATCGTGAGAACTTACATAGATCGGCTAACGCTGGAAGAC
CATCAATGCTTTTGGATCAAATGCCAAGAACTATTATCTAAAAGAACAACGTTAAACACA
CAACGCTGCGGACAACCTTACGCTACGGAAATTAAGATGAATTAATAAAGTCAAATCA
GTCTTAACTATGATCGACCAAATTAATTCCTTTGACCAACGAAAAACGTATTTTGTGTC
TGTGACGTTTCCGCAGATAACAGAATGGCTACATGTATTTATAAAAAATTAA

>5055_U56

ATGGAAACCGTCTACTGCACCTTTCGATCACAACTGTCACTTTCCGATATCAGCACCTTA
TGCAAGCTCATGAACATCGTCATACCGATCCCAGCTCACCACCATCTAATAGGTAGCGGC
AATTTAGGTCTTTATCCCATCGTCTCCTCCAACAAAGATTACGTCCACGTCCGAAACGTA
TTACGGACCATGGTCGTAACCATCTGCAAAAAGGTGGAAGGCAATCAACTGGTTTTGAGA
AAGCCGGTGACCGGCCATCAGTATGCCATAAAGAACACAGGACCCTTCCCCTGGGAAAAG
GGTGACACCTTGACATTGATTCCACCTTTATCGACGCACTCGGAAGAAAAGCTATTGAAA
CTAGGTGATTGGGAGCTCACTGTACCTTTAGTCGTTCCGACAGCCATTGCGGCCGAGATC
AACATCAGATTGCTATGCATTGGTCTTATAGCAGTTCACAGAGAATACAATGAGATGCAA
ACAATCATCGATGAGTTATGCTCTATCCAATACAGAGACGTAATTAATTAACCAGAC
ATAGTCAACGACAAACAATCAATGTACTCCATGAAAACAGCTTGCATCTCGCTGTCCATG
ATCACGGCCATGGCCCCAGACATCGTGAGAACTTACATAGATCGGCTAACGCTGGAAGAC
CATCAATGCTTTTGGATCAAATGCCAAGAACTATTATCTAAAAGAACAACGTTAAACACA
CAGCGCTGCGGACAACCTTACGCTACGGAAATTAAGATGAATTAATAAAGTCAAATCA
GTCTTAACTATGATCGACCAAATTAATTCCTTTGACCAACGAGAAAACGTATTTTGTGTC
TGTGACGTTTCCGCAGATAACAGAATGGCTACATGTATTTATAAAAAATTAA

>5814_U56

ATGGAAACCGTCTACTGCACCTTTCGATCACAACTGTCACTTTCCGATATCAGCACCTTA
TGCAAGCTCATGAACATCGTCATACCGATCCCAGCTCACCACCATCTAATAGGTAGCGGC
AATTTAGGTCTTTATCCCATCGTCTCCTCCAACAAAGATTACGTCCACGTCCGAAACGTA
TTACGGACCATGGTCGTAACCATCTGCAAAAAGGTGGAAGGCAATCAACTGGTTTTGAGA
AAGCCGGTGACCGGCCAGCAGTATGCCATAAAGAACACAGGACCCTTCCCCTGGGAAAAG
GGTGACACCTTGACATTGATTCCACCTTTATCAACGCACTCGGAAGAAAAGCTATTGAAA
CTAGGTGATTGGGAGCTCACTGTACCTTTAGTCGTTCCGACAGCCATTGCGGCCGAGATC
AACATCAGATTGCTATGCATTGGTCTTATAGCAGTTCACAGAGAATACAATGAGATGCAA
ACAATCATCGATGAGTTATGCTCTATCCAATACAGAGACGTAATTAATTAACCAGAC
ATAGTCAACGACAAACAATCAATGTACTCCATGAAAACAGCTTGCATCTCGCTGTCCATG
ATCACGGCCATGGCCCCAGACATCGTGAGAACTTACATAGATCGGCTAACGCTGGAAGAC
CATCAATGCTTTTGGATCAAATGCCAAGAACTATTATCTAAAAGAACAACGTTAAACACA
CAACGCTGCGGACAACCTTACGCTACGGAAATTAAGATGAATTAATAAAGTCAAATCA
GTCTTAACTATGATCGACCAAATTAATTCCTTTGACCAACGAAAAACGTATTTTGTGTC
TGTGACGTTTCCGCAGATAACAGAATGGCTACATGTATTTATAAAAAATTAA

U57

>2284_U57

ATGGAAAAATTGGCAGGCGACCGAAATTTTACCTAAGATCGAAGCACCTCTAAATATTTTC
AATGACATTAACACATACACAGCCGAACAACCTTTTGGACAATTTGCGAATTTATTTCTGGT
GACGATCCGAGCCGTTACAACATCAGTTTTGAAGCCTTACTCGGAATCTACTGCAACAAA
ATAGAATGGATTAACTTTTTACCACGCCGATCGCCGTTGCAGCGAACGTAATCCGCTTC
AATGATGTGAGTCAATGACCTCGGGAAGGTTCTTCTTTTATTCAATTACCTAGAGTC
GCCACAGGAAACGACGTAACCTGCTTCAAAAAGAAACCACCATCATGGTAGCCAAACACTCA
GAAAAACACCCCATAAACATATCGTTGATTTGAGCCCTGCCTGTCTGGAACATCTGGAA
AACACATTTAAAAACACAGTCATCGATCAGATTTTAAACATCAATGCGTTACATACAGTC
TTAAGATCTTTAAAGAATTCAGCCGATTCGCTCGAGCGAGGTCTGATTACGCATTATG
CAAACCTTATTGAGAAAATCTCCCCGCAATTTATCGTCTGACCATGAATGAGAACAAA
GTACATAATAACAAGCTCTGAGCCGAGTACAACGCAGCAACATGTTTCAGAGCCTGAAG
AACAGATTGTTAACGTCATTATTTTTTTTTGAACAGGAATAATAATTTTCATATATCTAT
AGAATCTAAACGACATGATGGAATCGGTCACGGAAAGCATTCTAAATGATACGAACAAC
TACACTTCAAAGAAAACGTCCCCCTAGATGGTGTTTTATTAGGACCGATCGGCTCTATC

CAAAAACCTACCAGCATACTCTCCCAGTACATCTCCACACAAGTCGTCTCCGCCCAATC
TCATATGGTCACTTTTATTATGGGCAAAGAAAACGCAGTGACTGCGATTGCATACCGTGCA
ATCATGGCCGATTTTACTCAATTCACCGTGAACGCCGGGACAGAACAACAAGACACTAAC
AACAAATCAGAAATCTTCGACAAAAACCGCGCGCACGCCGACCTAAAGCTGAACACGTTG
AACTGGGAGATAAATTAGTCGCATTCGACCACCTACACAAAGTTTACAAAAACACAGAC
GTCACGATCCGCTAGAACAGAGCTTACAACCTAACATTTCTTTTCCCTTTGGGTATCTAC
ATACCGAGCGAGACCGGTTTCAGTACAATGGAAACACGTGTGAAATTAACGACACCATG
GAAAACAACCTACCCACCAGCGTTTTTTTTCCACAATAAAGACCAAGTCGTGCAGCGAATT
GATTTTGCCGACATATTACCGTCGGTTTTGCCATCCCATTGTCCACGACTCGACCATCGTC
GAACGACTCATGAAAAGCGAACCATTGCCTACCGGCCACCGCTTTTCCCAACTATGTCAA
CTAAAAATTACCCGAGAAAACCCAGCCAAGATCTTACAGACCTTATACAACCTTATACGAA
AGTCGACAAGAAGTACCCAAAAACACCAACGTCTTAAAAAACGAATTAACATTGAAGAT
TTTTACAAACCGGACAATCCAACGCTGCCGACCGAAAAGACACCCCTTCTTCGATCTCAG
TATATCCAGAAAAACCGAGCCACAGAAGTACTCTGCACACCAAGAATAATGATAGGCAAC
ATACCTTTACCGCTAGCTCCAATCTCTTTCCACGAAGCCCGTACAAATCAAATGCTGGAA
CATGCAAAGACGAACTGCCAAAACATGACTTCACCCTCAAATTTGTCACCGAAAAGCTTG
ACGAGTGGCTCGTACCCAGAATTGGCTTACGTTATCGAGACCTTAGTGCATGGAAACAAG
CATGCTTTTATAATCCTAAAACAAGTAATTAGCCAGTGTATTTCTTATTGGTTTTAACATG
AAACATATACTTCTTTTTTGCAACAGCTTCGAGATGATCATGCTAATCTCTAACCACATG
GGCGACGAACTGATCCCGGGAGCAGCCTTGTCTACTACAGAAATCTTGTGTGCGTAATT
CGCCTAGTGAAGAGAACAATCTCTATCTCCAACCTCAACGAGCAACTTTGCGGCGAACCT
CTGGTGAATTTGCGCAACGCGTTGTTGACGGACGTCTGTTCTGCCCGTTTCGTCCATACC
ATGCCAGAAAACGACACGAATGCAAAAATAACAGCGGATGATACACCCTGACACAGAAC
ACCGTAAGAGTTAGAAATACGAAATATCCGATGTGCAAAGAATGAATCTAATAGACTCA
AGCGTCGTCTTACCGACAATGACAGACCATCGAACGAAACCACCATCCTGAGCAAGATC
TTTTACTTCTGCGTACTCCCGGCATATCAAATAACAAGGCCTGTGGCGCTGGCGTCAAC
GTAAAGGAACTAGTTCTAGACTTATTTCTACACGGAACCGTTTCATCAGTCCAGATGATTAT
TTCCAGGAGAATCCGATCACCAGCGACGTTCTAATGTCTCTGATCCGAGAAGGTATGGGC
CCTGGCTACACCGTAGCCAACACATCCTCTATCGCAAAACAGCTGTTTTAAATCGCTAATC
TACATTAATGAAAATACGAAAATACCTGGAAGTGGAGGTCTCCTTAGATCCTGCGCAGCGA
CACGGCAACTCCGTTTCAATCACTACAACACATTCTATAACAACGGGCTTTGCCTG
ATCTCACCGATCACACCCTAAGACGGTACTATCAACCAATCCATTTTCATCGATTCTTC
TCCGACCCGGGAATCTGCGGCACCATGAATGCTGATATCCAAGTTTTCTAAATACATTT
CCTCACTGTCAAAGAAACGACGGCGGTTTTCTCTCCCGCCCCATTAGCATTAGAATTT
TATAATTGGCAACGAACACCGTTTTCCGTGTACTCAGCCTTCTGCCACAATTCCTGTG
AGCATTATGACGCTTGCCGCCATGCACTCAAATTTGTCTCCGTTGCCATAGCGATCCAA
AGCAAAAACAAAATCCATCCGGGCTTTGCGGCCACACTAGTCCGGACGGATAATTTGAC
GTCGAGTGCCTATTATACAGTTCCAGAGCAGCCACATCTATAATTTTAGACGATCCACG
GTCACCGCGGAAGCTAAAGATATCGCAACCACTTACAACCTTACCCAGCACCTAAGTTTT
GTAGATATGGGCTTAGGTTTTAGCTCTACCACCGCCACTGCCAATCTTAAGCGAATTA
TCAGATATGGGAGCAAGATACAAAACCTTTTTCTCCGCCTTCCCGATACACGCGTTTACC
AACACGGACATAAATACGTGGATTTCGGCATCACGTCCGGATAGAAAAACCTAATCCCTCC
GAGAGCGAAGCACATAACATCATAACATTCGGCGGAATTAACAAAAACCCACCCTCCATA
CTACTGCATGGTCAACAAGCTATCTGCGAAGTTATACTGACCCCGGTACGACAAAACATT
AACTTTTTCAAATCGCCCCACAACCAAGAGGCAGGGAATCGTGTATGATGGGAACGGAC
CCGCACAACGAAGAGGCGGCTAGAAAAGCATTGTACGACCACACCCAAACAGACAGCGAT
ACATTCGCCGCAACCACAAATCCTTGGGCATCCCTATCAGGCTCCTTAGGCGATATTCTA
TACAACACGGCACACAGAGAACAACCTATGTTACAACCCCAAGACATACAGTCCCTAACGCT
CAATTTTTTACCGAATCTGACATCTTAAAAACAAACAAGATGATGTACAAAGTGATAAGC
GAATACTGCATGAAATCGAATCTCGTGTTTAAACAGCGATAGCGAAATACAATACTCGTGC
TCTGAGGGCACGGATAGCTTCGTAAGCAGACCATGCCAGTTCTTACAAAACGCTCTGCCT
CTTCACTGTTTCATCCAACCAAGCTCTATTAGAGAGTCGGTCTAAAACCGGCAATACGCAG
ATCAGCGAAACCCATTATTGTAATTACGCCATAGGAGAAACCATACTTTCCAACCTCATT
ATCGAATCATCCATATAA

>5055_U57

ATGGAAAATTGGCAGGCGACCGAAATTTTACCTAAGATCGAAGCACCTCTAAATATTTTC
AATGACATTAACATAACAGCCGAACAACTTTTTGACAATTTGCGAATTTATTTTCGGT

GACGATCCGAGCCGTTACAACATCAGTTTTGAAGCCTTACTCGGAATCTACTGCAACAAA
ATAGAATGGATTAACTTTTTACCACCCCGATCGCCGTTGCAGCGAACGTAATCCGCTTC
AATGATGTGAGTCGAATGACCCTCGGGAAGGTTCTCTTTTATTCAATTACCCAGAGTC
GCCACAGGAAACGACGTAACCGCTTCAAAAAGAAACCACCATCATGGTAGCCAAACACTCA
GAAAAACACCCCATAAACATATCGTTCGATTTGAGCGCTGCCTGTCTGGAACATCTGGAA
AACACATTTAAAAACACAGTCATCGATCAGATTTTAAACATCAATGCGTTACATACAGTC
TTAAGATCTTTAAAGAATTCAGCCGATTCGCTCGAGCGAGGTCTGATTCACGCATTCATG
CAAACCTTATTGAGAAAATCCCCCCGCAATTTATCGTCTGACCATGAATGAGAACAAA
GTACATAATAAACAAGCTCTGAGCCGAGTACAGCGCAGCAACATGTTTTCAAAGCCTAAAG
AACAGATTGTTAACGTCAATTATTTTTTTTTGAACAGGAATAATAATACTTCATATATCTAT
AGAATTCATAACGACATGATGGAATCGGTTACGGAAGCATTCTAAATGATACGAACAAC
TACACTTCCAAAGAAAACGTCCCCCTAGATGGCGTTTTATTAGGACCGATCGGCTCTATC
CAAAAACACCAGCATACTCTCCAGTACATCTCCACACAAGTCGTCTCCGCCCAATC
TCATATGGTCACTTTATTATGGGCAAAGAAAACGCAGTGACTGCGATTGCATACCGTGCA
ATCATGGCCGATTTTACTCAATTCACCGTGAACGCCGGGACAGAACAACAAGACACTAAC
AACAAATCAGAAATATTCGACAAAAACCGCGCGTACGCCGACCTAAAGCTGAACACGTTG
AAATTGGGAGATAAATTAGTCGCATTCGACCACCTACACAAAGTTTACAAAAACACAGAC
GTCAACGATCCGCTAGAACAGAGCTTACAACCTGACATTCTTTTTCCCTTTGGGTATCTAC
ATACCGAGCGAGACCGTTTTTCAGTACAATGGAAACACGTGTGAAATTAACGACACCATG
GAAAACAACCTACCCACCAGCGTTTTTTTTCCACAATAAAGACCAAATCGTGCAGCGAATT
GATTTTGCCGACATATTGCCGTGCGTTTTGCCATCCATTGTCCACGACTCGACCATCGTC
GAACGACTCATGAAAAGCGAACCATTGCCTACCGGCCACCGCTTTTTCCCAACTATGTCAA
CTAAAAATTACCCGAGAAAACCCAGCCAGGATCTTACAGACCTTATACAACCTTATACGAA
AGTCGACAAGAAGTACCCAAAAACACCAACGTCTTAAAAAACGAATTAACATTGAAGAT
TTTTACAAACCGGACAATCCAACACTGCCGACCGAAAGACACCCCTTCTTCGATCTCAG
TATATCCAGAAAACCGAGCCACAGAAGTACTCTGCACACCAAGAATAATGATAGGCAAC
ATACCTTTACCGTTAGCTCCAGTCTCTTTCCACGAAGCCCGTACAAATCAAATACTGGAA
CATGCAAAGACGAAC TGCCAAAAC TACGACTTCACCCTCAAATTTGTCACCGAAAGCTTG
ACGAGTGGCTCGTACCCAGAATTTGGCTTACGTTATCGAGACCTTAGTGCATGGAAACAAG
CATGCTTTTTATGATCCTAAAACAAGTAATTAGCCAATGTATTTCTTATTGGTTTAACATG
AAACATATACTTCTTTTTTTGCAACAGCTTCGAGATGATCATGCTAATCTCTAACCACATG
GGCGACGAAC TGATCCCGGAGCAGCTTTCACTCACTACAGAAATCTTGTGTGCGTAATT
CGCTTAGTGAAGAGAACAATCTCTATCTCCAACCTCAACGAGCAACTTTGCGGCGAACCT
CTGGTGAATTTGCGCAACGCGTTGTTGACGCGACGTCTGTTCTGCCCGTTCTGTCATACC
ATGCCAGAAAACGACACGAATGCAAAAATAACAGCGGATGATACACCCTGACACAGAAC
ACCGTAAGAGTTAGAAATTCGAAATATCCGATGTGCAAAGAATGAATCTAATAGATTCA
AGCGTCTCTTACCGACAATGACAGACCATCGAACGAAACCACCATCCTGAGCAAGATA
TTTTACTTCTGCGTACTCCCGGCATATCAAATAACAAGGCCTGTGGCGCTGGCGTCAAC
GTAAAGGAAC TAGTTCTAGACTTATTTCTACACGGAACCGTTTCATCAGTCCAGATGATTAT
TTCCAGGAGAATCCGATTACCAGCGACGTTCTAATGTCTCTGATCCGAGAAGGTATGGGC
CCTGGCTACACCGTAGCCAACACATCCTGTATCGAAAACAGCTGTTTTAAATCGCTAATC
TACATTAATGAAAATACGAAAATAC TGGAAGTGGAGGTCTCCTTAGATCCTGCGCAGCGA
CACGGCAACTCCGTTCATTTTCAATCACTACAACACATTCTATACAACGGGCTTTGCCTG
ATCTCACCAATCACCACCC TAAGACGGTACTATCAACCAATCCCATTTTCATCGATTCTTC
TCCGACCCAGGAATCTGTGGCACCATGAATGCTGATATCCAAGTTTTCTCAAATACATTT
CCTCACTATCAAAGAAACGACGGCGGTTTTCCACTCCCGCCCCATTAGCATTAGAATTT
TATAATTGGCAACGAACACCGTTTTTCCGTGTA CTAGCCTTCTGCCCAATTCCTGTTG
AGCATTATGACGCTTGCCGCCATGCAC TCAAATTTGCTCTCCGTTGCCATAGCGATCCAA
AGCAAAAACAAAATCCATCCGGGCTTTGCGGCCACACTAGTCCGGACGGATAATTTTGAC
GTCGAGTGCC TATTATACAGTTCCAGAGCAGCCACATCTATAATTTTAGACGATCCACG
GTCACCGCGGAAGCTAAAGATATCGCAACCCTTACAACCTTACCCAGCACCTAAGTTTT
GTAGATATGGGCTTAGGTTTTAGCTCTACCACCGCCACTGCCAATCTTAAGCGAATTTAA
TCAGATATGGGGAGCAAGATACAAAACCTTTTTCTCCGCCTTCCCGATACACGCGTTTACC
AACGCGGACATAAATACGTGGATTTCGACATCACGTGCGGATAGAAAACCTAATCCCTCC
GAGAACGAAGCATAAATACATAACATTCGGCGGAATTAACAAAACCCACCCCTCCATA
CTACTGCATGGTCAACAAGCTATCTGCGAAGTTATACTGACCCCGGTCACGACAAACATT
AATTTTTTCAAATCGCCCCACAACCAAGAGGCAGGGAATCGTGTATGATGGGAACGGAC
CCGCACAACGAAGAGGCGCTAGAAAAGCATTGTACGACCACACCCAAACAGACAGCGAT

ACATTGCGCCGCAACCACAAATCCTTGGGCATCTCTACCAGGCTCCTTAGGCGATATTCTA
TACAACACGGCACACAGAGAACAACATATGTTACAACCCCAAGACATACAGTCCCAACGCT
CAATTTTTTACC GAATCTGACATCTTAAAAACAAACAAGATGATGTACAAAGTGATAAGC
GAATACTGCATGAAATCGAACTCGTGTTTAAACAGCGATAGCGAAATACAATACTCGTGC
TCTGAGGGCACGGATAGCTTCGTAAGCAGACCATGCCAGTTCTTACAAAACGCTCTGCCT
CTTCACTGTTTCAATCCAACCAAGCTCTATTAGAGAGTCCGGTCTAAAACCGGCAATACGCAG
ATCAGCGAAACCCATTATTGTAATTACGCCATAGGAGAAACCATACTTTTCCAACCTCATT
ATCGAATCATCTATATAA

>5814_U57

ATGGAAAATTGGCAGGCGACCGAAATTTTACCTAAGATCGAAGCACCTCTAAATATTTTC
AATGACATTTAAAACATACACAGCCGAACAACTTTTTGAACAATTTGCGAATTTATTTTCGGT
GACGATCCGAGCCGTTACAACATCAGTTTTTGAAGCCTTACTCGGAATCTACTGCAACAAA
ATAGAATGGATTAACTTTTTTACCACGCGGATCGCCGTTGCAGCGAACGTAATCCGCTTC
AATGATGTGAGTCAATGACCCTCGGGAAGGTTCTCTTTTATTCAATTACCTAGAGTC
GCCACAGGAAACGACGTAAC TGCTTCAAAGAAAACACCATCATGGTAGCCAAACACTCA
GAAAAACACCCCATAAACATATCGTTCGATTTGAGCCCTGCCTGTCTGGAACATCTGGAA
AACACATTTAAAACACAGTCATCGATCAGATTTTAAACATCAATGCGTTACATACAGTC
TTAAGATCTTTTAAAGAATTCAGCCGATTCGCTCGAGCGAGGTCTGATTCACGCATTCATG
CAAACCTTATTGAGAAAATCTCCCCCGCAATTTATCGTCTGACCATGAATGAGAACAAA
GTACATAATAAACAAGCTCTGAGCCGAGTACAACGCAGCAACATGTTTCAGAGCCTGAAG
AACAGATTGTTAACGTCATTATTTTTTTTTGAACAGGAATAATAATTTTCATATATCTAT
AGAATTC TAAACGACATGATGGAATCGGTCACGGAAGCATTCTAAATGATACGAACAAC
TACACTTCCAAAGAAAACGTCCCCCTAGATGGTGTTTTTATTAGGACCGATCGGCTCTATC
CAAAAACCTACCAGCATACTCTCCAGTACATCTCCACACAAGTCGTCTCCGCCCAATC
TCATATGGTCACTTTATTATGGGCAAAGAAAACGCAGTGACTGCGATTGCATACCGTGCA
ATCATGGCCGATTTTACTCAATTCACCGTGAACGCCGGGACAGAACAACAAGACACTAAC
AACAAATCAGAAATCTTCGACAAAACCGCGCGCACGCCGACCTAAAGCTGAACACGTTG
AAACTGGGAGATAAATTAGTCGCATTCGACCACCTACACAAAAGTTTACAAAAACACAGAC
GTCAACGATCCGCTAGAACAGAGCTTACAACCTAACATTCTTTTTCCCTTTGGGTATCTAC
ATACCGAGCGAGACCGGTTTTTCAAGTACAATGGAAAACACGTGTGAAATTTAAACGACACCATG
GAAAACAACCTACCACCAGCGTTTTTTTTCCACAATAAAGACCAAGTCGTGCAGCGAATT
GATTTTGCCGACATATTACCGTTCGGTTTTGCCATCCATTGTCCACGACTCGACCATCGTC
GAACGACTCATGAAAAGCGAACCATTGCCTACCGGCCACCGCTTTTTCCAACCTATGTCAA
CTAAAAATTACCCGAGAAAACCCAGCCAAGATCTTACAGACCTTATACAACCTTATACGAA
AGTCGACAAGAAGTACCCAAAACACCAACGTCTTAAAAAACGAATTTAAACATTGAAGAT
TTTTTACAAACCGGACAATCCAACGCTGCCGACCGAAAAGACACCCCTTCTTCGATCTCACG
TATATCCAGAAAACCGAGCCACAGAAGTACTCTGCACACCAAGAATAATGATAGGCAAC
ATACCTTTACCGTAGCTCCAATCTCTTTCCACGAAGCCCGTACAAATCAAATGCTGGAA
CATGCAAAGACGAAC TGCCAAAACATGACTTCAACCTCAAATTTGTCACCGAAAAGCTTG
ACGAGTGGCTCGTACCCAGAATTGGCTTACGTTATCGAGACCTTAGTGCATGGAAAACAAG
CATGCTTTTATAATCTTAAAACAAGTAATTAGCCAGTGTATTTCTTATTGGTTTAAACATG
AAACATATACTTCTTTTTTTGCAACAGCTTCGAGATGATCATGCTAATCTCTAACCACATG
GGCGACGAAC TGATCCCGGAGCAGCCTTTGCTCACTACAGAAATCTTGTGTGCGTAATT
CGCCTAGTGAAGAGAACAATCTCTATCTCCAACCTCAACGAGCAACTTTGCGGCGAACCT
CTGGTGAATTTGCGCAACGCGTTGTTTCGACGGACGTCTGTTCTGCCCGTTTCGTCCATACC
ATGCCAGAAAACGACACGAATGCAAAAATAACAGCGGATGATACACCCTGACACAGAAC
ACCGTAAGAGTTAGAAATTACGAAATATCCGATGTGCAAAGAATGAATCTAATAGACTCA
AGCGTCTGCTTTCACCGACAATGACAGACCATCGAACGAAACCACCATCCTGAGCAAGATC
TTTTACTTCTGCGTACTCCCGGCATATCAAATAACAAGGCCTGTGGCGCTGGCGTCAAC
GTAAAGGAAC TAGTTCTAGACTTATTCTACACGGAACCGTTTCATCAGTCCAGATGATTAT
TTCCAGGAGAATCCGATCACCAGCGACGTTCTAATGTCTCTGATCCGAGAAGGTATGGGC
CCTGGCTACACCGTAGCCAACACATCCTCTATCGAAAACAGCTGTTTAAATCGCTAATC
TACATTAATGAAAATACGAAAATACTGGAAGTGGAGGTCTCCTTAGATCCTGCGCAGCGA
CACGGCAACTCCGTTCAATTTTCAATCACTACAACACATTCTATACAACGGGCTTTGCCTG
ATCTCACCGATCACCACCTAAGACGGTACTATCAACCAATCCATTTTCATCGATTCTTC
TCCGACCCGGGAATCTGCGGCACCATGAATGCTGATATCCAAGTTTTCTAAATACATTT
CCTCACTGTCAAAGAACGACGGCGGTTTTCTCTCCCGCCCCATTAGCATTAGAATTT

TATAATTGGCAACGAACACCGTTTTCCGTGACTCAGCCTTCTGCCACAATTCCTGTG
AGCATTATGACGCTTGCCGCCATGCACTCAAAATTGTCTCCCGTTGCCATAGCGATCCAA
AGCAAAAACAAAATCCATCCGGGCTTTGCGGCCACACTAGTCCGGACGGATAATTTTCGAC
GTCGAGTGCCTATTATACAGTTCAGAGCAGCCACATCTATAATTTTAGACGATCCACG
GTCACCGCGGAAGCTAAAGATATCGCAACCCTTACAACCTTACCCAGCACCTAAGTTTT
GTAGATATGGGCTTAGGTTTTAGCTCTACCACCGCCACTGCCAATCTTAAGCGAATTTAAA
TCAGATATGGGGAGCAAGATACAAAACCTTTTTCTCCGCCTTCCCGATACACGCGTTTACC
AACACGGACATAAAATACGTGGATTCGGCATCACGTCCGGATAGAAAAACCTAATCCCTCC
GAGAGCGAAGCATAAACATCATAACATTCGGCGGAATTAACAAAAACCCACCTCCATA
CTACTGCATGGTCAACAAGCTATCTGCGAAGTTATACTGACCCCGGTCACGACAAAACATT
AACTTTTTCAAATCGCCCCACAACCAAGAGGCAGGGAATCGTGTATGATGGGAACGGAC
CCGCACAACGAAGAGGCGGCTAGAAAAGCATTGTACGACCACACCCAAACAGACAGCGAT
ACATTCGCCGCAACCACAAATCCTTGGGCATCCCTATCAGGCTCCTTAGGCGATATTCTA
TACAACACGGCACACAGAGAACAACCTATGTTACAACCCCAAGACATACAGTCTTAACGCT
CAATTTTTTACC GAATCTGACATCTTAAAAACAAAACAAGATGATGTACAAAGTGATAAGC
GAATACTGCATGAAATCGAATCGTGTTTAAACAGCGATAGCGAAATACAATACTCGTGC
TCTGAGGGCACGGATAGCTTCGTAAGCAGACCATGCCAGTTCTTACAAAACGCTCTGCCT
CTTACTGTTTCATCCAACCAAGCTCTATTAGAGAGTCGGTCTAAAACCGGCAATACGCAG
ATCAGCGAAACCCATTATTGTAATTACGCCATAGGAGAAACCATACCTTTCCAACCTCATT
ATCGAATCATCCATATAA

U69

>2284_U69

ATGGACAACGCTGTGGAGACACCTCAGGGTCAAAAACTCAGCCGATAAAATTTGCCACCA
GTCAGGAAAAAGTTAAGAAAACATGAGGGACTCGGAAAAGGTGTTAAACGAAAACCTTTTT
GCCGAAGATAGCTCTCCCTTAAAGAAAACAGAATTCGGCCTGCAGTGATATGGAAAACCTT
TCTTCGCCTGTAAAGTCTGGATGCGAGTCGCGAAGTGCTTCTCTCGATGAAAAGTTTCGGA
AAATGTA AACACGAAATTTGCTTGCATTGTTCTGCGATAGAGGAATTTGCTTTGTACGAG
TCGCTTCTAGACTCGCCGATGAAACTGTGCAATGCCACACCATCTTCAGCTCAGACAAA
TGGAACCTGGAGCTAGAGAAAATTATAGCTTCAAAGCAGATCTTTCTAGACATGAGTGAG
AATGCTGAACCTTGC GGCTACGGCGAAACCTTTGTGTAACCTGAGAATTTTCGAAAAGATC
AGCTCGCCGTTTTTTGTTTGATGTGCAAAGCGAAGAGCGTTTCGTATTCAGTGGTCTACGTC
CCTCACAAACAAAGA ACTCTGTGGACAGTTTTGTCAACCTGAGAAAACCTATGGCTCGAGTT
CTCGGAGTGGGTGCTTACGGGAAGGTGTTTGATCTAGATAAAAGTGGCCATAAAGACGGCC
AATGAAGATGAGAGTGTCAATTCGGCTTTCATAGCTGGTGTTCATCCGTGCAAAATCAGGA
GCCGACTTATTATCTCACGAGTGTGTTATTAATAACCTATTGATTTCAAATTCGGTTTTGT
ATGAGTCATAAAGTGTCTTTGTACGACTTATGATATTGATCTCCATAAGTTTCGAAGAT
TGGGATGTCAGGAATGTAATGAATTACTACAGTGTGTTTTGTAAGTTAGCTGATGCTGTA
AGTTTTCTAAATCTGAAATGTAGAATTAATCATTTTCGATATCTCACCTATGAATATATTT
TTAAATCATAAAAAAGAGATCATCTTTGATGCCGTGTTGGCGGATTACAGCTTGTCCGAG
ATGCATCCCAATTATAACGGCACGTGTGCTATTGCTAAAGAGTATGACAAAAATCTTCAA
CTTGTGCCAATTAGTCGTAACAAATCTGTGACATGTTTAATCCTGGATTTTCGACCACTT
GTCGCCAATGCAATGATATTGGTCAATGTGTGCGGGCTTTTTGATGGTGAAAATAATCCT
CTTAGACATTGTAACCTGGATCTGTGTGCTTTTTGCCAGGTTGTATTATCGTGTGTGCTA
AGAATGACAGATAAACGTGGATGCCGCGAAGCTCAGTTATACTACGAGAAAAGGTTGTTT
GCCTTGGCTAACGAGGCCCTGTCGATTGAATCCTCTCAAATATCCATTTGCTTACAGGGAT
GCTTGCTGTAAAGTATTGGCTGAGCATGTAGTATTACTAGGGTTATTGTTTTACCGAGAC
GTAGTTGAGATATATGAAAAACTATACGATTTTTCTAGATGAGAGAGGGGAATTTGGGTCA
CGAGACCTTTTTGAGGCAACTTTTTTAAATAATAGTAACTTACCAGACGTCAGCCAATC
AGAGAGGTTCTCGGCTCTTACAGTCCGAGTATGGAGAAAACCTTTTACATGACCTT
AGAGAGTTGTTCTTGATCAATCTACTGCGGATCTGGATAAAGATACATCATCTCTCTTT
CATATGTGA

>5055_U69

ATGGACAACGGTGTGGAGACACCTCAGGGTCAAAAACTCAGCCGATAAAATTTGCCACCT
GTCAGGAAAAAGTTAAGAAAACATGAGGGACTCGGAAAAGGTGTTAAACGAAAACCTTTTT

GCCGAAGATAGCCCTCCCTTAAAGAAACAGATTTCCGCCTGTAGTGATATGGAAACACTT
TCTTCGCCTGTAAAGTCTGGATGCGAGTCGCGAAGTGCTTCTCTCGATGAAAGTTTCGGA
AAATGTA AACACGAAATTTGCTTGCATTGTTCTGCGATAGAGGAATTGCTTTGTACAGAG
TCGCTTCTAGACTCGCCGATGAAACTGTGCAATGCCACACCATCTTCAGCTCAGACAAA
TGAAACTGGAGCTAGAGAAAATTATAGCTTCAAAGCAGATCTTTCTAGACATGAGTGAG
AATGCTGAAC TTGCGGCC TACGGCGAAACTTTGTGTAACCTGAGAATTTTCGAAAAGATC
AGCTCGCCGTTTTTTGTTTGATGTGCAAAGCGAAGAGCGTTTCGTATTCAGTGGTCTACGTC
CCTCACAAACAAAGA ACTCTGTGGACAGTTTTGTCAACCTGAAAAAACTATGGCTCGAGTT
CTCGGAGTGGGTGCTTACGGGAAGGTGTTTGATCTAGATAAAAGTGGCCATAAAGACGGCC
AATGAAGATGAGAGTGTCAATTTTCGGCTTTTCATAGCTGGTGTATCCGTGCAAAAATCAGGA
GCCGACTTATTATCTCACGAGTGTGTTATTAATAACCTATTGATTTCAAATTCGGTTTTGT
ATGAGTCATAAAGTGTCTTTGTACGTTACTTATGATATTGATCTCCATAAGTTTCGAAGAT
TGGGATGTCAGGAATGTAATGAATTACTACAGTGTGTTTTGTAAGTTAGCTGGTGTCTATA
AGGTTTCTAAATCTGAAATGTAGAATTAATCATTTTCGATATCTCACCTATGAATATCTTT
TTAAATCATAAAAAAGAGATCATCTTTGATGCCGTGTTGGCGGATTACAGCTTGTCCGAG
ATGCATCCTAATTATAACGGCACGTGTGCTATTGCTAAAGAGTATGACAAAAATCTTCAA
TTTGTGCCAATTAGTCGTAACAAATTTCTGTGACATGTTAATCCTGGATTTTCGACCACTT
GTCGCCAATGCAATGATATTGGTCAATGTATGCGGGGCTTTTGATGGTGAAAATAATCCT
CTCAGACACTGTAAC TTGGATCTGTGTGCTTTTGCCAGGTTGTATTATCGTGTGTGCTA
AGAATGACAGATAAACGTGGATGCCGCGAAGCTCAGTTATACTACGAGAAAAGGTTGTTT
GCCTTGCTAACGAGGCC TGTGATTGAATCCTCTCAAATATCCATTTGCTTACAGGGAT
GCTTGTGTAAAGTATTGGCTGAGCATGTAGTATTACTAGGGTTATTGTTTTACCGAGAC
GTGGTTGAGATATATGAAAACTATAACGATTTTCTAGATGAGAGAGGGGAATTTGGGTCA
CGAGACCTTTTTGAGGCACTTTTTTAAATAATAGTAACTTACCAGACGTCAGCCAATC
AGAAAGGTCTCGCGTCTCTACAGTCCGAGTATGGAGAAAACTTTTACATGACCTT
AGAGAGTTGTTCTTGATCAATTTCTACTGCGGATCTGGATAAAGATAACATCATCTCTCTTT
CATATGTGA

>5814_U69

ATGGACAACGCTGTGGAGACACCTCAGGGTCAAAAACTCAGCCGATAAAATTTGCCACCA
GTCAGGAAAAAGTTAAGAAAACATGAGGGACTCGGAAAAGGTGTTAAACGAAAACCTTTTT
GCCGAAGATAGCTCTCCCTTAAAGAAACAGAATTCGCCTGCAGTGATATGGAAACACTT
TCTTCGCCTGTAAAGTCTGGATGCGAGTCGCGAAGTGCTTCTCTCGATGAAAGTTTCGGA
AAATGTA AACACGAAATTTGCTTGCATTGTTCTGCGATAGAGGAATTGCTTTGTACAGAG
TCGCTTCTAGACTCGCCGATGAAACTGTGCAATGCCACACCATCTTCAGCTCAGACAAA
TGAAACTGGAGCTAGAGAAAATTATAGCTTCAAAGCAGATCTTTCTAGACATGAGTGAG
AATGCTGAAC TTGCGGCC TACGGCGAAACTTTGTGTAACCTGAGAATTTTCGAAAAGATC
AGCTCGCCGTTTTTTGTTTGATGTGCAAAGCGAAGAGCGTTTCGTATTCAGTGGTCTACGTC
CCTCACAAACAAAGA ACTCTGTGGACAGTTTTGTCAACCTGAAAAAACTATGGCTCGAGTT
CTCGGAGTGGGTGCTTACGGGAAGGTGTTTGATCTAGATAAAAGTGGCCATAAAGACGGCC
AATGAAGATGAGAGTGTCAATTTTCGGCTTTTCATAGCTGGTGTATCCGTGCAAAAATCAGGA
GCCGACTTATTATCTCACGAGTGTGTTATTAATAACCTATTGATTTCAAATTCGGTTTTGT
ATGAGTCATAAAGTGTCTTTGTACGTTACTTATGATATTGATCTCCATAAGTTTCGAAGAT
TGGGATGTCAGGAATGTAATGAATTACTACAGTGTGTTTTGTAAGTTAGCTGGTGTCTATA
AGGTTTCTAAATCTGAAATGTAGAATTAATCATTTTCGATATATCACCTATGAATATCTTT
TTAAATCATAAAAAAGAGATCATCTTTGATGCCGTGTTGGCGGATTACAGCTTGTCTGAG
ATGCATCCTAATTATAACGGCACGTGTGCTATTGCAAAGAGTATGACAAAAATCTTCAA
TTTGTGCCAATTAGTCGTAACAAATTTCTGTGACATGTTAATCCTGGATTTTCGACCACTT
GTCGCCAATGCAATGATATTGGTCAATGTATGCGGGGCTTTTGATGGTGAAAATAATCCT
CTCAGACACTGTAAC TTGGATCTGTGTGCTTTTGCCAGGTTGTATTATCGTGTGTGCTA
AGAATGACAGATAAACGTGGATGCCGCGAAGCTCAGTTATACTACGAGAAAAGGTTGTTT
GCCTTGCTAACGAGGCC TGTGATTGAATCCTCTCAAATATCCATTTGCTTACAGGGAT
GCTTGTGTAAAGTATTGGCTGAGCATGTAGTATTACTAGGGTTATTGTTTTACCGAGAC
GTGGTTGAGATATATGAAAACTATAACGATTTTCTAGATGAGAGAGGGGAATTTGGGTCA
CGAGACCTTTTTGAGGCACTTTTTTAAATAATAGTAACTTACCAGACGTCAGCCAATC
AGAAAGGTCTCGCGTCTCTACAGTCCGAGTATGGAGAAAACTTTTACATGACCTT
AGAGAGTTGTTCTTGATCAATTTCTACTGCGGATCTGGATAAAGATAACATCATCTCTCTTT
CATATGTGA

U70

>2284_U70

ATGGATCTTGATCAAATATCTGAAACACTGAGTTCTGTGGCCGAAGAAGAGCCTTTAACC
ATGTTTTTACTTGATAAACTGTATGCAATACGGGAAAAGATCAAGCAAGTTCCATTTTTCG
ATTGTTTCGCTTGTGTTCATGTTTACTGCATGCTAATAAAAATATAACGCTTCTAACAATAAT
TGCATTCTGGGCCGTAAACTTATTGAGGAAATGCAGCAGTTTTTTGTGCGGCACAAGAGTG
GATGGATCGGAAGACATTTCTATGGATCTGAGTGAATTGTGCAAGCTCTATGATTACTGT
CCATTATTGTGTTCTGCTCTGTGTCTGTCGCGCTTGTGTATCTGTGAACAAGTTATTTAAA
ATCGTAGAGCGTAAACTCGGGGGCAGTCGGAAAATCCCCTCTGGCACGCGTTACGGAAA
TATACCGTGACGGCGACTAAGTTGTATGACATCTATACGACAAGAAGTTTTTTAGAGTAT
AAAGGACAGCAGTTTTTTGGGGAAGCGGTGATTTACGGCGCGAAACATGAGCGTGTTATT
AGACACCTCGTAGCGACCTTTTACGTTAAAAGGGAAGTTAAGGAAACGTTGGGATTGCTG
CTTGATCCTTCTTCTGGAGTTTTTGGTGCCTCTTAGACGCGTGTTTTGGAATTTTCATTC
AATGAGGATGGATTCTTGATGGTTAAGGAAAAAGCGCTGATTTTTGAGATTAAATTTCAA
TATAAATATTTACGGGATAAAGAAGATCACTTTGTTTCTGAACTATTAAAAAACCCACG
GAGAAATCCTTTTCGGATTTTATTTTATCTCATCCGGTGCCTGTCATAGAGTTTTCGAGAA
AGAGGGAAGATTCCGTCATCTAGAGAATATTTAATGACGTATGATTTTCAATATCGTCCT
CAGAGAAAGTTGCGTACTTGCCCCACTCCAGCTATTTTGGCACCTCATATCAAACAACG
CTGTGTTTGAACGAAACACAGAAGTCTACGGTAATCGTTTTTTGATTGTAAGAGCGACTTG
TGTGAGCAGAAGCTGTCTGTGTTTTCAGAAGGCTGTGTTTACTGTGAATGTATTTCGTAAAT
CCGAAACACAGGTATTTTTTTTTCAGAGTCTGTTACAACAATATGTAATGACTCAGTTTTAT
ATTAATGATCATAATAATCCAGAATATATCGAAAGTACGGAAGTGCCTTCTGTTACATT
GTGACGGCTTTCTTTCAGGAGAAGAACAGAGGAAGAAAGGAGTTTGCATTTGGTGATTGAT
GAGACGGAATATATAGAAGAAGAAATACCTTTGGCCTTGATTGTGACTCCGGTAGCACCG
AATCCGGAATTTACTTGTGTTGTTATAACAGACATATGCAATCTGTGGGAAAATAATATT
TGCAAGCAGACCAGTTTACAGGTATGGGCGCAAAGTGCTGTAAACCAGTATCTTGGCGCA
TGTGTAAGAAAACCGAAAACACCTTGA

>5055_U70

ATGGATCTTGATCAAATATCTGAAACACTGAGTTCTGTGGCCGAAGAAGAGCCTTTAACC
ATGTTTTTACTTGATAAACTGTATGCAATACGGGAAAAGATCAAGCAAGTTCCATTTTTCG
ATTGTTTCGCTTGTGTTCATGTTTACTGCATGCTAATAAAAATATAACGCTTCTAACAATAAT
TGTATTCTGGGCCGTAAACTTATTGAGGAAATGCAGCAGTTTTTTGTGCGGCACAAGAGTG
GATGGATCGGAAGACGTTTCTATGGATCTGAGTGAATTGTGCAAGCTGTATGATTACTGT
CCATTATTGTGTTCTGCTCTGTGTCTGTCGCGCTTGTGTGTCTGTGAACAAGTTATTTAAA
ATCGTAGAGCGTAAACTCGGGGGCAGTCGGAAAATCCCCTCTGGCACGCGTTACGGAAA
TATACCGTGACGGCGACTAAGTTGTATGACATCTATACGACAAGAAGTTTTTTAGAGTAT
AAAGGACAGCAGTTTTTTGGGGAAGCGGTGATTTACGGCGCGAAACATGAGCGTGTTATT
AGACACCTCGTAGCGACCTTTTACGTTAAAAGGGAAGTTAAGGAAACGTTGGGATTGCTG
CTTGATCCTTCTTCTGGAGTTTTTGGTGCCTCTTAGACGCGTGTTTTGGAATTTTCATTC
AATGAGGATGGATTCTTGATGGTTAAGGAAAAAGCGCTGATTTTTGAGATTAAATTTCAA
TATAAATATTTACGGGATAAAGAAGATCACTTTGTTTCTGAACTATTAAAAAACCCACG
GAGAAATCCTTTTCGGATTTTATTTTATCTCATCCGGTGCCTGTCATAGAGTTTTCGAGAA
AGAGGGAAGATTCCGTCATCTAGAGAATATTTAATGACGTATGATTTTCAATATCGTCCT
CAGAGAAAGTTGCGTACTTGCCCCACTCCAGCTATTTTGGCACCTCATATCAAACAACG
CTGTGTTTGAACGAAACACAGAAGTCTACGGTAATCGTTTTTTGATTGTAAGAGCGACTTG
TGTGAGCAGAAGCTGTCTGTGTTTTCAGAAGGCTGTGTTTACTGTGAATGTATTTCGTAAAT
CCGAAACACAGGTATTTTTTTTTCAGAGTCTGTTACAACAATATGTAATGACTCAGTTTTAT
ATTAATGATCATAATAATCCAGAATATATCGAAAGTACGGAAGTGCCTTCTGTTACATT
GTGACGGCTTTCTTTCAGGAGAAGAACAGAGGAAGAAAGGAGTTTGCATTTGGTGATTGAT
GAGACGGAATATATAGAAGAAGAAATACCTTTGGCCTTGATTGTGACTCCGGTAGCACCG
AATCCGGAATTTACTTGTGTTGTTATAACAGACATATGCAATCTGTGGGAAAATAATATT
TGCAAGCAGACCAGTTTACAGGTATGGGCGCAAAGTGCTGTAAACCAGTATCTTGGCGCA
TGTGTAAGAAAACCGAAAACACCTTGA

>5814_U70

ATGGATCTTGATCAAATATCTGAAACACTGAGTTCTGTGGCCGAAGAAGAGCCTTTAACC
ATGTTTTTACTTTGATAAACTGTATGCAATACGGGAAAAGATCAAGCAAGTTCCATTTTCG
ATTGTTTCGCTTGTGTATGTTTACTGTCATGCTAATAAAATATAACGCTTCTAACAATAAT
TGTATTTCTGGGCCGTAAACTTATTGAGGAAATGCAGCAGTTTTTGTGCGGCACAAGAGTG
GATGGATCGGAAGACGTTTCTATGGATCTGAGTGAATTGTGCAAGCTGTATGATTACTGT
CCATTATTGTGTTCTGCTCTGTGTCGTGCGCCTTGTGTGTCTGTGAACAAGTTATTTAAA
ATCGTAGAGCGTAAACTCGGGGGCAGTCGGAAAATCCCCTCTGGCACGCGTTACGGAAA
TATACCGTGACGGCGACTAAGTTGTATGACATCTATACGACAAGAAGTTTTTTAGAGTAT
AAAGGACAGCAGTTTTTTGGGGAAGCGGTGATTTACGGCGCGAAACATGAGCGTGTTATT
AGACACCTCGTAGCGATCTTTTACGTTAAAAGGGAAGTTAAGGAAACGTTGGGATTGCTG
CTTGATCCTTCTTCTGGAGTTTTTGGTGCCTCTTAGACGCGTGTTTTGGAATTTTCATTC
AATGAGGATGGATTTTTGATGGTTAAGGAAAAGCGCTGATTTTTGAGATTAAATTTCAA
TATAAATATTTACGGGATAAAGAAGATCACTTTGTTTCTGAACTATTAAAAAACCCACG
GAGAAATCCTTTTCGGATTTTATTTTATCTCATCCGGTGCCTGTCATAGAGTTTTCGAGAA
AGAGGGAAGATTCGGTCACTTAGAGAATATTTAATGACGTATGATTTTCAATATCGTCCT
CAGAGAAAGTTGCGTACTTGCCCACTCCAGCTATTTTGGCACCTCATATCAAACAACCTG
CTGTGTTTGAACGAAACACAGAAGTCTACGGTAATCGTTTTTATTGTAAGAGCGACTTG
TGTGAGCAGAAGCTGCTGTGTTTTCAGAAGGCTGTGTTTACTGTGAATGTATTTCGTAAT
CCGAAACACAGGTATTTTTTTTTCAGAGTCTGTTACAACAATATGTAATGACTCAGTTTTAT
ATTAATGATCATAATAATCCAGAATATATCGAAAGTACGGAAGTGCCTTCTGTTACATT
GTGACGGCTTTCTTTCAGGAGAAGAACAGAGGACGAAAGGAGTTTGCATTTGGTGATTGAT
GAGACGGAATATATAGAAGAAGAAATACCTTTGGCCTTGATTGTGACTCCGGTAGCACCG
AATCCGGAATTTACTTGTGTTTATAACAGACATATGCAATCTGTGGGAAAATAATATT
TGCAAGCAGACCAGTTTACAGGTATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCA
TGTGTAAGAAAACCGAAAACACCTTGA

U71

>2284_U71

ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAACACC
TTGATTGATTACAAGGGAAATCCTATATCACTAGCGGATGAATTTACTGTTTTAACCGAT
ACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAGGCCCTTACTTGAAAAAGTACAG
ACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAATCAAAA
AAATGA

>5055_U71

ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAACACC
TTGATTGATTACAAGGGAAATCCTATATTACTAGAGAATGAATTTACTGTTTTAACCGAT
ACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAGGCCCTTACTTGAAAAAGTACAG
ACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAATAA

>5814_U71

ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAACACC
TTGATTGATTACAAGGGAAATCCTATATTACTAGAGAATGAATTTACTGTTTTAACCGAT
ACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAGGCCCTTACTTGAAAAAGTACAG
ACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAATAA

U72

>2284_U72

ATGGCCTCCAGTCGTGTAGACACCATTAATTTACGGATCTGGCTGGTCTCTATAATCTGT
GCAGCCTTGTCAATTCATAAACGTCACGGTCCATCTGATAGCCATAAACTTTCCAAATTTA
GGATTTCCGTGCGCCTACTTTGAAATTAATGACCTCAAAGCGGTTAATCTCTCCGCTAAC
AATCAAATATATCAGATGACGCATCACTTTACATAAAACCCCGTGCAGATTATATGCTAC
GTACTGATAATGGCTATGTTATTCCTCTGATTATTATTTATTACATTGTCTGTGCA
AAAGTGTCTCTCTAACAAAACATCTAACGTTAACCAAACAACGAGGGATATTACCTGG

ATGGGGGACACGTCATCCTGCTTTCAATTCATCTTAATCATGGATACATTTCAATTATTC
GTCACCGCCTTGCTTTTCAGACTGGTAGCCCTAGGCGCATTTCCTATTGCATCTTCTTC
GTGTGTTTCACCACCTTTAACGTTACGCTGATAACGCAATTTCAATCAGCGGATAAGTCA
TTTTTCGCCTTTTCAGAAAATTCATCCTAACCTAAAGGGGACCGTACAATTCAGACCGTT
GTTATCAATCTAACCGAATCATGCTAGGTTACTCTACAATGTTTCTCGGCATTACTACT
TGCTTGGGCGTCGGAAATAGTATCTACATACGTTCCATAACCGTTGCTTACTCTTCCATC
AATACCTTTCTTGTCATGGCTTGTCATATATAGTATAGTAATAGAAGCCGTCCTGGTTTCGC
TACGTTAAACCTTGTTTCGGCTATTACGTCGGTATGTTTTGTGGCGCCGTAGGCCTATCT
TTTCCATTCTCCAGTACGAAACTTTCTTTGAATCCGAGTGGAGCACAGGACTTATTATC
AATTTAGCGGTCATCGCCATAATTTTCGATCGGTTTTATAATTTGTGACTTGTACGCTAC
CTAGTTAAAAAGAAGCGTCGGTACAAGCAACTACTTAATACTGAAAGCTCGAGCCTAATG
GATGAAAACGAGTGA

>5055_U72

ATGGCCTCCAGTCGTGTAGACACCATTAATTTACGGATCTGGCTGGTCTCTATAATCTGT
GCAGCCTTGTCATTCATAAACGTCACGGTCCATCTGATAGCCATAAACTTTCCAAATTTA
GGATTTCCGTGCGCCTACTTTGAAATTAATGACCTCAAAGCGGTTAATCTCTCCGCTAAC
AATCAAATATATCAGATGACGCATCAACTTTACATAAAACCCCGTGCAGATTATATGCTAC
GTACTGATAATGGCTATGTTATTCCTCCTGATTATTATTTATTACATTGTCTGCTGTGCA
AAAGTGTTCCTCCTAACAAAACATCTAACGTTAACCAAACAACGAGGGATATTACCTGG
ATGGGGGACACGTCATCCTGCTTTCAATTCATCTTAATCATGGATACATTTCAATTATTC
GTCACCGCCTTGCTTTTCAGACTGGTAGCCCTAGGCGCATTTCCTATTGCATCTTCTTC
GTGTGTTTCACCACCTTTAACGTTACGCTGATAACGCAATTTCAATCAGCGGATAAGTCA
TTTTTCGCCTTTTCAGAAAATTCATCCTAACCTAAAGGGGACCGTACAATTCAGACCGTT
GTTATCAATCTAACCGAATCATGCTAGGTTACTCTACAATGTTTCTCGGCATTACTACT
TGCTTGGGCGTCGGAAATAGTATCTACATACGTTCCATAACCGTTGCTTACTCTTCCATC
AATACCTTTCTTGTCATGGCTTGTCATATATAGTATAGTAATAGAAGCTGTCCTGGTTTCGC
TACGTTAAACCTTGTTTCGGCTATTACGTCGGTATGTTTTGTGGCGCCGTAGGCCTATCT
TTTCCATTCTCCAGTACGAAACTTTCTTTGAATCCGAGTGGAGCACAGGACTTATTATC
AATTTAGCGGTCATCGCCATAATTTTCGATCGGTTTTATAATTTGTGACTTGTACGCTAC
CTAGTTAAAAAGAATCGTCGGTACAAGCAACTACTTAATACTGAAAGCTCGAGCCTAATG
GATGAAAACGAGTGA

>5814_U72

ATGGCCTCCAGTCGTGTAGACACCATTAATTTACGGATCTGGCTGGTCTCTATAATCTGT
GCAGCCTTGTCATTCATAAACGTCACGGTCCATCTGATAGCCATAAACTTTCCAAATTTA
GGATTTCCGTGCGCCTACTTTGAAATTAATGACCTCAAAGCGGTTAATCTCTCCGCTAAC
AATCAAATATATCAGATGACGCATCAACTTTACATAAAACCCCGTGCAGATTATATGCTAC
GTACTGATAATGGCTATGTTATTCCTCCTGATTATTATTTATTACATTGTCTGCTGTGCA
AAAGTGTTCCTCCTAACAAAACATCTAATGTTAACCAAACAACGAGGGATATTACCTGG
ATGGGGGACACGTCATCCTGCTTTCAATTCATCTTAATCATGGATACATTTCAATTATTC
GTCACCGCCTTGCTTTTCAGACTGGTAGCCCTAGGCGCATTTCCTATTGCATCTTCTTC
GTGTGTTTCACCACCTTTAACGTTACGCTGATAACGCAATTTCAATCAGCGGATAAGTCA
TTTTTCGCCTTTTCAGAAAATTCATCCTAACCTAAAGGGGACCGTACAATTCAGACCGTT
GTTATCAATCTAACCGAATCATGCTAGGTTACTCTACAATGTTTCTCGGCATTACTACT
TGCTTGGGCGTCGGAAATAGTATCTACATACGTTCCATAACCGTTGCTTACTCTTCCATC
AATACCTTTCTTGTCATGGCTTGTCATATATAGTATAGTAATAGAAGCTGTCCTGGTTTCGC
TACGTTAAACCTTGTTTCGGCTATTACGTCGGTATGTTTTGTGGCGCCGTAGGCCTATCT
TTTCCATTCTCCAGTACGAAACTTTCTTTGAATCCGAGTGGAGCACAGGACTTATTATC
AATTTAGCGGTCATCGCCATAATTTTCGATCGGTTTTATAATTTGTGACTTGTACGCTAC
CTAGTTAAAAAGAATCGTCGGTACAAGCAACTACTTAATACTGAAAGCTCGAGCCTAATG
GATGAAAACGAGTGA

U73

>2284_U73

ATGGAGAACAGGTTTCAGAACGATACTCTGTTCTCAGAGTGGTTCGGTCAGAGTTTGTCC

GATGTAAGATTTCCGGACAATGTCACCGTTTATTCTCAAGCCGACTCGGCGGTGAGTTTT
GAAAACGTCAGACAGCCGATAAAACTGGTTCGCGCAGCGATGGGTTCAGGAAAAACA
GCCTTGATTCATTTTTTGAACAGGTACCCAGGAACTATCTGTCCTTTTGATTTCTTGC
CGTAAAACGTTTCGCTGCTGAAATTTTGCACCGTTTTACTCTCAATGGGTGAAGGATTTT
GAATTATATTGCGATATTACGGAGCGGCAGATAAAACAACAGAAAAGTCATTGTCCAGATA
GAAAGCTTGCATCGTCTTACAGAAAATTACGATGTTTTAATATTAGATGAAATTATGTCTG
ATTATAAAGCAGTTTTATTCCAAAACGATGACAAAAACCAAAGAAGTTGATTGTAAGTTT
TTATCGCTGATCAAAAATTCCAGTCATGTAATAGCCATGGATGCGACGCTGACTCGGCAT
GTGGTGGAGTTTTTTGCCGCGTTCAAGCCAGATACTCAAATCGCCCTCATCAGAAATACA
TTTGTGTCCGCGATTTTTCCAATAGGATCGCTTACTTTTTGTGACACGTTTTTTGGTAAA
GAGTTTTCTTTTTTGCACGCTTGAAGACAAGCTACGATGGGATAAGAAATTGTGTTTG
TTTTGCAGCACAGTGTGGCAGCGGAATACATGCATGATTTGATACGGTCTCGGTTTCCA
CTCAAAAAGTTCCTTGTGTTGACGCTCAAACAAGGGAAATGCTCTTCCATAGAATCGTGG
ATACGGTATGACGTAGTGATTTATACGTCGGTCGTGACTGTGCGTCTCAGTTTCGAACCG
GTTTATTTTCAGCTCTTTGTTTGTTCATTTAGTTGGCCAAAGGAGGGCCGGACATGGTT
TCAATATTTCAATCCATTGGTCTGTGAGGCGAGTTATCGATGAAGATATATACATTTAC
ATGAATCCCGTATTAATTAGAAGCTATGATCCTTTGGCACCTATTGCCATACCTCCGTC
AGTGATTGGAGTGTGCTGAGCAGTCTATCATCTCAGAGAGTTGCATCGACTTCCGCGGA
AAGTGCAGCGGAGCGCATAAATAAATTTTTGTTCCGTAAGTCTCTTTTTTCGGTAC
AGGCATTATATCGAAAAGACTACTATTACAAGTTTATCTGATAGTTTATTTCTGCTATGC
AGTCTGTTATGTGAGAAATTCATTAAGGTGGACATCGTTGGAAATGGATTCCCGATGCGC
AAAGAGTTTTTCTGAGTTTCTTGCAGATTCTAGTTGAGGAGTGTCATTTTATCGAAAAG
AAGATTACTTTGCCGGGTGACGACATGACGTTCCAAGAGATTATAAGTAGCAGAGAACT
ATCATGAACGGGGATTTTTATGAAAACGGTGATCAACTATTACATAAGGACTATATTACC
GACATGGGAAAGTTCAGAGCGACTTTCTTATCTCCGGGCGTTGACATTTTCATCGCATCG
GACATTGTATATGATCTAAAAAATGAAAGCAAACGCTATGTTTTTGTGAACGCTGTTG
CAAAAATGTGTATCTGCCGGTGTGGAAGCACGGAAATCGAAAGAGTTTTTTGTGAGCGG
ATTAAGTGCTATATGCTGCCGAAGAGTTTTCTGTGCGACGAGTACTTTGTTCTAGGTGAC
ATTTCTGGGTTTACGAGTGGGGAATGTTAATAGATCTGGCCTTCTGGCGGAAATGATC
CGAAAAGATTTGAACTGAAATCGTGCACGGATACTACTACTGACATTTCTGAAGATGAT
CTCCTTCTGTGTGCGGCCAGGAGGTCAAGTGACATCTTACAAATAATGCAATTAGTGT
ACGGTGCATGTGCAATTTTTTCAAAAATACAGTCTGCAGACTCTGCAGCTTTTTTAACAAG
TTGAGAGGCATGCGGATTGTGACGGGTGTGTTCTCTATAGAAAATTCAGCATTTCCATC
CTCAGGCTTTTCTTTAAATGTGCATTTAATATGACTCTGTCCGCCAGCAGACCTCGGTAC
ATTCCCGGAAGGCGTATCGCAATCTAACAAAAAATGATATGGAGAACATGCTAGACAAT
TGGGAGATTTTCGCGCACCAATCTTAAGACTTGCAAAGAACTACGTAAAGCCCTCACCGAG
GCCTCGAGAGCAAGCGGAAACAACAATTTATAAACTGCAAGGCTCGGATATAAGCCTC
TCCGTTAGTGAAGTTGGAGTTTTTCGGCAACATGCATCTCCGGGGTGTGTGTGTCATCT
TAG

>5055_U73

ATGGAGAACAGGTTTCAGAACGATACTCTGTTCTCAGAGTGGTTCGGTCAGAGTTTGTCC
GATGTAAGATTTCCGGACAATGTCACCGTTTATTCTCAAGCCGACTCGGCGGTGAGTTTT
GAAAACGTCAGGCAGCCGATAAAACTGGTTCGCGCAGCGATGGGTTCAGGAAAAACA
GCCTTGATTCATTTTTTGAACAGGTACCCAGGAACTATCTGTCCTTTTGATTTCTTGC
CGTAAAACGTTTCGCTGCTGAAATTTTGCACCGTTTTACTCTCAATGGGTGAAGGATTTT
GAATTATATTGCGATATTACGGAGCGGCAGATAAAACAACAGAAAAGTCATTGTCCAGATA
GAAAGCTTGCATCGTCTTACAGAAAATTACGATGTTTTAATATTAGATGAAATTATGTCTG
ATTATAAAGCAGTTTTATTCCAAAACGATGACAAAAACCAAAGAAGTTGATTGTAAGTTT
TTATCGCTGATCAAAAATTCCAGTCATGTAATAGCCATGGATGCGACGCTGACTCGGCAT
GTGGTGGAGTTTTTTGCCGCGTTCAAGCCAGATACTCAAATCGCCCTCATCAGAAATACA
TTTGTGTCCGCGATTTTTCCAATAGGATCGCTTACTTTTTGTGACACGTTTTTTGGTAAA
GAGTTTTCTTTTTTGCACGCTTGAAGACAAGTTACGATGGGATAAGAAATTGTGTTTG
TTTTGCAGCACCGTGTGGCAGCGGAATACATGCATGATTTGATACGGTCTCGGTTTCCA
CTCAAAAAGTTCCTTGTGTTGACGCTCAAACAAGGGAAATGCTCTTCCATAGAATCGTGG
ATACGGTATGACGTAGTGATTTATACGTCGGTCGTGACTGTGCGTCTCAGTTTCGAACCG
GTTTATTTTCAGCTCTTTGTTTGTTCATTTAGTTGGCCAAAGGAGGGCCGGACATGGTT
TCAATATTTCAATCCATTGGTCTGTGAGGCGAGTTATCGATGAAGATATATACATTTAC

ATGAATCCCGTATTAATTAGAAGCTATGATCCTTTGGCACCTATTGCCATACCTCCGTGC
AGTGATTGGAGTGTGCGTGTGAGCAGTCTATCATCTCAGAGAGTTGCATCGACTTCCGCGGA
AAGTGCAGCGGAGCGCATAAATAAATTTCTGTTCCGTAAGTCTCTTTTTTCGGTAC
AGGCATTATATCGAAAAGACTACTATTACAAGTTTATCTGATAGTTTATTTCTGCTATGC
AGTCTGTTATGTGAGAATTCATTAAGGTGGACATCGTTGGAAATGGATTCCCGATGCGC
AAAGAGGTTTTCTGAGTTTCTTGAGATTCTAGTTGAGGAGTGTCAATTTTATCGAAAAG
AAGATTACTTTGCCGGGTGACGACATGACGTTCCAAGAGATTATAAGTAGCAGAGAACT
ATCATAAACGGGGATTTTTATGAAAACGGTGATCAACTATTACATAAGGACTATATTACC
GACATGGGAAAGTTCAGAGCGACTTTTTTATCTCCGGGCGTTGACATTTTCATCGCATCG
GACATTGTATATGATCTAAAAAATGAAAGCAGACGCTATGTTTTTGTGAACGTCTGGTTG
CAAAAATGTGTGCTGCCGGTGTGGAAAGCACGGAAATCGAAAGAGTTTTTTGTGAGCGG
ATTAAGTGCATATGCTGCCGAAGAGTTTTCTGTGCGACGAGTACTTTGTTCTAGGTGAC
ATTTCTGGGGTTTATGAGTGGGGAATGTTAATAGATCTGGCCTTCTGGCGGAAATGATC
CGAAAAGATTTGAACTGAAATCGTGCACGGATACTACTACTGACATTTCTGAAGATGAT
CTCCTTCTGTGTGCCGCCAGGAGGTCAAGTGACATCTTACAAATAATGCAATTAGTGTTT
ACGGTGCATGTGCAATTTTTTCAAAAATACAGTCTGCAGACTCTGCAGCTTTTTTAACAAG
TTGAGAGGCATGCGGATTGTGACGGGTGTGTTCTCTATAGAAAATTCAGCATTTCCATC
CTCAGGCTTTTCTTTAAATGTGCATTTAATATGACTCTGTCCGCCAGCAGACCTCGGTAC
ATTCCCGGGAAGGCGTATCGCAATCTAACAAAAAATGATATGGAGAACATGCTAGACAAT
TGGGAGATTTGCGGCACCAATCTTAAGACTTGCAAAAGAACTACGTAAAGCCCTCACCGAG
GCCTCGAGAGCAAGGCGGAAACAACAATTTATAAACTGCAAGGCTCGGATATAAGCCTC
TCCGTTAGTGAAGTTGGAGTTTTTCGGGCAACATGCATCTCCGGGGGTGTGCGTGTCACT
TAG

>5814_U73

ATGGAGAACAGGTTTCAGAACGATACTCTGTTCTCAGAGTGGTTCGGTCAGAGTTTGTCC
GATGTAAGATTTCCGGACAATGTCACCGTTTATTCTCAAGCCGACTCGGCGGTGAGTTTT
GAAAACGTCAGGCAGCCGATAAAACTGGTTCGCGCAGCGATGGGTTCAGGAAAAACAACA
GCCTTGATTCAATTTTTTGAACAGGTACCCAGGGAACATCTGTCTTTTTGATTTCTTGC
CGTAAAACGTTTCGCTGCTGAAATTTTTGCACCGTTTTACTCTCAATGGGTTGAAGGATTTT
GAATTATATTGCGATATTACGGAGCGGCAGATAAAACAACAGAAAAGTCATTGTCCAGATA
GAAAGCTTGCATCGTCTTACAGAAAATTACGATGTTTTAATATTAGATGAAATTATGTCG
ATTATAAAGCAGTTTTTATTCAAAACGATGACAAAACCAAAGAAGTTGATTGTAAGTTT
TTATCGTGTATCAAAAATTCAGTCATGTAATAGCCATGGATGCGACGCTGACTCGGCAT
GTGGTGGAGTTTTTTGCCGCTTCAAGCCAGATACTCAAATCGCCCTCATCAGAAATACA
TTTGTGTCCGCAATGTTTTCCAATAGGATCGCTTACTTTTTGTGACACGTTTTTTAGTAAA
GAGTTTTCTTTTTTGCACGCTTGAAGACAAGTTACGATGGGATAAGAAAATTGTGTTT
TTTTGCAGCACCGTGTGGCAGCGGAATACATGCATGATTTGATACGGTCTCGGTTTCCA
CTCAAAAAGTTCTCTTGTGAGCTCTAAACAAGGGAATGCTCTTCCATAGAATCGTGG
ATACGGTATGACGTAGTGATTTATACGTCCGTCGTGACTGTGCGTCTCAGTTTCGAACCG
GTTTATTTTCAGCTCTTTGTTTGTTTACATTCAGTTGGCCAAAGGAGGGCCGGACATGGTT
TCAATATTTCAATCCATTGGTGTGTGAGGCGAGTTATCGATGAAGATATATACATTTAC
ATGAATCCCGTATTAATTAGAAGCTATGATCCTTTGGCACCTATTGCCATACCTCCGTGC
AGTGATTGGAGTGTGCGTGTGAGCAGTCTATCATCTCAGAGAGTTGCATCGACTTCCGCGGA
AAGTGCAGCGGAGCGCATAAATAAATTTCTGTTCCGTAAGTCTCTTTTTTCGGTAC
AGGCATTATATCGAAAAGACTACTATTACAAGTTTATCTGATAGTTTATTTCTGCTATGC
AGTCTGTTATGTGAGAATTCATTAAGGTGGACATCGTTGGAAATGGATTCCCGATGCGC
AAAGAGGTTTTCTGAGTTTCTTGAGATTCTAGTTGAGGAGTGTCAATTTTATCGAAAAG
AAGATTACTTTGCCGGGTGACGACATGACGTTCCAAGAGATTATAAGTAGCAGAGAACT
ATCATAAACGGGGATTTTTATGAAAACGGTGATCAACTATTACATAAGGACTATATTACC
GACATGGGAAAGTTCAGAGCGACTTTTTTATCTCCGGGCGTTGACATTTTCATCGCATCG
GACATTGTATATGATCTAAAAAATGAAAGCAGACGCTATGTTTTTGTGAACGTCTGGTTG
CAAAAATGTGTGCTGCCGGTGTGGAAAGCACGGAAATCGAAAGAGTTTTTTGTGAGCGG
ATTAAGTGCATATGCTGCCGAAGAGTTTTCTGTGCAACGAGTACTTTGTTCTAGGTGAC
ATTTCTGGGGTTTATGAGTGGGGAATGTTAATAGATCTGGCCTTCTGGCGGAAATGATC
CGAAAAGATTTGAACTGAAATCGTGCACGGATACTACTACTGACATTTCTGAAGATGAT
CTCCTTCTGTGTGCCGCCAGGAGGTCAAGTGACATCTTACAAATAATGCAATTAGTGTTT
ACGGTGCATGTGCAATTTTTTCAAAAATACAGTCTGCAGACTCTGCAGCTTTTTTAACAAG

TTGAGAGGCATGCGGATTGTGACGGGTGTGTTCTCTATAGAAAAATTCAGCATTTCCATC
CTCAGGCTTTTTCTTTAAATGTGCATTTAATATGACTCTGTCCGCCAGCAGACCTCGGTAC
ATTCCCAGGAAGGCGTATCGCAATCTAACAAAAAATGATATGGAGAACATGCTAGACAAT
TGGGAGATTTTCGCGCACCAATCTTAAGACTTGCAAAAGAACTACGTAAAGCCCTCACCGAG
GCCTCGAGAGCAAGGCGGAAACAACAATTTATAAACTGCAAGGCTCGGATATAAGCCTC
TCCGTTAGTGAAGTTGGAGTTTTTCGGGCAACATGCATCTCCGGGGGTGTGTGTGCATCT
TAG

U74

>2284_U74

ATGCATCTCCGGGGGTGTGTGTGCATCTTAGCCTTTATTGTGTGTACAATGACTGGGAG
AAAAAGATCTACAGAGTACCGATTTTTTCAGTGTCTGTTTTTAGAAGCGGAGACTCGTTTCG
TTAAAAACGTTTTTGTATTAGGGGACAGAGCCTAGACCAAGAATCGTTAAATCAGATAGAA
GTTACAAGAAAAGAAACAATGCTTTGGGACCTTCAGGAACAAAGTAATATGATGGACAAA
AAAATAGCAGCAATCAGCAACCTCATTATGAATAACGGAGAGCTTGTTAGGAAACTTTCA
AAATTTTTTCGTGCCCTTAACCGTTGTTCTTGGAGATGACGGGTAGAGATTTTTGGAAGCA
TACGTCTGTGGCGAGGAACCATGTTGCCTCTGGACACGGTACCTGTTATCTTACGGTGT
ATCGGAGACTACGCGGCCCTTGGACACGAAGCATCTTTTGGAGTAACGAATGTACACAGGCT
TCCAAAAAATTCGCTTTGGATACAGTGTGATGGACTTTCACTTTTCGCTGACTGTGTCT
GACGTGAAAAATTTGTTTTTTCACACACGGACACCGGTGAGGCCGTATGTGAAAAAATGAAA
CAAATTTTTTCTTTTTCTGTGTGTGCTTTTCGGCGGTGAGCAAGTTCTCTTAGTGACGCCA
AAAAATGCTTATGCCCTTATTGTTTCGACGATGATTTGTGTTTGCTTTTACTGCAGAGCGTG
TTTGCTTTTCTGCACGATAAGATATTCGGAGTTTTATAAACAGGTGTTAGTTTCAGCTTTGT
GAATACATCGGTCCCGACTTATGGCCCTTTGGGAATGAAAGATCAGTTTCTTTTATAGGA
TATCCAAATCTGTGGCTTTTGTCTGTATCGGACTTAGAAAGGCGGGTTCCGGATACAAC
TATATCTGTGCGAAATCTCTCATTTTTGCGGTTTTGGCGCCTATCCTGGGACCTCGTGGT
AGACACGTGGTCCCCGTAGTCAGGGAGCTAAATATTGAAATGCCTGGAAGTGAACCTTCT
TTGCAGCGTTTTTAGGTTCAATAGTCAGTATGTTTCAAGCGAGTCCCTGTGTTTTTCAGACA
GGTCCCAGGACACACATCTATTCTTTAGTGACAGTGATATGTATGTAGTGACTTTGCCA
GATTGTCTGAGGCTATTATTGAAGTCCACGGTTTCTAAGGCCTTTTTACCATGTTTTGAT
GAAAACGCCACGGAATCAATCTGCTTTTAAAAATTTATGTCAAGGCTGCAACACAGATCG
TACGCTTTTATTCGATGCGGTGATCTTTATGTTAGATGCTTTTTGTTTTCCGCTTTTCAGCGA
GCATGTACCCTAATGGGAATGCGATGGCTGTTAGTGCGGGACTTACACATGTTTTATTTA
ACCTGTGACGGTAAAGATACACATGTTGTGCATGCCTTTACTGCAAACGGCCGTAGAAAAAC
TGTTGGGAAAAAATACGGAATCAAGCAAAGACCTACATTTTCAGTGTGCGGAAATTTTCG
CGTGTGGATTTATCGTCTACGCCAGATTTTTTTTTGAGCAGTGGGTTGTCACAATCTAAA
GAAGCGCACTGGACAGTAACGGCGAGTAAATATTTATCTGCATGTATTCGGACCAATCAG
ACAGGTCTTTGCTTTGCCAGTATAACTGTTTATTTTTCAGGATATGATGTGTGTTTTTATA
GCTAATAGGTATAATGTTTCTTATTGGATCGAAGAGTTTCGATCCTAATGATTATTGTCTG
GAATATCATGAAGGACTTCTGGACTGCAGTAGATATACGGCCGTGATGTCTGAGGATGGA
CAGCTTGTGAGACAAGCACGTGGAATTGCGTTGACTGACAAAATAAACTTTTTCTTATTAC
ATTCTCGTTACACTGAGAGTGTGAGGAGATGGGTGGAGAGTAAATTTGAAGACGTTGAA
CAGACACAGTTTATTAGGTGGGAAAACAGGATGCTCTGTGAACACATTCATTTACTGCAT
TTGAATTAA

>5055_U74

ATGCATCTCCGGGGGTGTGCGTGTGCATCTTAGCCTTTATTGTGTGTACAATGACTGGGAG
AAAAAGATCTACAGAGTACCGATTTTTTCAGTGTCTGTTTTTAGAAGCGGAGACTCGTTTCG
TTAAAAACGTTTTTGTATTAGGGGACAGAGCCTAGACCAAGAATCGTTAAATCAGATAGAA
GTTACAAGAAAAGAAACAATGCTTTGGGACCTTCAGGAACAAAGTAATATGATGGACAAA
AAAATAGCAGCAATCAGCAACCTCATTATGAATAACGGAGAGCTTGTTAGGAAACTTTCA
AAATTTTTTCGTGCCCTTAACCGTTGTTCTTGGAGATGACGGGTAGAGATTTTTGGAAGCA
TACGTCTGTGGCGAGGAACCATGTTGCCTCTGGACACGGTACCTGTTATCTTACGGTGT
ATCGGAGACTACGCGGCCCTTGGACACGAAGCATCTTTTGGAGTAACGAATGTACACAGGCT
TCCAAAAAATTCGCTTTGGATACAGTGTGATGGACTTTCACTTTTCGCTGACTGTGTCT
GACGTGAAAAATTTGTTTTTTCACACACGGACACCGGTGAGGCCGTATGTGAAAAAATGAAA

CAAATTTTTCTTTTTCTGTGTGTGCTTTCCGGCGGTGAGCAAGTTCTCTTAGTGACGCCA
AAAAATGCTTATGCCTTATTGTTTCGACGATGATTTGTGTTTGCTTTTACTGCAGAGCGTG
TTTGCTTTCCCTGCACGATAAGATATTCGGAGTTTATAAACAGGTGTTAGTTCAGCTTTGT
GAATACATCGGTCCCGACTTATGGCCCTTTGGGAATGAAAGATCAGTTTCTTTTATAGGA
TATCCAAATCTGTGGCTTTTGTCTGTATCGGACTTAGAAAGGCGGGTTCCGGATACAAC
TATATCTGTGCGGAAATTCCTCTCATTTTTGCGGTTTGGCGCCTATCCTGGGACCACGTGGT
AGACACGTGGTCCCCGTAGTCAGGGAGCTAAATATTGAAATGCCTGGAAGTGAACTTCT
TTGCAGCGTTTTCAGGTTCAATAGTCAGTATGTTTTCAAGCGAGTCCCTGTGTTTTTCAGACA
GGTCCCAGGACACACATCTATTCTTTAGTGACAGTGATATGTATGTAGTGACTTTGCCA
GATTGTCTGAGGCTATTATTGAAGTCCACGGTTCCTAAGGCCTTTTTACCATGTTTTGAT
GAAAACGCCACGGAAATCAATCTGCTTTTAAAATTTATGTCAAGGCTGCAACACAGATCG
TACGCTTTATTCGATGCGGTGATCTTTATGTTAGATGCTTTTTGTTTTCCGCTTTTCAGCGA
GCATGTACCCTAATGGGAATGCGATGGCTGTTAGTGCGGGACTTACACATGTTTTATTTA
ACCTGTGACGGTAAAGATACACATGTTGTATGCCTTTACTGCAAACGGCCGTAGAAAAC
TGTTGGGAAAAAATACGGAAATCAAGCAAAGACCTACATTTTCAGTGTGCGGAAATTTTCG
CGCTGTGGATTTATCGTCTACGAAAGATTTTTTTTTGAGCAGTGGGTTGTCACAATCTAAA
GAAGCGACTGGACAGTAACGGCGAGTAAATATTTATCTGCATGTATTCGGACCAATCAG
ACAGTCTTTGCTTTGCCAGTATAACTGTTTTATTTTCAGGATATGATGTGTGTTTTTATA
GCTAATAGGTATAATGTTTCTTATTGGATCGAAGAGTTCGATCCTAATGATTATTGTCTG
GAATATCATGAAGACTTCTGGACTGCAGTAGATATACGGCCGTGATGTCTGAAGATGGA
CAGCTTGTGAGACAAGCACGTGGAGTTGCGTTGACTGACAAAATAAACTTTTTCTTATTAC
ATTCTCGTTACACTGAGAGTGTGAGGAGATGGGTGGAGAGTAAATTTGAAGACGTTGAA
CAGACACAGTTTATTAGGTGGGAAAACAGGATGCTCTGTGAACACATTCATTTACTGCAT
TTGAATTAA

>5814_U74

ATGCATCTCCGGGGTGTGTGTGTCATCTTAGCCTTTATTGTGTGTACAATGACTGGGAG
AAAAAGATCTACAGAGTACCGATTTTTTCAGTGTCTGTTTTTAGAAGCGGAGACTCGTTCC
TTAAAAACGTTTTTGTATTAGGGGACAGAGCCTAGACCAAGAATCGTTAAATCAGATAGAA
GTTACAAGAAAAGAAACAATGCTTTGGGACCTTCAGGAACAAAGTAATATGATGGACAAA
AAAATAGCAGCAATCAGCAACCTCATTATGAATAACGGAGAGCTTGTAGGAAACTTTCA
AAATTTTTCGTGCCCTAACCCTTGTCTTGGAGATGACGGGTTAGAGATTTTGAAGCA
TACGCTGTGGCGAGGAATCATGTTGCCTCTGGACACGGTACCTGTTATCTTACGGTGT
ATCGGAGACTACGGCCCTTGGACACGAAGCATCTTTTGGTAACGAATGTACACAGGCT
TCCAAAAAATTCGCTTTGGATACAGTGTGATGGACTTTCACTTTTCGCTGACTGTGTCT
GACGTGAAAATTTGTTTTTACACACGGACACCGGTGAGGCCGTATGTGAAAAAATGAAA
CAAATTTTTCTTTTTCTGTGTGTGCTTTCCGGCGGTGAGCAAGTTCTCTTAGTGACGCCA
AAAAATGCTTATGCCTTATTGTTTCGACGATGATTTGTGTTTGCTTTTACTGCAGAGCGTG
TTTGCTTTCCCTGCACGATAAGATATTCGGAGTTTATAAACAGGTGTTAGTTCAGCTTTGT
GAATACATCGGTCCCGACTTATGGCCCTTTGGGAATGAAAGATCAGTTTCTTTTATAGGA
TATCCAAATCTGTGGCTTTTGTCTGTATCGGACTTAGAAAGGCGGGTTCCGGATACAAC
TATATCTGTGCGGAAATTCCTCTCATTTTTGCGGTTTGGCGCCTATCCTGGGACCTCGTGGT
AGACACGTGGTCCCCGTAGTCAGGGAGCTAAATATTGAAATGCCTGGAAGTGAACTTCT
TTGCAGCGTTTTTAGGTTCAATAGTCAGTATGTTTTCAAGCGAGTCCCTGTGTTTTTCAGACA
GGTCCCAGGACACACATCTATTCTTTAGTGACAGTGATATGTATGTAGTGACTTTGCCA
GATTGTCTGAGGCTATTATTGAAGTCCACGGTTCCTAAGGCCTTTTTACCATGTTTTGAT
GAAAACGCCACGGAAATCAATCTGCTTTTAAAATTTATGTCAAGGCTGCAACACAGATCG
TACGCTTTATTCGATGCGGTGATCTTTATGTTAGATGCTTTTTGTTTTCCGCTTTTCAGCGA
GCATGTACCCTAATGGGAATGCGATGGCTGTTAGTGCGGGACTTACACATGTTTTATTTA
ACCTGTGACGGTAAAGATACACATGTTGTATGCCTTTACTGCAAACGGCCGTAGAAAAC
TGTTGGGAAAAAATACGGAAATCAAGCAAAGACCTACATTTTCAGTGTGCGGAAATTTTCG
CGCTGTGGATTTATCGTCTACGCCAGATTTTTTTTTGAGCAGTGGGTTGTCACAATCTAAA
GAAGCGACTGGACAGTAACGGCGAGTAAATATTTATCTGCATGTATTCGGACCAATCAG
ACAGTCTTTGCTTTGCCAGTATAACTGTTTTATTTTCAGGATATGATGTGTGTTTTTATA
GCTAATAGGTATAATGTTTCTTATTGGATCGAAGAGTTCGATCCTAATGATTATTGTCTG
GAATATCATGAAGACTTCTGGACTGCAGTAGATATACGGCCGTGATGTCTGAGGATGGA
CAGCTTGTGAGACAAGCACGTGGAATGCGTTGACTGACAAAATAAACTTTTTCTTATTAC
ATTCTCGTTACACTGAGAGTGTGAGGAGATGGGTGGAGAGTAAATTTGAAGACGTTGAA

CAGACACAGTTTATTAGGTGGGAAAACAGGATGCTCTGTGAACACATTCATTTACTGCAT
TTGAATTAA

U81

>2284_U81

ATGGCCCTACTACAGTGGATGTTAGATCACGTTTCAGGATGAAGAAAAAACTGTGAAAAC
CTTTC AATAGATGATCAGCATTACACTCTTCGGCATAAACAGAGACTGGTTGAGTTTCTTG
CAACTCTCCAAGTTAGAAATTACACACCTTAAACATGTATATAAAATTGGTGGACAACGAT
AGGGCTCATCTAACGGTCCACCCCTCTTCGGGTAACGTCCACGCTTGGAGTTTTTTATGC
AAACCCGACGATGTAAAAGTTGTGATTTTGGGACAAGATCCGTATCCCGACGGCAGAGGT
TGCGGCTTGGCCCTTCGGTACAGTGAAGAGTGCTCGATTCCAGAATCTCTAAAGAATATA
TTTAAAGAACTGGAAAGAAGCGTGCCGAATTTTTCCCCCCTGACAACGGCTGTTTAAAC
TCCTGGTGTAGTGAAGGAGTCCATTGCTAAACTCGATATTCAGTGTAGTTCATGGATTA
CCAATGTCCCACGAGGCATTTGGCTGGCAAACACTGAGCTATAAGATTATCAGCAAACCTA
TCGGAAGAAATGGACTCTCTCGTTTTCTTATTGTGGGAAATCATGCCCGGAAACTCTCT
TATCTAATAGACGCACGGAAACATCTCGTCTTAGAAAGTGCACATCCATCACCCAAAGTG
AAATCTGCAAGGATGCCATTTATTGGTTGCAATCATTTCGTGCGAGCAAATTTATTTCTT
ACGGAGCACGGGAAAGACCCAATCAATTGGAACATTCTGAACGAATAA

>5055_U81

ATGGCCCTACTACAGTGGATGTTAGATCACGTTTCAGGATGAAGAAAAAACTGTGAAAAC
CTTTC AATAGATGATCAGCATTACACTCTTCGGCATAAACAGAGACTGGTTGAGTTTCTTG
CAACTCTCCAAGTTAGAAATTACACACCTTAAACATGTATATAAAATTGGTGGACAACGAT
AGGGCTCATCTAACGGTCCACCCCTCTTCGGGTAACGTCCACGCTTGGAGTTTTTTATGC
AAACCCGACGATGTAAAAGTTGTGATTTTGGGACAAGATCCGTATCCCGACGGCAGAGGT
TGCGGCTTGGCCCTTCGGTACAGTGAAGAGTGCTCGATTCCAGAATCTCTAAAGAATATA
TTTAAAGAACTGGAAAGAAGCGTGCCGAATTTTTCCCCCCTGACAACGGCTGTTTAAAC
TCCTGGTGTAGTGAAGGAGTCCATTGCTAAACTCGATATTCAGTGTAGTTCATGGATTA
CCAATGTCCCACGAGGCATTTGGCTGGCAAACACTGAGCTATAAGATTATCAGCAAACCTA
TCGGAAGAAATGGACTCTCTCGTTTTCTTATTGTGGGAAATCATGCCCGGAAACTCTCT
TATCTAATAGACGCACGGAAACATCTCGTCTTAGAAAGTGCACATCCATCACCCAAAGTG
AAATCTGCAAGGATGCCATTTATTGGTTGCAATCATTTCGTGCGAGCAAATTTATTTCTT
ACGGAGCACGGGAAAGACCCAATCAATTGGAACATTCTGAACGAATAA

>5814_U81

ATGGCCCTACTACAGTGGATGTTAGATCACGTTTCAGGATGAAGAAAAAACTGTGAAAAC
CTTTC AATAGATGATCAGCATTACACTCTTCGGCATAAACAGAGACTGGTTGAGTTTCTTG
CAACTCTCCAAGTTAGAAATTACACACCTTAAACATGTATATAAAATTGGTGGACAACGAT
AGGGCTCATCTAACGGTCCACCCCTCTTCGGGTAACGTCCACGCTTGGAGTTTTTTATGC
AAACCCGACGATGTAAAAGTTGTGATTTTGGGACAAGATCCGTATCCCGACGGCAGAGGT
TGCGGCTTGGCCCTTCGGTACAGTGAAGAGTGCTCGATTCCAGAATCTCTAAAGAATATA
TTTAAAGAACTGGAAAGAAGCGTGCCGAATTTTTCCCCCCTGACAACGGCTGTTTAAAC
TCCTGGTGTAGTGAAGGAGTCCATTGCTAAACTCGATATTCAGTGTAGTTCATGGATTA
CCAATGTCCCACGAGGCATTTGGCTGGCAAACACTGAGCTATAAGATTATCAGCAAACCTA
TCGGAAGAAATGGACTCTCTCGTTTTCTTATTGTGGGAAATCATGCCCGGAAACTCTCT
TATCTAATAGACGCACGGAAACATCTCGTCTTAGAAAGTGCACATCCATCACCCAAAGTG
AAATCTGCAAGGATGCCATTTATTGGTTGCAATCATTTCGTGCGAGCAAATTTATTTCTT
ACGGAGCACGGGAAAGACCCAATCAATTGGAACATTCTGAACGAATAA

U82

>2284_U82

ATGGAACTTTTAGTATTTGTAATGTCACCTCATATTACTAACCTTCTCAAAGCAATACCT
CTTTTCAGCCATAATCTTTCTATTTTGA AAAACTTGACGACTGTATCGCAGCTGTAATA
AACTGCACGAAATCCGAAGTACCTCTATTACTAGAACCAATCTACCAGCCTCCGGTATAT
AATGAAGACGTTATGTCAATACTGCTACAACCTCCGGCAAAAAAAAAGCCTTTTAGCCGT

ATAATGGTAACCGATGAATTTCTCAGCGACTTCTTACTCCTACAGGATAACCCAGAACAA
CTACGCACATTGTTTTCGCGTAATAAGGGACCCAGAATCTCGGGACAACCTGGCTAAACTTT
TTCAATGGCTTCCAGACATGTTACCTTCCGTTCGGAATAACAACCTGCATCAGAGATAAC
TG TAGAAAATATTTCGCCCGAAAAAATTATGTACGTCAATAACTTTTTTGTGATAACATT
GCGGGTCTCGAGTTTAAACATTTTCAGAAAACACAGACAGTTTTTACAGCAACATTGGTTTT
TTATTATACCTGGAGAATCCTGCTAAAGGCGTCAAAAAATTATTAGGTTCCCTTTTAAAC
TCTTTGACTATTTTTGATACGATTTTGAATTGTTTAAAGTATTTCCACTTGAAAACCGGA
GTAGAGCTCGACCTGCTAAAACATATGGAACCTACAATTCTAAACTACCTTTCCGAAGT
TCCCGCCCTACAATTCTGATTAGAAACACATAA

>5055_U82

ATGGAACTTTTACTATTTGTAATGTCACTCATATTACTAACCTTCTCAAAGCAATACCT
CTTTTCAGCCATAATTCCTTCTATTTTGAAAACTTGACGACTGTATCGCAGCTGTAATA
AACTGCACGAAATCCGAAGTACCTCTATTACTAGAACCAATCTACCAGCCTCCGGTATAT
AATGAAGACGTTATGTCAATACTGCTACAACCCCGACAAAAAAAAGCCTTTTAGCCGT
ATAATGGTAACCGATGAATTTCTCAGCGACTTCTTACTCCTACAGGATAACCCAGAACAA
CTACGCACATTGTTTTCGCGTAATAAGGGACCCAGAATCTCGGGACAACCTGGCTAAACTTT
TTCAATGGCTTCCAGACATGTTACCTTCCGTTCGGAATAACAACCTGCATTAGAGATAAC
TG TAGAAAATATTTCGCCCGAAAAAATTACGTACGTCAATAACTTTTTTGTGATAACATT
GCGGGTCTCGAGTTTAAACATTTTCAGAAAACACAGACAGTTTTTACAGCAACATTGGTTTT
TTATTATACCTGGAGAATCCTGCTAAAGGCGTCAAAAAATTATTAGGTTCCCTTTTAAAC
TCTTTGACTCTTTTTGATACGATTTTGAATTGTTTAAAGTATTTCCACTTGAAAACCGGA
GTAGAGCTCGACCTGCTAAAACATATGGAACCTACAATTCTAAACTACCTTTCCGAAGT
TCCCGCCCTACAATTCTGATTAGAAACACATAA

>5814_U82

ATGGAACTTTTAGTATTTGTAATGTCACTCATATTACTAACCTTCTCAAAGCAATACCT
CTTTTCAGCCATAATTCCTTCTATTTTGAAAACTTGACGACTGTATCGCAGCTGTAATA
AACTGCACGAAATCCGAAGTACCTCTATTACTAGAACCAATCTACCAGCCTCCGGTATAT
AATGAAGACGTTATGTCAATACTGCTACAACCTCCGGCAAAAAAAAAGCCTTTTAGCCGT
ATAATGGTAACCGATGAATTTCTCAGCGACTTCTTACTCCTACAGGATAACCCAGAACAA
CTACGCACATTGTTTTCGCGTAATAAGGGACCCAGAATCTCGGGACAACCTGGCTAAACTTT
TTCAATGGCTTCCAGACATGTTACCTTCCGTTCGGAATAACAACCTGCATCAGAGATAAC
TG TAGAAAATATTTCGCCCGAAAAAATTATGTACGTCAATAACTTTTTTGTGATAACATT
GCGGGTCTCGAGTTTAAACATTTTCAGAAAACACAGACAGTTTTTACAGCAACATTGGTTTT
TTATTATACCTGGAGAATCCTGCTAAAGGCGTCAAAAAATTATTAGGTTCCCTTTTAAAC
TCTTTGACTATTTTTGATACGATTTTGAATTGTTTAAAGTATTTCCACTTGAAAACCGGA
GTAGAGCTCGACCTGCTAAAACATATGGAACCTACAATTCTAAACTACCTTTCCGAAGT
TCCCGCCCTACAATTCTGATTAGAAACACATAA

ACGGCGAGCGCGACGGCCGCGACGTGACGCGCGGACCGACCCGTCCGACACGGGTGACG
CGTAAGGGGCGTGTGGCAACCGAAGGCGTCCATCTCCCCGCACGCCCCGAATCTGAGGAA
CAGATAGACGGCCACCACGGGCGCCAGGAGAGCAGCGACGACCACCAGCGCGGGGGAGC
GGACGAGGACACCGCGATGACGGCGCGCACCGCCACGCGAATGACAAAACAGAGCCCCAG
CAGCGCGGAGAGCACGAGGAAAGGAAACAGACCGACTCCGGGGCGCCACGAGCACGCACAA
GAGAGCCAGGTGCGAAGAAGAGACGAGGAGGGAACGGAGCAGGGCGGTAGCGGAAGGAGC
TGCGGGGGGGCGACGCAGACGTACGGCGGTAGAGGCCGACATGATTCCCTGCCCGTGCATC
CCCCGTGTCGGTCCCCGGGCCCGATCCCCGCCTGTGGGTCCCTCCCCCTCATCTGTTATTT
CCTTCCCCTCTGCCACCGATGACGCCCCGTGACGACGAGCCGTCCGTCCGCCCTCGGTGC
CAGCCAAGTCCCGCAGAGGAACCCCCACGTGTGTCGTCCCCGTCCGCCCGCCCGTCCGTCC
GACACCCCGCTGTCCGCTGTCTCGCGCCCCCTCCGCTCCGCCCGTCCCCCGCCGTCCACT
GCGCGTGTTCGTTTCTTCCCTCTCTTCCCTCCTCCTCCCCGTCCACTCTCCCTCCCCG
TTGTCTCCGCCCTCTCCCGTCTCTCCCTCGTCTCCAGATCTCCGTTTCATCCCCCCCATC
AGATCTCCGGAACCTCTAGCAAAACCGCGGGTGTCTCCGACCATCCCGCGGGCGTTCCCG
CCGGCGCTCTCGTCCGCGCCGGCCCGTTTCGAAACAGTCCCGTCCGTTCCCTCATCGGCA
TCTCCGTCCGCGCCGTGCGTCGGAAAGTCGCAGCCGCCCGCCGCACACACGGCGTGA

>5814_DR1

TACAACCGCATCTTCTTCAGGGCTGGCCTACGAAGTCTTCTCTATGGTTTCGACCCTAAAC
CCTACCATCCTTCGGCCGACAGCAAGTTGCTACCGCTGGGTCTGATCACGCTGTCCGCCT
GTTCCACGCGCGTGTGCGAGCCGACACACTGGAGCGGCTACGACGCGACAGATCTGTCTC
TCAGCTGGCTGACGGGATCGTCCCCTTGGCTCGTGATGCTACAAGGACAGGGGGGTCTC
TGTTCTGCCACGACGTGCTGCAAGGTCGACTCTATGTCCTGTGCGACTCCGTGTGCTCT
TTCTAAAGACGGCCCTCCGCCACTGCGAGGCAATCTATCGCGCACCGCTGTGGCGCAACC
GACCCCTGCCGAGCCTGTGGACGTGTGCGGACCCCGACAAGGCCTTCTTGCCGACATTAC
TGGCCAGAAGCGCCGACGCGGCTGGCCGCTTTCTACGCCCTGTGGAGACTGCATCTGG
GATCGCGCTCAGAGCTCTCCACCCCGTGTGGAGTGGGAGAGAACAGAGCTTGTCTGA
CGACTGGAGACGCGGGCGGCCGTCTACGCACCTCCCGTCCGACTCGCGGTCCAGCGCG
CTACGTGAGTGGCGCCACGGACACGCGGGACGCGGCGACAGAGGAGGCGCGGGGGGAA
AAAAGGAGGCGGAGGAGACGGGCCAGCGTGCAGGAACTTTCGCGAGGTGCAGACCGC
TGCTGGACGAACGTGCGGAGAGGGCGGCTGGCTTCCCTTCGCGTTCCTCTCGACATCTC
CGCACGTCTGGCTGATCCTAACGGAAGGAGGCCCGTCTAGCGATTGACGTGAACGACA
CCTCCGTGTGGCGCATCGCGGACGACCTGGAGCTGCTGCGTGCCTGGGGAGCCTGCTCC
TCCACTCAGGCCCTCCGGCTTCTCTCCGTCCCCCGAGCGGGAGCGGCGAGGCGGCCGGAG
AGCCGGGTACGAGGAGGAAGAGAGAAGAGGAGAGCGTGCAGGCGAGCGCGACGCGCG
CGACGTGACGCGCGGACCGACCCGTCCGACACGGGTGACGCGTAAGGGGCGTGTGGCAA
CCGGAGGCGTCCATCTCCCCGCACGCCCCGAATCTGAGGAACAGACAGACGGCCACCACG
GGCGCCAGGAGAGCAGCGACGACCACCAGCGCGGCGGTAGCGGACGAGGACACCGCGATG
ACGGCGCGCACCGCCACGCGAATGACAAAACAGAGCCCCAGCAGCGCGGAGAGCACGAGG
AAAGGAAACAGACCGACTCCGGGCGCCACGAGCACGCACAGGAGAGCCAGGTGCGAAGAA
GAGACGAGGAGGGAACGGAGCAGGGCGGTAGCGGAAGGAGCTGCGGGGGGGCGACGCAGA
CGTACGGCGGTAGAGGCCGACATGATTCCCTGCCCGTGCATCCCCCTGTCCGTCCCCGGC
CCGATCCCCGCTGTGGGTCCCTCCCCCTCATCTGTTATTTCTTCCCTCTGCCACCGA
TGACGCCCGTGCAGCAGAGCCGTCCGTCCGCCCTCGGTGCCAGCCAAGTCCCGCAGAGG
AACCCCCACGTGTTGTCCCCGTCCGCCCGCCCGTCTGTCGACACCCCGCTGTCCGCTG
TCTCGCGCCCCCTCCGCTCCGCCCGTCCCCCGCCGTCCACTGCGCGTGTCCGTTTCTTCC
TCTCTTCTCTCTCTCCCCGTCTACTCTCCCTCCCCGTTGTCTCCGCCCTCTCCCGTCT
CTCCCTCATCTCCAGATCTCCGTTTCATCCCCCCCATCAGATCTCCGGGACTCCGAGCGA
AACCGCAGGTGTCTCCGGGCATCCCGCGGGCTTCCCGCGGGCGCCCTCGTCCGCGCCGG
CCCGTTCCAAAACAGTCCCGTCCGTTCCCTCATCGGCATCTCCGTCCGCGCCGT

DR6

>2284_DR6

CCCTAGACTCCTGCCTGTACGTATGCTGCGGATACGGAGAAAACTTCAACCCGTGGGTT
TCGTAAAGTCGTATGTAACCAACTCCAGCTCGACACGCTCCGCGTGCTTTTAGTGGGCA
AAGACGGAGCCGTGTACGTCCACCACATGAGGGCGGCAAGACTCTGCCGACTGGCGTCGA

ACACAACGGAGTTTACAAGACGAGGGCTGCAACGGGACGCCGTGACGTATGAAGAAGACC
TAGAGCTGCCAGACCGGCGTTTGTGCGGAACGAACGCCCGACATCTGTTTCGACGTGATCG
CCGCGGCCGCCGACGAACACGACCTGCTGACCGTCCGGCGGCCTGTGTGACAGCGCACGCCG
GAGTGTCTTGCAGCTTACTAGAGACCGTGGGAGATCCGTGGACGGCGGTTCCGGCCGCGC
GCATGACTCTGACCGTGCCGCAGGTTTCAGTACCGGTTATGGCCCCGAAGCCCGGAGAGACC
TCCGCCGGCACCTATACGCGGGACACCCCTGGGACCGTGGCTCGTGTGCGGGCGTTCTCT
CTCGAGAGAGGGAGACGCAGAGGCCGTGCGCTCCGATAGACAGCGGGCGGCGGATTGTGG
ATAACGTGCCACGCCGGGGCCACGCGAGGTGGAGATCGCTTGGGTGGTCTTGACTTTGG
CGGGACCTCTGTTATCGTTCTGGCCCCGATACCGGCAAGATCTGTCTGTGCGAAACTCGT
TCTCCACCTGTGGAGGATGGGACCGCGGGCCATGAGAGGACACTGGACGTATTCCGGCCC
CGGGTAGACATCTCCCCGGGGACGCGTGGCCGCTGTGCGAACACGTGAGACCGCCGGTGG
GAAAGCTTCCCCGGAAGAGAGCGTACCTGGATTAG

>5055_DR6

ATGTCGACGCGACACACACAGAGGACGGGCGGACGCATCTCGGTCCGGCGCGACGAGGCG
AGACTGGCCCACGCGGAGCCGGTGCCTCGAGTGGCTGCTCCTGGCCCCGCGGACAGG
CCGTCCAAACTGTACGGCTATGCGAGCCGGCATCGCGGAGAACTTATCCACCTACCGTGG
CCGCCGTGCTGGTGTCTAGAACTCCACCACGATCCGTACAGAGACGCCAGAAGTGCCACC
GTGTGGGGCCACCGTGGGGTTGGCCGCCGACGCACGTGAGACCCAAATGCGTTCAAGAC
TGCGCCCTAGACTCATGTCGTGTACGTATGCTGCGGATACGGAGAGAACTTCAACCCGTG
GGTTTCGTAAAGTTCGTATGTAACCAACTCCCAGCTCGACACGCTCCGCGTGTCTTGGTG
GGCAAAGACGGAGCCGTGTACGTCCACCACATGAGGGCGGCAAGACTCTGCCGACTGGCG
TCGAACACAACGGAGTTTACAAGACGAGGGCTGCAACGGGACGCCGTGACGTATGAAGAA
GACCTAGAGCTGCCAGACCGGCGTTTGTGCGGAACGAACGCCCGACATCTGTTTCGACGTG
ATCGCCGCGGCCGCCGACGAACACGACCTGCTGACCGTCCGGCGGCCTGTGTGACAGCGAC
GCCGGAGTGTATGCAACTTACTAGAGACCGTGGGAGATCCGTGGACGGCGGTTCCGGCC
CGCGCATGACTCTGACCGTGCCGCAGGTTTCAGTACCGGTTATGGCCCCGAAGCCCGGAGAG
ACCTCCGCCGGCACCTATACGCGGGACACCCCTGGGACCGTGGCTAGTGTGCGGGCGTTC
TCTCTCGAGAGAGGGAGACGCAGAGGCCGTGCGCTCCGATAGACAGCGGGCGGCGGATTG
TGGATAACGTGCCACGCCGGGGCCACGCGAGGTGGAGATCGCTTGGGTGGTCTTGACTT
TGGCGGGACCTCTGTTATCGTTCTGGCCCCGATACCGGCAAGATCTGTCTGTGCGAAACT
CGTTCCTCCACCTGTGGAGGATGGGACCGCGGGCCATGAGAGGACACTGGACGTATTCCG
CCCCGGGTAGACATCTACCCGGGGACGCGTGGCCGCTGTGCGAACACGTGAGACCGCCGG
TGGGAAAGCTTCCCCGGAAGAGAGCGTACCTGGATTAG

>5814_DR6

ATGTCGACGCGACACACACAGAGGACGGGACGGACGCATCTCGGTCCGGCGCGACAGGGCG
CGACTGGCCCAGGCGGAGCCCGTGCCTGCTTCGAGTGGCTGCTCCTGGCCCCGCGGACAGG
CCGTCCAAACTGTACGGCTATGCGAATCGGCATCGCGGAGAACTGATCCACCTACCGTGG
CCGCCGTGCTGGTGTCTAGAACTCCACCACGATCCGTACAGAGACGCCAGAAGTGCCACC
GTGTGGGGCCACCGTGGGGTTGGCCGCCGACGCACGTGAGACCCAGATGCGTTCAAGAC
TGCGCCCTAGACTCATGTCGTGTACGTATGCTGCGGATACGGAGAGAACTTCAACCCGTG
GGTTTCGTAAAGTTCGTATGTAACCAACTCCCAGCTCGACACGCTCCGCGTGTCTTGGTG
GGCAAAGACGGAGCCGTGTACGTCCACCACATGAGGGCGGCAAGACTCTGCCGACTGGCG
TCGAACACAACGGAGTTTACAAGACGAGGGCTGCAACGGGACGCCGTGACGTATGAAGAA
GACCTAGAGCTGCCAGACCGGCGTTTGTGCGGAACGAACGCCCGACATCTGTTTCGACGTG
ATCGCCGCGGCCGCCGACGAACACGACCTGCTGACCGTCCGGCGGCCTGTGTGACAGCGAC
GCCGGAGTGTATGCAACTTACTAGAGACCGTGGGAGATCCGTGGACGGCGGTTCCGGCC
GCGCGCATGACTCTGACCGTGCCGCAGGTTTCAGTACCGGTTATGGCCCCGAAGCCCGGAGA
GACCTCCGCCGGCACCTATACGCGGGACACCCCTGGGACCGTGGCTAGTGTGCGGGCGTT
CTCTCTCGAGAGAGGGAGACGCAGAGGCCGTGCGCTCCGATAGACAGCGGGCGGCGGATT
GTGGATAACGTGCCACGCCGGGGCCACGCGAGGTGGAGATCGCTTGGGTGGTCTTGACTT
TTGGCGGGACCTCTGTTATCGTTCTGGCCCCGATACCGGCAAGATCTGTCTGTGCGAAAC
TCGTTCTCCACCTGTGGAGGATGGGACCGCGGGCCATGAGAGGACACTGGACGTATTCCG
GCCCGGGTAGACATCTACCCGGGGACGCGTGGCCGCTGTGCGAACACGTGAGACCGCCGG
GTGGGAAAGCTTCCCCGGAAGAGAGCGTACCTGGATTAG

U11

>2284_U11

ATGGATCTGCAAAGACATCCGATTCGGTTTTGCGTGGCTAGATCGAGACAAAGTTGAGTAT
CTTACAAATTTTTCTCAGCAATTTGGAAAGACTGGATAATGTAGATTTGCGAGAGCATCCC
CATGTGACTAATTTCTTGTGTCGTGAGAGAGGGAGAAGATGTAGACGATTTAAAAACATTG
TATAACCTACTAGTGTTATGGCTTATGTATCACTACGTCTTATCTAAAAGGAAGCCGGAT
TATGATGCTATATGGCAAGACATCACGAACTCCAAAGTGTCTGTAACGAGTACTTAAAA
TCCAAAGGTCGAATAAAGGAATTTTTGAAAATATGTTACGAACAAAGAAAAGTTTGAA
TCGCAATTCAGTGATTTTAATCGCGCTTTACTGTGTTTAGGAACTCTATTAAGTGGGGT
AGCAATGTTGCGATCGATACTCCTTATGTAAATCTTACTGCAGAAGACAGCAGCGAGATA
GAAAATAATTTGCAAGATGCTGAAAAAACATGCTGTGGTATACCGTCTATAACATAAAT
GACCCCTGGGACGAAAACGGTTACTTAATAACGAGTATTAATAAATTAATTTATCTCGGT
AAGTTATTTTTTAGCGTTAACTCAGTCCTGGTCAAAGCTAGAAAAGGTTGCTATGAGTCAA
ATTGTAATCACGCAAAATCATCTCTCGGGTCATTTGAGGAGGCACGACAATTTTAATATT
GTATATTCATAGGGTTTTGTCAGACTCCTCTGACTGGTCAAAGAGTAGAGAGTTTTCTG
AAAATAATCACCAGTGATTATGATATTATCAAAAGTAGTCTGGAATCACACAGCGCATCA
AAAGCATTTTCGATGCTGAGATTGGGCCTTATTCTTTAATGGATTTTCGTCCCTTTACAC
GGCGATATACCTTCAAATTTGACTTTACCTAGTATGTCTATAGATACAAAGAAATCATCT
GTAGATCCGGCTCGTCTGAAAAAAGTAAATCCAGAAGTTTGATACTTTCTTAAGAATG
CAGAGACAACCGAAATTTCTAGAGTTGGATAGCGTTGATAATGCCGGGGGAAAAATTTTA
ATAAAGGAAGCAACACTCGGGGGTGAAGAAGTTAAAGCAACAACGCCTGCTTCGTCTGTC
TCTTTAATGTCCGGAGTTGAGTCGCCGTCGTCTTTCACTTCTACCAACCTGGATCTGCCG
TTGTCGTCTTTCACTTCTACTAATCTGGATCTGCGAGATAAGTCGCACGGTATTTATAAA
ATTGGCCCTTCGGGGATTTTAGATTTTAATGTTAAATTTCCACCTAATGCGCAATTGAAT
ACGAACGGCGTGGATTTACTACAGGATAAACTTCGATCGGGAGTCCCAGTAGCGGTATT
ACCGATGTGGTAAATGGTTTAGCTAATCTCAATCTGCATCAGAATAAATCGGAAGTTTTG
CCACCGTGGAGCAGAAACACAGCGGCAATGCGGATTTTTTAGATCCGGTGCATCGCTTT
GTTCCTGAGCAGACAGGGACCCCTTCGTGTTGAATAATCCGACGTGGCGGGATCAGAA
GCGAAGCATACTACTACAGCACGGAGACCGCGTGTCAACCCGTAATGTTTTTCTCATT
AAAGATTTGAGAGGCAAAGACGGTTTTTAGGAAACAGAAGCAGTCAGATATTCGAAAAGC
TTAACTAAGGAAAGAAATGATAAAGCTATAATGCACTCACGCGAGGTGACCGGAGATTCT
GGCGGTGCGACTGAACTGTAGCTGCTCGGAATTCCTCGGCTTGAGAAAAATTAAGCAA
GCGAATGGTTTTTTTTCCCGGGTTAAATAAGAAAAATGATCGTGACGTATTAAGAGGGGGG
AAAGGAAATAGCAAGGACTTCTTGCAATTTGGCGGCAATGCAAAAAAAAAAAGAAATGTCG
GGAAAGTTTAATGTCGATAAAGAAATGACGCGAAACGGGCAAGAGCCATCACGTAGTTTA
ATGGGAGATGCTAGAAATGCCGGAGATGAACAATATATTCAAGCGGGTCTCGGGCAGCGA
GTTAACAATCTTCTAAGTCAATTTACAAATCTGATTAGTTTAGGCGAGAAGGGCATCGAA
GACATTTTGCAGAATCAGCGCGGGACCGATTTAAAGTTGGCTACAGAAAAAAGTTCGGGA
CGCGAATCGGAGGAAGCTAACGTACAAAAAATTTTGAATTTAGTAATCCTCAAGATATG
TTTAAAAATTTTAGGTTGCAAAACAATCTCGATTCCGTTTCACTCTCCGTTTAGGCTACCG
GATGTTGATTTGTCGCGAGTTAGATTTCCGCGTCATTTAAGGACGCGTTAGACTTGAAG
CTTCCGGGTAACGGAGAACGAGAAATAGATCTCGCTCTTGAAAAAGTGAAGGTAGGCGAG
ACGGAAACCTCAGATTTAAAAGTCCGTCAGGATGAAAGTTTTGTTCTGCGCAATTAATG
AAGGTTGAGACACCTGAAGAAAAAGATGATGTAATTGAACGGATGGTTCTGAGGATACGT
CAAGACGGGGAACTGATGAAAACACCGTCTCTGGACCGGGAGTCGCTGAGTCTTTGGAT
ATAGAAGCCAAAGGCGAGTCAGCGATCGCGTCTGTA

>5055_U11

ATGGATCTGCAAAGACATCCGATTCGGTTTTGCGTGGCTAGATCGAGACAAAGTTGAGCTT
CTTACAAATTTTTCTCAGCAATTTGGGAAGACTGGAGAATGTAGATTTGCGAGAGCATCCC
CATGTGACTAATTTCTTGTGTCGTGAGAGAGGGAGAAGATGTAGACGATTTAAAAACATTG
TATAACCTACTAGTGTTATGGCTTATGTATCACTACGTCTTATCTAAAATGAAGCCGGAT
TATAATACATATATGGCAAGACATAACGAACTCCAAAGTGTCTGTAACGAGTACTTAAAA
TCCAAAGGTCGAATAAAGGAATTTTTGAAAATATGTTACGAACAAAGAAAAGTTTGAA
TCGCAATTCAGTGATATTAATCGCGCTTTACTGCGTTTTAGGAACTCTATTAAGTGGGGT
AGCAATGTTGCGATCCATACTCCTTATGTAAATCTTACTGCAGAAGACAGCAGCGAGATA

GAAAATAATTTGCAAGATGCTGAAAAAACATGCTGTGGTATAACCGTCTATAACATAAAT
GACCCCTGGGACGAAAACGGTTTCTTAGTGACGAGTATTAATAAATTAATTTATCTCGGT
AAGTTATTTTTAGCGTTAAAACAGTCCGGTCAAAGCTAGAAAAGGTTGCTATGAGTCAA
ATTGTAATCACGCAAAATCATCTCTCGGGTCATTTGAGCAGGAACGACAATTTTAATATT
GTATATTCTCATAGGGTTTTGCAGACTCCTCTGACTGGTCAAAGAGTAGAGAGTTTTCTG
AAAATAATCACCAGTGATTATGATATTATCAAAAAGTAGTCTGGAATCACACAGCGCGTCCG
AAAGCATTTTCGGTGTCTGAGATTGGGCCTCATTCTTTAATGGATTTTCGTCTCTTTAGAC
GGCAAGATACCTTCAAATTTGACTTTACCTAGTATGTCTATAGATACAAAGAAATCATCT
GTAGATCCGGCTCGTCTGAAAAAAGTGAATCCAGAAGTTTTGATACTTTCTTAATGCAG
AGACAACCGAAATTTCTAGAGTTGGATAGCGTTGATAATGCCGGGGGAAAAATTTAATA
AAGGAAGCAACACTCGGGGTGAAGACGTTAAAGCGACAACGCCTGCTTCCTCTGTCTCT
TTAATGTCCGGAGTTGAGTCGCCGTCTTTTCACTTCTACCAATCTGGATCTGCCGTTG
TCGTCTTTCACTTCTACTAATCTGGATCTGCGAGATAAGTCGCACGGTAATTATAAAATT
GGCCCTTCGGGGATTTTAGATTTTAATGTTAAATTTCCACCTGATGCGCAATTGAATACG
AACGATGTGGATTTACTACAGGATAAAACTTCGATCGGGAGTCCTAGTAGCGGTATTACC
GATGTGGTAAATGGTTTTCGCTAATCTCAATCTTCATCAGAATAAATCGAATGTTTCGCCA
CCGTGGAGCAGAAACACAGCGCGAATGCGGACTTTTTAGATCCGGTGCATCGCTTTGTT
CCTGAGCAGACAGGGACACCCCTTCGTGTTGAATAATTCCGACGTGGCGGGATCAGAAGAG
AACCATACGACTTACAGTACGGAGACCGCGTTTTCACCCGTAACGTTTTTCTCATTAAA
GATTTGAGAGGCAAAGACGGTTTTTAGGAAACAGAAGCAGTCAGATATTCGAAAAGCTTA
ACTAAGGAAAGAAATGATAAAGCTATAATGCACCTCACGCGAGGTGACCGGAGATTCTGGC
GATGCGACTGAACTGTGGGTGCTCGGAATTTCCCGCGCTTGAGAAAAATTAAGCAAGCA
AATGATTTTTTTGCCGGGTAAATAAGAAAAATGATCGTGACGTATTAAGAGGGGGGAAA
GGAAATAGCAAGGACTTGCATTTTCGGCGCAATGCAAAAAAAGAAATGTCCGGGAAAG
TTAATGTGATAAAGAAATGACGCGAAACGGACAAGAGCCATCACGTAGTTAATGGGA
GATGCTAGAAATGCCGGAGATGGACAATATATTCAAGCGGGTCTCGGGCAGCGAGTTAAC
AATCTTCTAAGTCAATTTACAAATCTGATTAGTTTAGGCGAGAAGGGCATCGAAGACATT
TTGCAGAATCAGCGCGGGACCGAGTTAAAGTTGGCTACAGAAAACAAGTCGGGACGCGAA
TCGGAGGAAGCTAACGTAAAAAGAATTTCTTGAAGTTAGTAATCCTCAAGATATGTTAAA
AATTTTAGGTTGCAAAACAATCTCGATTCCGTTTCAGTCTCCGTTTAGGCTACCGGATGTT
GATTTGTCTCGCGAGTTAGATTCCCGCTCATTTAAGGACGCGTTAGACTTGAAGCTTCCG
GGTAACGGAGAACGAGAAATAGATCTCGCTCTTGAAAAAGTGAAGGTAGGCGAGACGGAA
ACCTCAGATTTAAAAGTCCGTCAGGATGAAAGTTTTGTTCTGCGCAATTAATGAAGGTT
GAGACACCTGAAGAAAAAGATGATGTAATTGAACGGATGGTTCTGAGGATACGTCAAGAC
GGGAAACTGATGAAAACACCGTCTCTGGGCCGGGAGTCGCTGAGTCTTTGGATATAGAA
GCCAAAGGCGAGTCAGCGATCGCGTCTGTA

>5814_U11

ATGGATCTGCAAAGACATCCGATTCCGTTTGCCTGGCTAGATCGAGACAAAGTTGAGCTT
CTTACAAATTTTCTCAGCAATTTGGGAAGACTGGAGAATGTAGATTTGCGAGAGCATCCC
CATGTGACTAATTTCTGTGTCTGTGAGAGAGGGAGAAGATGTAGACGATTTAAAAACATTG
TATAACCTACTAGTGTATGGCTTATGTATCACTACGTCTTATCTAAAATGAAGCCGAAT
TATGATGCTATATGGCAAGACATAACGAAACTCCAAAGTGTCTGTAACGAGTACTTAAAA
TCCAAAGGTCGTAATAAAGGAATTTTTGAAAATATGTTACGAACAAAGAAAAGTTTCAA
TCGCAATTCAGTGATATTAATCGCGCTTTACTGCGTTTAGGAACTCTATTAAGTGGGGT
AGCAATGTTGCGATCCATACTCCTTATGTAAATCTTACTGCAGAAGACAGCAGCGAGATA
GAAAATAATTTGCAAGATGCTGAAAAAACATGCTGTGGTATAACCGTCTATAACATAAAT
GACCCCTGGGACGAAAACGGTCACTTAGTGACGAGTATTAATAAATTAATTTATCTCGGT
AAGTTATTTTTAGCGTTAAATGAGTCCGGTCAAAGCTAGAAAAGGTTGCTATGAGTCAA
ATTGTAATCACGCAAAATCATCTCTCGAGTCATTTGAGCAGGAACGACAATTTTAATATT
GTATATTCTCATAGGGTTTTGCAGACTCCTCTGACTGGTCAAAGAGTAGAGAGTTTTCTG
AAAATAATCACCAGTGATTATGATATTATCAAAAAGTAGTCTGGAATCACACAGCGCGTCCG
AAAGCATTTTCGGTGTCTGAGATTGGGCCTCATTCTTTAATGGATTTTCGTCTCTTTAGAC
GGCAAGATACCTTCAAATTTGACTTTACCTAGTATGTCTATAGATACAAAGAAATCATCT
GTAGATCCGGCTCGTCTGAAAAAAGTGAATCCAGAAGTTTTGATACTTTCTTAATGCAG
AGACAACCGAAATTTTTAGAGTTGGATAGCGTTGATAATGCCGGGGGAAAAATTTAATA
AAGGAAGCAACACTCGGGGTGAAGACGTTAAAGCGACAACGCCTGCTTCCTCTGTCTCT
TTAATGTCCGGAGTTGAGTCGCCGTCTTTTCACTTCTACCAATCTGGATCTGCCGTTG

TCGTCTTTCACCTTCTACTAATCTGGATCTGCGAGATAAGTCGCACGGCAATTATAAAAATT
GGCCCTTCGGGGATTTTAGATTTTAAATGTTAAATTTCCACCTGATGCGCAATTGAATACG
AACGATGTGGATTTACTACAGGATAAAACCTTCGATCGGGAGTCCTAGTAGCGGTATTACC
GATGTGGTAAATGGTTTCGCTAATCTCAATCTTCATCAGAATAAATCGAATGTTTCGCCA
CCGTGGAGCAGAAACACAGCGGCGAATGCGGACTTTTTAGATCCGGTGCATCGCTTTGTT
CCTGAGCAGACAGGGACACCCCTTCGTGTTGAATAATTCCGACGTGGCGGGATCAGAAGAG
AAGCATACGACTTACAGTACGGAGACCGGCGTTTTACCCCGTAACGTTTTTCTCATTAAA
GATTTGAGAGGCAAAGACGGTTTTTAGGAAACAGAAGCAGTCAGATATTCGGAAAAGCTTA
ACTAAGGAAAGAAATGATAAAGCTATAATGCACCTCACGCGAGGTGACCCGGAGATTCTGGC
GATGCGACTGAAACTGTGGGTGCTCGGAATTCGCCGGCGTTGAGAAAAATTAAGCAAGCA
AATGATTTTTTTGCCGGGTAAATAAGAAAAATGATCGTGACGTATTAAGAGGGGGGAAA
GGAAATAGGAAGGACTTGCAATCTGGCGGCAATGCAAAAAAAAAAAGAAATGTCGGGAAAAG
TTAATGTCGATAAAGAAATGACGCGAAACGGACAAGAGCCATCACGTAGTTAATGGGA
GATGCTAGAAATGCCGGAGATGGACAATATATTCAAGCGGGTCTCGGGCAGCGAGTTAAC
AATCTTCTAAGTCAATTAACAAATCTGATTAGTTTAGGCGAGAAGGGCATCGAAGACATT
TTGCAGAATCAGCGCGGGACCGAGTTAAAGTTGGCTACAGAAAACAAGTCGGGACGCGAA
TCGGAGGAAGCTAACGTACAAAAATTTTGAATTTAGTAATCCTCAAGATATGTTAAA
AATTTTAGGTTGCAAAACAATCTCGATTCGGTTCAGTCTCCGTTTAGGCTACCGGATGTT
GATTTGTCTCGCGAGTTAGATTCGCGTCAATTAAGGACGCGTTAGACTTGAAGCTTCCG
GGTAACGGAGAACGAGAAATAGATCTCGCTCTTGAAAAAGTGAAGGTAGGCGAGACGGAA
ACCTCAGATTTAAAAGTCCGTCAGGATGAAAGTTTTGTTCTGCGCAATTAATGAAGGTT
GAGACACCTGAAGAAAAAGATGATGTAATTGAACGGATGGTTCTGAGGATACGTCAAGAC
GGGAAAAC TGATGAAAACACCGTCTCTGACCGGGAGTCGCTGAGTCTTTGGATATAGAA
GCCAAAGGCGAGTCAGCGATCGCGTCGTGA

U13

>2284_U13

ATGGCGCACGCTAAAAACGGGTACGACGAAAACTTTTAACTTCGACAGACGACCCCATATA
TTGTCCAACACTTTTACCATGCGCCCCGACAAGTAAGATTGCAGACGCTGAAATCATTTC
AGAGAACATGATTATATCGCAAGCAAAACACAGACGGATTCAAAAAAATATCGCCCCCTG
TCTGTGATTTTAGACAAAAC TGCCCTCTTTGAATTTTACGGTATAAGAGACAACAACGAA
AAAGCTATCGTCTACCCATCGATCCAGATTTCTTATTATGTGACTCGGAACATAATTGC
ACCCTATCACCTTTCTTATAA

>5055_U13

ATGGCGCACGCTAAAAACGGGTACGACGAAAACTTTTAACTTCGACAGACGACCCCATATA
TTGTCCAACACTTTTACCATGCGCCCCGACAAGTAAGATTGCAGACGTTGAAATCATTTC
AGAGATCATGATTATATCGCAAGCAAAACACAGACGGATTCAAAAAAATATCGCCCCCTG
TCTGTGATTTTAGACAAAAC TGCCCTCTTTGAATTTTACGGTATAAGAGACAATGAAAAA
GCTATCGTCTACCCATGATCCAGATTTCTTATTATATGACTCGGAAAATAATTGCACC
CTATCACCTTTCTTATAA

>5814_U13

ATGGCGCACGCTAAAAACGGGTACGACGAAAACTTTTAACTTCGACAGACGACCCCATATA
TTGTCCAACACTTTTACCATGCGCCCCGACAAGTAAGATTGCAGACGTTGAAATCATTTC
AGAGATCATGATTATATCGCAAGCAAAACACAGACGGATTCAAAAAAATATCGCCCCCTG
TCTGTGATTTTAGACAAAAC TGCCCTCTTTGAATTTTACGGTATAAGAGACAATGAAAAA
GCTATCGTCTACCCATGATCCAGATTTCTTATTATATGACTCGGAAAATAATTGCACC
CTATCACCTTTCTTATAA

U14

>2284_U14

ATGGTCACACACGCGATGGAAGAACC AAAGACATTCAACATTCCAACCTTTCGTCTTAGAC
GAAAATTGTAACCTCATACCCGAAGTTCTCTCACGCGCCAACGCTAAATATATCAAAGAT
GTACTCATCCAAGACTCCTACAATACGGTTTTGTCTCGCAAACAGCTTTATTCCCATGACG

ATTCAAACCACAGAGCAAATTATGGTTATCGTCCAAAAGTTTAAAGTTTTACGTGCACGA
AATCTGTTTCGAAAAAGTACTCCGCCTCGGCGTACATATAAACAGATTTTATGCCGGAAAA
AAACAAGTCAAACATATGATTACAATGATGAAAAGTCTTTTCGACACCGAAGAAGCCATG
CGACGTCTCGACAGAGCTCTTATGGGGCTGTTTGTTCGATGCTCGTGGTAGTTTCATATATG
CCTCTCATAGCCTTGTCCTTGCACGAAAACGGCCTTCCAGATTCCAAATTCGCAAAAAGCC
GTCAAATTAATAAACACAACCTTAAACTCCTTTTCATAATCGACCGGACGCAGATATCGAA
CAGTATGCACAAAAATTACGAGCATACAATTACCTCTATAAAAATACCCATTATACTTTG
AAAGAAGCTGTTGACGTTTATTCCGAAAATCTTAAAGATCTCTCTATTGGAGTTAATAAAA
AAGCCACATTACTGTTTACTTCTTTCAGACGACGGGTATTTATCTCATATTTTTAACGAC
ATGCTATTTCTTAACTTCTACTTGGAAATATGATTTATAACTCCAAAGTACAAATCAGGCGT
TTGACCCTTGGATCAAATACGAAATCAACTCCATATTAGAAGATTGCGTATTGGTGGGA
TTCCAACCTCCAGATCTAAAGGACACAATACTAGATCTAGCGGCACTCACATCAAATATG
AATCTCGTCGACCCAGATAAAGACCTTTTTTCTCACTACAAGCTCATACTGGAAAAACTT
TTTGAATCTCTATTTTCGCGACTAAGGCAAACATTTGCATTTTACCAACCTTCATAAAG
AGTCACCTAATAGAGTTCGAGGATGTCTTAAAAAGAAGCAATGATGATGAAGACCTCAAT
TATCTGCTTCTGAAATCTAACGATTCTGATGATGAATACGACGAAGACAAACTACCAATA
CAAGTCGATCCAGGCAGAGTAGACAACGTGCTTACAGATTCTAACTTTTTTAACGTAACC
GCAGATAATGCTTTCTCGTCTATAGCGATCATGCCAATTTCTTATGATAAGATCATAGAT
GTGGAAGAAAATGTAATTCAGTCCTAGAAGTTGAGATGCAGAGCCTCTCAGCAGTTGTC
TACGGAGCAGTTGCAAACAAATACGGCTTAAGCCTACCACAAGTCATCAAAGGTTAAAT
CAAAATGAAGGACGGGCTCTTCTCGTGCATCCTCCAGCCACAGTACTTCTACCATCCCA
TACTCCCCCGCAAAGCGGTGCTCCACTCCAACCTTCTATCCTAAGGCAAAGAGCACCA
ATACGTTCCAACAGCAGATCATCATCGGTCTCTTTCTCACAAGACGACGATAATCGTAGT
CATTACTCTGACGAAACAAATATCAGTGATTATTTCATATCCCATGGCAGATCTAGATTTA
GAAGACGAAGAACCATGGAGGACCATCCACACTCCCCTCAATCAGCATCATCTAACAAC
TCCATGTCTCGTCGAAGCCGAGCTCTACAAAACGAACAAAGAAGACGTACGCCACAATG
GCCCCACCATCTCCGGCAAGAGGACAAAATAATGCCCGCTCAGACGCAGAACCAGACGC
TCAACGAAACGAACGACGAAAGGTTATAA

>5055_U14

ATGGTCCACACGCGGATGGAAGAACCAAAGACATTCAACATTCCAACTTTTCGTCTTAGAC
GAAAATTGTAACCTTCATACCCGAAGTTCTCTCACGCGCAACGCTAAATATATCAAAGAT
GTACTCATCCGAGACTCCTACAATACGGTTTTGTCTTGCAAACAGCTTTATTCCCATGACG
ATTCAATCCGTAGAGCAAATTATGACTATCATCACAAGTTTAAAGTTTTACGTGCACGA
GATCTGCTCACAAAAATAGTCCGCCTCGGCGTACATATAAACAGATATTATGCCGGAAAA
AAACAAGTCAAACATATGATTACAATGATGAAAAGTCTGTTTCGATACCGAAGAAGCCATG
CGACATCTCGACAGAGCTCTTATGGGGCTGTTTGTTCGATACTCCTAATAGTTTCATATATG
CCTCTCATAGCCTTGTCCTTGCACGAAAACGGCTTTCCAGATTCCAAATTCGCAAAAAGCC
GTCAAATTAATATACACAACCTTAAACTCCTTTTCATAATCGACCGGACGCAGATATCGAG
CAATATACACAAAAATTAAAAGCATACAATTACCTCTATAAAAATACCCATTATACTTTG
AAAGAAGCTGTTGACATTTATTCCGAAAATCTTAAAGATCTCTCTATTGGAGTTAATAAAA
AAGCCACATTACTGTTTACTTCTTTCAGACGACGGATATTTATCTCATATTTTTCAACGAC
ATGCTATTTTTTAACTTCTACTTGGAAATATGATTTATAACTGCAAAATAGAAATCAGGCGT
TTGAACACTTGGATCAAACAAGAAATCAACTCCATGATGGAAGATTGCGTATTGACGGGA
TTCCAACCTCCAGATCTGAAGGACACAATCTAGATCTAGCGGCACTCACATCAAATATG
AATCTCGTCAGCCAGATAAAGACCTTTTTTCTCACTACAAGCTCATATTAGCAAAATTA
TTTGAACCTCTCTATTTACGCGACTAAGGCAAACATTTGCATTTTACCAACCTTCATAAAG
AGTCACCTAATAGAGTTCGAGGATGTCTTAAAAAGAAGCAATGATGACGAAGACCTCAAC
TATCTGCTTCTGAAATCTAACGATTCTGATGACGAATACGACGAAGACAAACTACCAATA
CAAGTCGATCCAGGCAGAGTAGACAATGTGCTTACAGATTCTAACTTTTTTAACGTAACC
GCAGATAATGCTTTCTCGTCTATAGCGATCATGCCAATTTCTTATGATAAGATCATAGAT
GTGGAAGGAAATGAAATTCAGTCCTAGAAGTTGAGATGCAGAACCTCTCAGCAGTTGTC
TACGGAGCAGTTGCAAACAAATACGGCTTAAGCCTACCACAAGTCATCAGAAGGTTAAAT
CAAAATGAAGGACGGGCTCTTCTCGTGCATCCTCCAGCCACAGTACTTCTACCATCCCA
TACTCCCCCGCAAAGCGATCGCTCCACTCCAACCTTCTATTCTAAGGCTAAGAGCACCA
ATACGTTCCAACAGCAGATCATCATCGGTCTCTTTCTCACAAGACGACGATAATCGTAGT
CATTACTCTGACGAAACAAATATCAGTGATTATTTCATATCCCATGGCAGATCTAGATTTA
GAAGACGAAGAACCATGGAGGACCATCCACACTCCCCTCAATCAGCATCATCTAACAAC

TCCATGTCTCGTCTGAAGCCGAGCTCTACAAAACGAACAAAGAAGACGTACGCCCCACAATG
GCCCCACCATCTCCGGCAAGAGGACAAAATAATGCCCGCCTCAGACGCAGAACCAGACGC
TCAACGAAAACGAACGACGAAAGGTTATAA

>5814_U14

ATGGTCACACGCGCGATGGAAGAACCAAAGACATTCAACATTCCAACCTTTCGTCTTAGAC
GAAAATTGTAACCTTCATACCCGAAGTTCTCTCACGCGCCAACGCTAAATATATCAAAGAT
GTACTCATCCGAGACTCCTACAATACGGTTTTGTCTTGCAAACAGCTTTATTCCCATGACG
ATTCAATCCGTAGAGCAAATTATGACTATCATCACAAGTTTTAAGTTTTACGTGCACGA
GATCTGCTCACAAAAATAGTCCGCCTCGGCGTACATATAAACAGATATTATGCCGGAAAA
AAACAAGTCAAACATATGATTACAATGATGAAAAGTCTGTTTCGATACCGAAGAAGCCATG
CGACATCTCGACAGAGCTCTTATGGGGCTGTTTTGTTCGATATTCTAATAGTTTCATATATG
CCTCTCATAGCCTTGTCCTTGACGAAAACGGCTTTCCAGATTCCAAATTCGCAAAAAGCC
GTCAAATTAATATACGCAACCTTAAACTCCTTTTCATAATCGACCCGGACGCAGATATCGAG
CAATATACACAAAAATTAAAAGCATACAATTACCTCTATAAAAATACCCATTATACTTTG
AAAGAAGCTGTTGACATTTATTCCGAAAATCTTAAAGATCTCTCTATTGGAGTTAATAAAA
AAGCCCACATTACTGTTTACTTCTTCAGACGACGGATATTTATCTCATATTTTCAACGAC
ATGCTATTTTTTAACTTCTACTTGGAAATATGATTTATAACTGCAAAATAGAAATCAGGCGT
TTGAACACTTGGATCAAACAAGAAATCAACTCCATGATGGAAGATTGCGTATTGGTAGGA
TTCCAACCTCCAGATCTGAAGGACACAATTCTAGATCTAGCGGCACTCACATCAAATATG
AATCTCGTCAGCCCAGATAAAGACCTTTTTTCTCACTACAAGCTCATATTAGCAAAATTA
TTTGAACCTCTCTATTTACGCGACTAAGGCAAACATTTGCATTTTACCAACCTTCATAAAG
AGTCACCTAATAGAGTTCGAGGATGTCTTAAAAAGAAGCAATGATGACGAAGACCTCAAC
TATCTGCTTCTGAAATCTAACGATTCTGATGACGAATACGACGAAGACAAACTACCAATA
CAAGTCGATCCAGGCAGAGTAGACAATGTGCTTACAGATTCTAACTTTTTTAAACGTAACC
GCAGATAATGCTTCTCGTCTATAGCGATCATGCCAATTTCTTATGATAAGATCATAGAT
GTGGAAGGAAATGAAATTCAGTCCTAGAAGTTGAGATGCAGAGCCTCTCAGCAGTTGTC
TACGGAGCAGTTGCAAACAAATACGGCTTAAAGCCTACCACAAGTCATCAGAAGGTTAAAT
CAAAATGAAGGACGGGCTCTTCTCGTGCATCCTCCAGCCACAGTACTTCTACCATCCCA
TACTCCCCCGCAAAGCGATCGCTCCACTCCAACCTTCTATTCTAAGGCTAAGAGCACCA
ATACGTTCCAACAGCAGATCATCATCGGTTTTCTTTCTCACAAGACGACGATAATCGTAGT
CATTACTCTGACGAAACAAATATCAGTGATTATTCATATCCCATGGCAGATCTAGATTTA
GAAGACGAAGAACCATGGAAGACCACCCGCACTCCCCTCAATCAACATCATCTAACAAC
TCCATGTCTCTTGAAGCCGAGCTCTACAAAACGAACAAAGAAGACGTACGCCCCACAATG
GCCCCACCATCTCCGGCAAGAGGACAAAATAATGCCCGCCCCAGACGCCTAACCCAGACGC
TCAACGAAAACGAACGACGAAAGGTTATAA

U15

>2284_U15

ATGGATGTGTGGAAGCGTCAACGGCTTCAAGAATGCCGTGAATTGTGTCTTTGCCTGTA
TTAATGTCACTGTGCAATATGTTGTCAAAAATCGAAATCGTATACGTTAAATATCTATTT
AAAATGGACTTTTTATACTATGTATAGATTTCTTTTACCGGCTCTTACGTTGAGTATGACG
GTTACAAAATCCGTAGTTATTGAAATGTTATTTATTTTAAAAAGATGGGAAGATATTGAT
CAATTTTTTTAGATTGAACATCCGGAAGTCAACGACTGTATGGTCGTAGCTCAGTTCCACC
CATATTCCTATAAAAAGGAAGTTAATTGTGTTGTTGTATATGTTAACAGTAGACAAGAA
AAACAGCTTTTCTCAATATGATATATGCGTTTTTTAGAGAAGAGTCACCTTAGACTTGGT
GATGATGAAGAAAAGAATGCTATTCGCTTTTTTTCTTACGTTGATGATTTACACCTAACG
CGGGATATTTTATTGGATATGATCTACAAGCTCAAGAATACGGAAATCAATCAAACAATG
GAACTTTTTATTATCGTACAATGAATTGGCTAGATAA

>5055_U15

ATGGATGTGTGGAAGCGTCAACGGCTTCAAGAATGCCGTGAATTGTGTCTTTGCCTGTA
TTAATGTCACTGTGCAATATGTTGTCAAAAATCGAAATCGTATACGTTAAATATCTATTT
AAAATGGACTTTTTATACTATGTATAGATTTCTTTTACCGGCTCTTACGTTGAGTATGACG
GTTACAAAATCCGTAGTTATTGAAATGTTATTTATTTTAAAAAGATGGGAAGATATTGAT
CAATTTTTTTAGATTGAACATCCGGAAGTCAACGACTGTATGGTCGTAGCTCAGTTCCACC

CATATTCCTATAAAAAGGAAGTTAATTGTGTTGTTGTATATGTTAACAGTAGACAAGAA
AACAGCTTTTCCCTCAATATGATATATGCGTTTTTAGAGAAGAGTCACCTTAGACTTGGT
GATGATGAAGAAAAGAATGCTATTCGCTTTTTTTTCTTACGTTGATGATTTACACCTAACG
CGGGATATTTTATTGGATATGATCTACAAGCTCAAGAATACGGAAATCAATCAAACAATG
GAACTTTTATTATCGTACAATGAATTGGCTAGATAA

>5814_U15

ATGGATGTGTGGAAGCGTCAACGGCTTCAAGAATGCCGTGAATTGTGTCCTTTGCCTGTA
TTAATGTCACGTGTCGAATATGTTTTCAAAAATCGAAATCGTATACGTTAAATATCTATTT
AAAATGGACTTTTCTACTATGTATAGATATATTTTACCGGCTCTTACGTTGAGTATGACG
GTTACAAAATCCCTAGTTATTGAAATGTTATTTATTTTAAAAAGATGGGAAGATATTGAT
CAATTTTTTTAGATTGAACATCCGGAAAGTAAGCGACTGTTTCATCGTAGCTCAGTTCAAC
CATATTCCTATAAAAAGGAAGTTAATTGTGTTGTTGTATATGTTAACAGTAGACAAGAA
AACAGCTTTTCCCTCAATATGATATATGCGTTTTTAGAGAAGAGTCACCTTAGACTTGGT
GATGATGAAGAACAGAATGCTATTCGTTTTTTTTTCTTACGTTGATGATTTACACCTAACG
CGGGATATTTTATTGGAGATCATCCACAAGCTCAAGAATACGGAAATCAATCAAACAATG
GAACTTTTATTATCGTACAATGAATTGGCTAGATAA

U19

>2284_U19

ATGGCATTACCCGGTGATGAACGGCGCGCATGTTGCAGTTTAAAGATAAGATGATTTCT
TCGGCGGGGTTTCGCTCTGAAATTTGAGAAAGTTGTACAGGAGGCTATGGCTTCCGGAATA
GTTCTTCAACATATCACTTGTATTAAGTTTCGGATTTGTGACAATTCCGATATTTTGTCT
GACCGTCAGTTGAGATGTTTTATTAATAAACGGGTTGTATCCTTTTGAGGGTAGGATAAGT
ATGTTTGGTGTAAACGGAAGAATGGGAAGGTGCGTCTGCGGCTCCGGAGAGACAGGTTGTT
TTTTTGTATCGTCAACTGGTCAAGTGCCTGGATATGAAGATGGTGTGGTATTTTATTTG
TCACCTACGTTTTTTCAGATTTTTTGGACAACCTGCCATGGAGTTCTCGTGTGAGAACGCGATC
TTGAGTAATTTTTATCGCACAGAAGTCCCGCGATGAATACAGTGATCAATTTCAAAAATAT
TTTACGAGAATGAGACATACCCCAATTTTTTTTGCAGGTGTATTGCCTAGACGGTTTCAG
AAGGTAGAGAGCGGTGATTGTGTGGAAGAAGATGCTCGTGCTTCCATGAGACCAATTCAG
TGTGACTCTTTTTGGTGTAAAGGATACATTTTGTGCGCCGACTGAAGAATTGCTGCAGCCA
TCTGCGAATAAAGATGTTGGGGGCAAGGTGTGTATGGCTTTATCGTGTCAAGAAGACAAC
TCCGCTCGACATTGCACAATCTATGGACTGACCAAACTCGTGGTATAAAAAATTATGTTT
TCGAGGCATACACAGACAGACCGGCTCGAGGTGATGTGTAACGCTGCCACCCAAACCGGA
GATGTGGTGGATAATCTTCAGAACTCTGTTTTCTCGGTGAAAATTTGGTCCATCAGTCA
ATTTTGGAAACGGAAGTCAAAACGACCGCAAAAATACTTTTTGATGTCTCTGATCCTCGA
ATCGACTCGGTGTACGATACGACAGTGTTCGGTGCTATGGCAACCGATGATGTAGGTAGT
GAAAATGTTTCAGGGCGGTGCGTCTTTGGCGCAAGAAAAACCTTTAAAAGGTTATTGTATC
ATTGCTACTCCATCAGAATGCAAACCGAACATTCATTGGCTTAAAGAGCCAGAAAATGAT
GTTTCAGGAATCGGCAGCTGTGCTGAGGTAG

>5055_U19

ATGGCATTACCCGGTGATGAACGGCGCGCATGTTGCAGTTTAAAGATAAGATGATTTCT
TCGGCGGGGTTTCGCTCTGAAATTTGAGAAAGTTGTACAGGAGGCTATGGCTTCCGGAATA
GTTCTTCAACATATCACTTGTATTAAGTTTCGGATTTGTGACAATTCCGATATTTTGTCT
GACCGTCAGTTGAGATGTTTTATTAATAAACGGGTTGTATCCTTTTGAGGGTAGGATAAGT
ATGTTTGGTGTAAACGGAAGAATGGGAAGGTGCGTCTGCGGCTCCGGAGAGACAGGTTGTT
TTTTTGTATCGTCAACTGGTCAAGTGCCTGGATATGAAGATGGTGTGGTATTTTATTTG
TCACCTACGTTTTTTCAGATTTTTTGGACAACCTGCCATGGAGTTCTCGTGTGAGAACGCGATC
TTGAGTAATTTTTATCGCACAGAAGTCCCGCGATGAATACAGTGATCAATTTCAAAAATAT
TTTACGAGAATGAGACATACCCCAATTTTTTTTGCAGGTGTATTGCCTAGACGGTTTCAG
AAGGTAGAGAGCGGTGATTGTGTGGAAGAAGATGCTCGTGCTTCCATGAGACCAATTCAG
TGTGACTCTTTTTGGTGTAAAGGATACATTTTGTGCGCCGACTGAAGAATTGCTGCAGCCA
TCTGCGAATAAAGATGTTGGGGGCAAGGTGTGTATGGCTTTATCGTGTCAAGAAGACAAC
TCCGCTCGACATTGCACAATCTATGGACTGACCAAACTCGTGGTATAAAAAATTATGTTT
TCGAGGCATACACAGACAGACCGGCTCGAGGTGATGTGTAACGCTGCCACCCAAACCGGA

GATGTGGTGGATAAATTCCTCAGAACTCTGTTTCTCGGTGAAAATTTGGTCCATCAGTCA
ATTTTGGAAACGGAAGTCAAAACGACCGCAAAAAATACTTTTGTATGTCTCTGATCCTCGA
ATCGACTCGGTGTACGATACGACAGTGTTCAGTGTCTATGGCAACCGATGATGTAGATAGT
GAAAATGTTTCAGGGCGGTGCGTCTTTGGCGCAAGAAAAACCTTTAAAAGGTTATTGTATC
ATTGCTACTCCATCAGAATGCAAACCGAACATTCATTGGCTTAAGAGCCCAGAAAATGAT
GTTTCAGGAATCGGCAGCTGTGCTGAGGTAG

>5814_U19

ATGGCATTACCCGGTATGAACGGCGCGCATGTTGCAGTTTAAGGATAAGATGATTTCT
TCGGCGGGGTTTCGCTCTGAAATTTGAGAAAGTTGTACAGGAGGCTATGGCTTCCGGAATA
GTTCTTCAACATATCATTGTATTAAGTTTCGGATTTGTGACAATTCGGATATTTTGTCT
GACCGTCAGTTGAGATGTTTATTAATAAACGGGTTGTATCCTTTTGAGGGTAGGATAAGT
ATGTTTGGTGTAAACGGAAGAATGGGAAGGTGCGTCTGCGGCTCCGGAGAGACAGGTTGTT
TTTTTGTATCGTCAACTGGTCAAGTGTTCGGATATGAAGATGGTGTGGTATTTTATTTG
TCACCTACGTTTTTTCAGATTTTTGGACAACCTGCCATGGAGTTCTCGTGTGAGAACGCGATC
TTGAGTAATTTTATCGCACAGAAGTCCCAGCATGAATACAGTGTATCAATTTCAAAAATAT
TTTACGAGAATGAGACATACCCCAATTTTTTTGACAGGTGATTGCCTAGACGGTTTCAG
AAGGTAGAGAGCGGTGATTGTGTGGAAGAAGATGCTCGTGCTTCATGAGACCAATTCAG
TGTGACTCTTTTGGTGTAAAGGATACATTTTGTGCGGCCACTGAAGAATTGCTGCAGCCA
TCTGCGAATAAAGATGTTGGGGCAAGGTGTGTATGGCTTTATCGTGTCAAGAAGACAAC
TCCGCTCGACATTGCACAATCTATGGACTGACCAAACTCGTGGTATAAAAATTATGTTT
TCGAGGCATACACAGACAGACCGGTCTGAGGTGATGTGTAACGCTGCCACCCAAACCGGA
GATGTGGTGGATAAATTCCTCAGAACTCTGTTTCTCGGTGAAAATTTGGTCCATCAGTCA
ATTTTGGAAACGGAAGTCAAAACGACCGCAAAAAATACTTTTGTATGTCTCTGATCCTCGA
ATCGACTCGGTGTACGATACGACAGTGTTCGGTGTCTATGGCAACCGATGATGTAGGTTGT
GAAAATGTTTCAGGGCGGTGCGTCTTTGGCGCAAGAAAAACCTTTAAAAGGTTATTGTATC
ATTGCTACTCCATCAGAATGCAAACCGAACATTCATTGGCTTAAGAGCCCAGAAAATGCT
GTTTCAGGAATCGGCAGCTGTGCTGAGGTAG

U47

>2284_U47

ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCAACA
TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTACATCTCGCAACTC
GGCCTTTTTCTGGGCCTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAAATACACG
TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAAGA
GAACATTTTTACGACCTCCGAAACTTGTATAACAAGTTTTTGTGACACAAATCTATCCCTC
GACTGCTTCACTCAAATTTTAACATAACGTTTTCTCTTGGGACATTTCGAGATTCACAATGC
AAGTCTGCGGTTAACTTGTCTCCATTACAGAATTTACCGCGGACAGAAACCAAATAGTG
CTATCGTGCAGCGCAGCAAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT
CTGTTTGGCACACTATCTACATATAACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT
AAAATTTCTGGGAGCACAGTTATTTGACATAAACTGACCGAAGAATCCTTGCAGGATGGCA
ATGAGCAAATTTTCCAACCGAATCTGACACGGTCATTGACTTCGTTACGTCAGAAAACC
TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT
TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCTTTTGGGA
CGAACCATAGTAAACTTTGATTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT
CTCACATCACCATATCCCTAGTACCAACATACCAACCCAGAACCTCCCCCGTAACA
AAAACTCTACAAAACGATACAGACACCATAAAAAGTTACACCGAACACACCCACCATA
ACAACGCAAACAACGGAAGCATCAAAAAAATAGTTAAACGTTTCAGATTTTCTCGACCC
ATGTACACCCCAACCGACATTCCAACTCTTACAATCCGTCTTAACGCCACTATTAACC
GAACAAAACACCGAAAACCAAAAAGTCCACCAAAACCAACAAATTTTGAATAACACA
ATCAGAATTCCTAAAACCTTTGAGAGCGCTACAGTAACAACAACAACCAACTCAAAAGCTC
GAAAGCACCACTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA
CCAAAAACTCTATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAATAATTCACCGAC
GCCGAACACACCCTCCGATTTTAAAGTTTACCCTTGGCAAAAACACGGCACGCACATAC
ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCAGAGGACAACCTTGA
TCCTCAAACGAGCTACCTACCATTTCAGACGTTGTCTGTCACTCCAAAACAAAACCTACCG

TCGAATGTAAC TGCCAAAAC TGAAGTACACATAACTAACAAATGCTTTACCATCTAGTAAT
TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAAC TAGAATGT CAGCACCC
ACTCACGAAGAGATAAACCACACAGAAATAGCACGAATGACACCAATTCCTAACGCTCAC
ACATCGGAAAAAATACACCTCAACGGCCCTTCACCGCTGAAACGTTCTTAACGACATCT
TCAAAGCCTACCATCCTAACCTGGTCAAACCTTACTAACAACAACGCCTAAGGAACCATTA
ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA
ACTCAATCAGCCAAACTATCAAAAGCTAACATCTCGTCACAAACGACTAACATCTACCCC
CAAACAATCACGGGACGATCTACAGAGGTTTAA

>5055_U47

ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA
TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTACATCTCGCGACTC
GGCCTTTTTCTGGGCCTTTTTCGCGATAGTCATACACTCCGTTAATCTAATAAAAATACACG
TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAAGA
GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGT CAGACAAATCTATCCCTC
GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATT CACAATGC
AAGTCTGCGGTAACTTGTCTCCATTACAGAATTTACCGCGGACAGAAACCAAATAGTG
CTATCGTCGACGGCTGCAAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT
CTGTTTGCCACACTATCTACATATAACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT
AAAATTC TGGGAGCACAGTTATTTGACATAAAACTGACCGAAGAATCCTTGCGGATGGCA
ATGAGCAAATTTTCCAAC TCGAATCTGACACGGTCATTGACTTCGTT CACGTCAAAAAAC
TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT
TCAAATGATCAATATTTCAAACAGTCGCCAAAACCTATAAATGTTACC ACTTCCCTTTGGA
CGAGCCATCGTAAACTTTGATTTCGATACTAACTACTACCCATCATCGACGT CAGCGTCT
CTCACATCACCACATATCCCTAGTACCAACATACCAACCCAGCACCTCCCCCGTCACA
AAAACTCTACAAAAC TGCATACAGACACCATAAAAAGTTACACCGAACACACCCACTATA
ACAACGCAAACAACGGAAAGCATCAAAAAAATAGTTAAACGTT CAGATTTTCTCGACCC
ATGTACACCCCAACCGACATTTCAAAC TCTTACAATCCGTCTTAACGCCACTATTA AAAACC
GAACAAAACACCGAAAACCCAAAAGTCCACCAAAAACCAACAAATTTTGAAAATACCACA
ATCAGAAATCCCAAAACCTTTGAGAGCGCTACAGTAACAACAACGCAACCCAAAAGATC
GAAAGCACCACCTTCAACA AATAGGAATCAAGGAAATTAACGGCAATACCTATTCTTCA
CCAAAAAATATATTTATCTTAAAGAGCAAATCACAGCAGAGTACAACA AAATTCACCGAC
GCCGAACACACCACCTCCGATTTTAAAGTTTACC ACTTGGCAAAAACACGGCAGCACATAC
ATGAGTCACAACACAGAAGTACAAAACATGACCGGCAAATTCAGAGGACAACCTTGAAA
TCCTCAAACGAGCTACCTACCATT CAGACGTTGTCTGTCACTCCAAAACAAAAC TACCG
TCGAATGTAAC TGCCAAAAC TGAAGTACACATAACTAACAAACGCTTTACCATCTAGTAAT
TCATCATACTCAATCACTAAAGTCACTAAAGAGGTAAAGCAAAC TAGAATGT CAGCGTCC
ACTCACGAAGAGATAAACCACACAGAAATAGCACAAATGACACCAATTCCTAACGCTCAC
ACATCGGAAAAATCAACTACACCTCAACGGTCTTTCACCGCTGAAACGTTCTTAACGACA
TCTTCAAAGCCCGCCATCCTAACCTGGTCAAATTTATTATCAACAACGCCTAAGGAACCA
TTAACGAATACAAGTCTAAGGTGGACAGATCATATCACTACACAGCTAACGACTAGCAAT
AGAACTCAATCAGCCAAACTAACAAAAGCTAACATCTCGTCACAAACGACTAACATCTAC
CCCCAAACAATCACGGGACGATCTACAGAGGTTTAA

>5814_U47

ATGCATTTGGAGGTCATCGTCCAATCCTATAAAAAAAGCAAATACTATTTTAGTCACACA
TTCTACTTGTACAAATTCATTGTGGTCAACTCACCCGACATGCTTACATCTCGCAACTC
GGCCTTTTTCTGGGCCTTTTTCGCGATAGTCATGCACTCCGCTAATCTAATAAAAATACACG
TCTGATCCCTTAGAAGCTTTCAAACCGTCAACCGCCACAATTGGAGCGACGAACAAAAGA
GAACATTTTTACGACCTCCGAAACTTGTATACAAGTTTTTGT CAGACAAATCTATCCCTC
GACTGCTTCACTCAAATTTTAACTAACGTTTTCTCTTGGGACATTCGAGATT CACAATGC
AAGTCTGCGGTAACTTGTCTCCATTACAGAATTTACCGCGGACAGAAACCAAATAGTG
CTATCGTCGACGGCAGCAAACAAATCTATCATCGCTAGCAGTTTTTCTCTATTTTATCTT
CTGTTTGCCACACTATCTACATATAACCGCAGATCCGCCATGCGTAGAGTTACTACCATTT
AAAATTC TGGGAGCACAGTTATTTGACATAAAACTGACCGAAGAATCCTTGCGGATGGCA
ATGAGCAAATTTTCCAAC TCGAATCTGACACGGTCATTGACTTCGTT CACGT CAGAAACC
TTCTTTAATTACACCAGCTTTGTTTACTTCTTGCTCTATAACACAACATCATGCGTCCCT
TCAAATGATCAATATTTTAAACAGTCGCCAAAACCTATAAATGTTACTACTTCCCTTTGGA

CGAACCATAGTAAACTTTGATTTCGATACTAACTACGACACCATCATCGACGTCAGCGTCT
CTCACATCACCACATATCCCTAGTACCAACATACCAACCCAGAACCTCCCCCGTAACA
AAAACTCTACAAAACGTCATACAGACACCATAAAAAGTTACACCGAACACACCCACCATA
ACAACGCAAAACGGAAGCATCAAAAAAATAGTTAAACGTTTCAGATTTTCTCGACCC
ATGTACACCCCAACCGACATTCCAACCTTTACAATCCGTCTTAACGCCACTATTAACC
GAACAAAACACCGAAAACCCAAAAAGTCCACAAAACCAACAAATTTTGAAAATACCACA
ATCAGAATTCCCAAAAACCTTTGAGAGCGCTACAGTAACAACAACAACCAACTCAAAAAGCTC
GAAAGCACCACCTTCGCAACAATAGTAATCAAGGAAATTAGCGGCAATACCTATTCTTCA
CCAAAAAATCTATTTATCTTAAGAGCAAATCACAGCAGAGTACAACAAAATTCACCGAC
GCCGAACACACCCTCCGATTTTAAAGTTTACCCTTGGCAAAAACACGGCACGCACATAC
ATGAGCCACAACACAGAAGTACAAAACATGACCGGCAAATTCAGAGGACAACCTTGAAA
TCCTCAAACGAGCTACCTACCATTTCAGACGTTGTCTGTCACTCCAAAACAAAACTACCG
TCGAATGTAACGTCAAAACGTAAGTACACATAACTAACAATGCTTTACCATCTAGTAAT
TCATCACACTCAATCACTAAAGTCACTAAAGAGCTAAAGCAAACCTAGAATGTCAGCACCC
ACTCACGAAGAGATAAACCACACAGAAATAGCACGAATGACACCAATTCCTAACGCTCAC
ACATCGAAAAAATACACCTCAACGGCCCTTACCGCTGAAACGTTCTTAACGACATCT
TCAAAGCCTACCATCCTAACCTGGTCAAACCTACTAACAACAACGCCTAAGGAACCATTA
ACGAATACAAGTCTAAGGTGGACAGATCATATCACAACACAGCTAACGACTAGCAATAGA
ACTCAATCAGCCAACTATCAAAAAGCTAACATCTCGTCACAAAACGACTAACATCTACCCC
CAAACAATCACGGGACGATCTACAGAGGTTTAA

U54

>2284_U54

ATGCAACCCGCCACTCTACAATGGTCTTCGTACGTAACAGCTAAGAGTAACAACCTACC
GCCATATTAACCCGGCGAAGTGTGAGATTGTTTAAATGCGGATTAGGTATCACCCCCCT
CCATCCGCCGTAGTGTGCACCTGTAGAGACGAAAGCAGTTTTTGCATCATCGCCTTTTACT
TACATCGATCCAAAAGATTACGGCAACATACCCTTCGCGGTGCGTAACAACCTCAGATCTA
GATCTGGATCTCAGTAGCTTACCTATCGTCCTTAATATATTTGCACTTCCTTACGCAAAG
GTGAATATATCAAATTTACCGGTGCAACACATCGAAACCCACGATCGCCACATTATACCC
CACGGACAGTGCAGCGCAAGTTTGTATATACGGCCCTCTCACGCGTATTAGAATTA
GTGGCTGATATACGATGGATAGAACAACACCAGAAGAGCCAACACGATATTCATTTAAG
GCAGAAATATGGGTTAATCTCCAAAACACTCCGCTCGACCAAATTTTTTAAAAATGCTGAC
ATCGAATTTATATCACACAAAACGCTCTACATTTCTAGAATCTTACTCTGCGGAAATCTA
TTAATTTTGAAGGCAATATACGAAAACGACTATCTCCGTGACGATAATTTTTATCTGAA
CAACTTGACATACAAATTTCACTGACCCCCCAAACCGCAGACATTACACTGTGCGAGAAT
CAAGAAACACTTCTAAAATGCAACACCCATTCAATAACTGTCTGCGCAACTAAAAATATC
ATCTCTAACATGGTTACTCCTTTCCATTGCTCTTTCAACAATATCTTTAATTCAAATTC
AACTTTACGGGTTTATTTATACCCAAACTCTTGCTCGGAATCTCTACAACCGCCGGCATT
TGGGATGAGACACGGCCTCTCTACATCACCATGAAAGCATTAAAGAAGAACCCTCGCATC
AATTACTCGCAACCTATAGGCGTAGTATACTTCTTCCCCAAACAAATTTTACCACCGGGG
AATAATGTGGAATTCAACTGGGCAGAAATGAGCAAAATTTATGTAACACAGAATCTCCA
AATGGTCCCGTAAGAAACACCGTAACTTTCAAAAACCAAGCAGTATTGCGAACACCAAGT
CTAAGCACGGTTGCAAACCTAACGCCAGACATCAACATGGGAATTTTTTTGAGTTCACTC
CGCGTTGCGTTCGACACCGCACATATGGTACCCTTGCACCTTTCTCTAAAACCCGGAGAA
TCTACTCGCATGGAATTCATGCCGCCCGGAACCCACAAAACCTAACGATATTAGAAGGA
GACGTCGGCATTCATTTATTCTTGTGTCACAATCACTCCCATAGATCTTCCCCGTAA

>5055_U54

ATGCAACCCGCCACTCTACAATGGTCTTCGTACGTAACAGCTAAGAGTAACAACCTACC
GCCATATTAACCCGGCGAAGTGTGAGATTGTTTAAATGCGGATTAGGTATCACCCCCCT
CCATCCGCCGTAGTGTGCACCTGTAGAGACGAAAGCAGTTTTTGCATCATCGCCTTTTACT
TACATCGATCCAAAAGATTACGGCAACATACCCTTCGCGGTGCGTAACAACCTCAGATCTA
GATCTGGATCTCAGTAGCTTACCTATCGTCCTTAATATATTTGCACTTCCTTACGCAAAG
GTGAATATATCAAATTTACCGGTGCAACACATCGAAACCCACGATCGCCACATTATACCC
CACGGACAGTGCAGCGCAAGTTTGTATATACGGCCCTCTCACGCGTATTAGAATTA
GTGGCTGATATACGATGGATAGAACAACACCAGAAGAGCCAACACGATATTCATTTAAG

GCAGAAATATGGGTTAATCTCCAAAACACTCCGCTCGACCAAATTTTTAAAAATGCTGAC
ATCGAATTTTATATCACACAAAACGTCTACATTTCTAGAATCTTACTCTGCGGAAATCTA
TTAATTTTGAAGGCAATATACGAAAACGACTATCTCCGTGACGATAATTTTTATCCTGAA
CAACTTGACATACAAATTTCACTGACCCCCAAACCGCAGACATTACACTGTGCGAGAAT
CAAGAAACACTTCTAAAATGCAACACCCATTCAATAACTGTCTGCGCAACTAAAAATATC
ATCTCTAACATGGTTACTCCTTTCCATTGCTCTTTCAACAATATCTTTAATTCAAATTCA
AACTTTACGGGTTTATTTATACCCAAACTCTTGCTCGGAATCTCTACAACCGCCGGCATT
TGGGATGAGACACGGCCTCTCTACATCACCATGAAAGCATTAAAGAAGAACCCTCGCATC
AATTACTCGCAACCTATAGGCGTAGTATACTTCTTCCCAAACAAATTTTACCACCGGGG
AATAATGTGGAATTCAACTGGGCAGAAATGAGCAAATTTATGTAAACACAGAATCTCCA
AATGGTCCCGTAAGAAACACCGTAACTTTACAAAACCAAGCAGTATTGCGAACACCAAGT
CTAAGCACGGTTGCAAACCTAACGCCAGACATCAACATGGGAATTTTTTTGAGTTCACTC
CGCGTTGCGTTTCGACACCGCACATATGGTACCCTTGCACTTTTCTCTAAAACCCGGAGAA
TCTACTCGCATGGAATTCATGCCGCCCGGAACCCACAAAACCTAACGATATTAGAAGGA
GACGTGGCATTCAATTCATTCTTGTGACAATCACTCCCATAGATCTTCCCCGTAA

>5814_U54

ATGCAACCCGCCACTCTACAATGGTCTTCGTACGTACTACAGCTAAGAGTAACAACCTACC
GCCATATTTAAACCCGGCGAAGTGTGAGATTGTTTAAATGCGGATTAGGTATCACCCCCCT
CCATCCGCCGTAGTGTGCACTTGTAGAGACGAAAGCAGTTTTGCATCATCGCCTTTTACT
TACATCGATCCAAAAGATTACGGCAACATACCCTTCGCGGTGCGTAACAACCTCAGATCTA
GATCTGGATCTCAGTAGCTTACCTATCGTCCTTAATATATTTGCACTTCCTTACGCAAAG
GTGAATATATCAAATTTACCGGTGCAACACATCGAAACCCACGATCGCCACATTATACCC
CACGGACAGTGCAGCGGAAGTTTGTGATATACGGCCCTCTCACGCGTATTAGAATTTAA
GTGGCTGATATACGATGGATAGAACAACACCAGAAGAGCCAACACGATATTCATTTAAG
GCAGAAATATGGGTTAATCTCCAAAACACTCCGCTCGACCAAATTTTTAAAAATGCTGAC
ATCGAATTTTATATCACACAAAACGTCTACATTTCTAGAATCTTACTCTGCGGAAATCTA
TTAATTTTGAAGGCAATATACGAAAACGACTATCTCCGTGACGATAATTTTTATCCTGAA
CAACTTGACATACAAATTTCACTGACCCCCAAACCGCAGACATTACACTGTGCGAGAAT
CAAGAAACACTTCTAAAATGCAACACCCATTCAATAACTGTCTGCGCAACTAAAAATATC
ATCTCTAACATGGTTACTCCTTTCCATTGCTCTTTCAACAATATCTTTAATTCAAATTCA
AACTTTACGGGTTTATTTATACCCAAACTCTTGCTCGGAATCTCTACAACCGCCGGCATT
TGGGATGAGACACGGCCTCTCTACATCACCATGAAAGCATTAAAGAAGAACCCTCGCATC
AATTACTCGCAACCTATAGGCGTAGTATACTTCTTCCCAAACAAATTTTACCACCGGGG
AATAATGTGGAATTCAACTGGGCAGAAATGAGCAAATTTATGTAAACACAGAATCTCCA
AATGGTCCCGTAAGAAACACCGTAACTTTACAAAACCAAGCAGTATTGCGAACACCAAGT
CTAAGCACGGTTGCAAACCTAACGCCAGACATCAACATGGGAATTTTTTTGAGTTCACTC
CGCGTTGCGTTTCGACACCGCACATATGGTACCCTTGCACTTTTCTCTAAAACCCGGAGAA
TCTACTCGCATGGAATTCATGCCGCCCGGAACCCACAAAACCTAACGATATTAGAAGGA
GACGTGGCATTCAATTCATTCTTGTGACAATCACTCCCATAGATCTTCCCCGTAA

U65

>2284_U65

ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAATTTTCTGCAGAAT
GAATGTAACGGTTCAGAATTTGTAAAAAACATTCTATCGCGAATATCGTAGCGTTGCG
ACATCGTCTCCTACATTCCTGTTGAATAATAAACCTAAGAAATTTTGTGCTTGCATTGTGAG
ATTGTAATACTCAAGCGAAGTGAGAAATTTATGTTTACGCTTTCGCGTAAACGGCATAACAT
TTTGGGCAGTTTTTACCGGAAGAATGAAATTTAATAAGAAATCAGTTCCGGAAGGGCTC
TATTACTATATATTGGAATTGGGAAGCATAACCCCTTATCGATTTGAACTTTATCCCGAGA
TATAATTCCGACTGTGTTACAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT
TGCTCTATTGTGTGTTCCGAAGAGGCAATTCGCCTCACAGTAAAAGGGTCCGGGGACAAT
AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT
ATCTATACTTTTGCCTCTCTTACGATCTTTTCCCTAACTTGTATGACAAATCCATCTTT
CCATCTCTGGCAAAAATTTATTTTGTATGATAGCTTGCGAATCCGAAGATTGTGTCTTT
TGTAAGATCACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTTAAT
CAAGAAACCTGTTTTTGTACACACCGTGTAAGAAAAAATGGCTAATATTAACAATCCG

GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA
AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGAC
CCTTATGCGTTAAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG
CTGATTATTTTGTCTGTCTCTGTGTGTAAGCGTGTGGTAATGGACTAA

>5055_U65

ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAAATTTTCTGCAGAAT
GAATGTAAC TGGTTCAGAAATTTGTAACAAAAACATTCTATCGCGAATATCGTAGCGTTGCG
ACATCGTCTCCTACATTCTCGTTGAATAATAAGCCTAAGAAAATTTTGCATGCATTGCGAG
ATTGTAATACTCAAGCGAAGTGAAGAATTTATGTTTACGCTTTCGCGTAAATGGCATAACAT
TTTGGGCAGTTTTTAACCGGAAAAATGAAATTTAATAAGAAAAGCAGTTCCGGAAGGGCTC
TATTACTATATATTGGAATTGGGAAGCATAACCCCTATCGATTTGGGCTTTATTCCGAGA
TATAATTCCGACTGTGTTACAAACATGCGTTGTGTTACACCGGAGGTTATTTATGAAAAT
TGCTCTATTGTGTGTCCCGAAGAGGCAAATCGCCTCACGGTAAAAGGGTCCGGGGACAAT
AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT
ATCTATACTTTTTGACTCGCTTACGATCTTTTCCCTAACTTGTTATGACAAATCCACCTTT
CCATCTCTGGCGAAAAATATTTTTGATATGATAGCTTGCGAATCCGAAGATTGTGTCTTT
TGTAAGATCACAAACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTAAT
CAAGAAACCTGTTTTTGTACACACCGTGTAAAGAAAAAATGGCTAATATTAACAATCCG
GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA
ATAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTTCCGGTGACGAC
CCTTATGCGTTAAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG
CTGATTGTTTTGTCTGTCTCTGTGTGTAAGCGTGTGGTAATGGACTAA

>5814_U65

ATGGCGATATCGACTTTTTCAATTGGCGATTTGGGTTATCTTAGAAAATTTTCTGCAGAAT
GAATGTAAC TGGTTCAGAAATTTGTAACAAAAACATTCTATCGCGAATATCGTAGCGTTGCG
ACATCGTCTCCTACATTCTCGTTGAATAATAAACCTAAGAAAATTTTGCATACATTGTGAG
ATTGTAATACTCAAGCGAAGTGGAGAATTTATGTTTACGCTTTCGCGTAAACGGCATAACAT
TTTGGGCAGTTTTTACCGGAAGAATGAAATTTAATAAGAAAAGCAGTTCCGGAAGGGCTC
TATTACTATATATTGGAATTGGGAAGCATAACCCCTATCGATTTGAACTTTATCCCGAGA
TATAATTCCGACTGTGTTACAAGCATGCGTTGTGTTACACCGGAGCTTATTTATGAAAAT
TGCTCTATTGTGTGTCCCGAAGAGGCAAATTCGTCTCACAGTAAAAGGGTCCGGGGACAAT
AAATTGACTCCCTTAGGTGGGTGTGGAGCATGGTGTCTGAAAAATGGTGGCGATCTGTAT
ATCTATACTTTTTGACTCGCTTACGATCTTTTCCCTAACTTGTTATGACAAATCCATCTTT
CCATCTCTGGCAAAAAATATTTTTGATATGATAGCTTGCGAATCCGAAGATTGTGTCTTT
TGTAAGATCACAAACAAACATGTATCGCAAGCTGGACAGATTGTAGGGTGCCTCTAAT
CAAGAAACCTGTTTTTGTACACACCGTGTAAAGAAAAAATGGCTAATATTAACAATCCG
GAGTTAATCTCTCTGCTCTGTGATCAGGAAATTAATAAGATAGATATTATGTATCCCGAA
AGAAAAGCATCGTTATCACTGGACATTAATTCTTACGCTCATGGGTACTCCGGCGACGAC
CCTTATGCGTTAAAAATGTGTTAATTGGATACCAGTCAGGATCAGCGCAGCTCTGAGCAGG
CTGATTATTTTGTCTGTCTCTGTGTGTAAGCGTGTGGTAATGGACTAA

U71

>2284_U71

ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAAACACC
TTGATTGATTACAAGGGAAATCCTATATCACTAGCGGATGAATTTACTGTTTTAACCGAT
ACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAAAGCCCTTACTTGAAAAAGTACAG
ACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAATCAAAA
AAATGA

>5055_U71

ATGGGCGCAAAGTGCTGTAAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAAACACC
TTGATTGATTACAAGGGAAATCCTATATTACTAGAGAATGAATTTACTGTTTTAACCGAT
ACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAAAGCCCTTACTTGAAAAAGTACAG
ACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAATAA

>5814_U71

ATGGGCGCAAAGTGCTGTAACCAGTATCTTGCGGCATGTGTAAGAAAACCGAAAAACACC
TTGATTGATTACAAGGGAAATCCTATATTACTAGAGAATGAATTTACTGTTTTAACCGAT
ACCGAATCGGAAGAGGAAGGGATGGCAGATCTTGAAGGCCCTTACTTGAAAAAGTACAG
ACCACATTAGTTGCTAAGTGTGATACGGAAGCTGAAAAAAAATTGCCTTGTAATAA

U79

>2284_U79

ATGTATGCTGAAGAACGTGGATATGGATCATTGACGATGTTGAAGAAGTTTATGAACAA
ATTGTTACCGAGTCTATTACAGCTGAAAATGCTTCATTTCCGGCAGTGGCCGTTTAATTGAG
TTCTTAGGTGACTCTGGAACGTGTGAAACATTTTGTAAATGGATGGATATCACTGATTTGT
TGGACATCCGATACAAATTCAATGGGTCTCTAACCGGTGGATATTGGGATAGATGATGGA
AAATATAAACGTACAGTGTCTGGGGTGTGTTACTGTGTTTCCGAAATCGATCACGTCGATT
TCTCAGAATACTGAGGGCAGAGATAGGATTTTACTGTTTCCCAAGAAAACAGAAAACCTT
CAAATAACGTTTGTACTCTCCTAAAGTTTTTCAAAGATCAGGATATACGAAACCCGGGC
GATCTGAAATGTATGGAACAGATTGAAAAAGAATGTCGAAATTTAGACAGAAAAAACAC
GACGATGAACACAGAAAACGCTCGGGGAAGCAAAAAGAAAAAGAAAAGTGGAGGACACG
GACAAAAAAAAGATGATGACAGACGAAAACACGAAGAAAAGAAAACGAAACGACGAAGAC
AAGCAACCGGACAAAAGGAGGAATCGGACGAACCTCCTAAGGAGAAGAGACAGAAGTAC
CATGACATGAAACGTAATTTGGAAGAACAAGTACAGAGGATGGCATAATACCAACTTCT
ACGACACTCGTAAATGGAGCGGTTGAGGGTGCCTTACCGCCCTGTATTTCTATTGATAAT
CACGAAGATCAACAACATGATGAATTAGACAAGCGCGCTTATGCGCAGGGGACGAACCGA
GAAGTCTATCCAATGAAGATAATTATGGGAATTTTTCGGCTTAATAAGTCTTTAGAACAG
CTTAGGGCCAACTTGTGTAAGCAGCGGCGACATTGTAGAAAGATCGCTTTTGAAATTG
AAAGAGTGTGTTGGATAATGTAAGGATAATTTAATAAAAAATGAATGTGTTGATGTTGCT
GTTCCAAGTAAATGTTTAAAGTAAAAACAAAACATATCGAGCCAAAGAAAACAGATAGTGT
TCTGATTGTATCAGGCCGGTACCCGTGTGTGAGATTAACCGTTTATCGATGTGGGAATA
TTTGAAACGACGCCGAACGCAAGAAGACTTCGACAACCGACCAGGACAATCGTCGGATCA
ACAGATGGTGCATCGAGCAACAGCGTGTATTTTCGGGACAAAACCGGGGCAGAGCACGA
GGACGTGGCCGAGGGAGAGCTCCCAGGAGACGAAATTCGAATATAAATAATTCAGGGACA
CAAACCTGCAATTGTTATAGACGATAGCAGCGAAGCCGAAAACCTTGAAAATGAGGGGAGT
TTTAATGAAGATTTACTGGCAACCACAGTATTGGAACATTGTAA

>5055_U79

ATGTATGCTGAAGAACGTGGATATGGATCATTGACGATGTTTTAGAAGTTTATCAACAA
ATTGTTACCGAGTCTATTACAGCTGAAAATGCTTCATTTCCGGCAGTGGCCGTTTAATTGAG
TTCTTAGGTGACTCTGGAACGTGTGAAACATTTTGTAAATGGATGGATATCACTGATTTGT
TGGACATCCGATACAAATTCAATGGGTCTCTAACCGGTGGATATTGGGATAGATGATGGA
AAATATAAACGTACAGTGTCTGGGGTGTGTTACTGTGTTTCCGAAATCGATCACGTCGATT
TCTCAGAATACTGAGGGCAGAGATAGGATTTTACTGTTTCCCAAGAAAACAGAAAACCTT
CAAATAACGTTTGTACTCTCCTAAAGTTTTTCAAAGATCAGGATATACGAAATCCGGGC
GATCTGAAAGGTATGGAAGAGATTGAAAAAGAATGTCGAGCTTTAGACAGAAAAAACAC
GACGATGAACACAGAAAACGCTCGGGGAAGCAAAAAGAAAAAAAAGTGGAGGACACGGAC
AAAAAAAAGATGATGACAGACGAAAACAAGAAAAGAAAACGAAACGACGAAGACAAGCAA
CCGGACAAAAGGAGGAATCGGACGAACCTCCTAAGGAGAAGAGACAGAAGTACCATGAC
ATGAAACGTAATTTGGAAGAACAAGTACAGAGGATGGCATAATACCAACTTCTACGACA
CTCGTAAATGGAGCGGTTGAGGGTGCCTTACCGCCCTGTATTTCTATTGATAATCACGAA
GATCAACAACATGATGAATTAGACAAGCGCGCTTATGCGCAGGGGACGAACCGAGAAGGT
CTATCCAATGAAGATAATTATGGGAATTTTTCGGCTTAATAAGTCTTTAGAACAGCTTAGG
GCCAACTTGTGTAAGCAGCGGCGACATTGTAGAAAGATCGCTTTTGAAATTGAAAGAG
TGTTTGGATAATGTAAGGATAATTTAATAAAAAATGAATGTGCTGATGTTACTGTTCCA
AGTAAATGTTTAAAGTAAAAACAAAACATATCGAGCCAAAGAAAACAGATAGTGTGTTTCTGAT
TGTATCAGGCCGGTACCCGTGTGTGAGATTAACCGTTTATCGATGTGGGAATATTTGAA
ACGACGCCGAACGCAAGAAGACTTCGACAACCGACCAGGACAATCGTCGGAACAACAGAT
CGTATTTTCGGGACAAAACCGGGGCAGAGCACGAGGACGTGGCCGAGGGAGAGCTCCTAGG

AGACGAAATTCCAATATAAATAATTCAGGGATACAAACTGCAATTGTTATAGACGATAGC
AGCGAAGCCGAAAACTTTGAAAATGAAGGGAGTTTTAATGAAGATTTACTGGCAACCACA
ATATTGGAAACATTGTAA

>5814_U79

ATGTATGCTGAAGAACGTGGATATGGATCATTGACGATGTTGAAGAAGTTTATGAACAA
ATTGTTACCGAGTCTATTCAGCTGAAAATGCTTCATTTCCGGCAGTGGCCGTTTAAATTGAG
TTCTTAGGTGACTCTGGAACGTGTGAAACATTTTTGTAATGGATGGATATCACTGATTTGT
TGGACATCCGATACAAATTCATGGGTTCTCTAACGGTGGATATTGGGATAGATGATGGA
AAATATAAAACGTACAGTGTCTGGGGTTGTTTACTGTGTTTCGAAATCGATCACGTTCGATT
TCTCAGAATACTGAGGGCAGAGATAGGATTTTTGACTGTTTTCCAAGAAAACAGAAAACCTT
CAAATAACGTTTGTACTCTCCTAAAGTTTTTCAAAGATCAGGATATACGAAAACCCGGGC
GATCTGAAATGTATGGAACAGATTGAAAAAGAATGTGAAAATTTAGACAGAAAAAACAC
GACGATGAACACAGAAAACGCTCGGGGAAGCAAAAAGAAAAAGAAAAGTGGAGGACACG
GACAAAAAAAAGATGATGACAGACGAAAACACGAAGAAAAGAAAACGAAACGACGAAGAC
AAGCAACCCGACAAAAGGAGGAATCGGACGAACCTCCTAAGGAGAAGAGACAGAAGTAC
CATGACATGAAACGTAATTTGGAAGAACAAGTCCAGGATGGCATAATACCAACTTCT
ACGACACTCGTAAATGGAGCGGTTGAGGGTGCCTTACCGCCCTGATTTCTATTGATAAT
CACGAAGATCAACAACATGATGAATTAGACAAGCGCGCTTATGCGCAGGGGACGAACCGA
GAAGTCTATCCAATGAAGATAATTATGGGAATTTTCCGGCTTAATAAGTCTTTAGAACAG
CTTAGGGCCAAACTTGTGTCAGCAGCGCGACATTGTAGAAAAGATCGCTTTTGAATTTG
AAAGAGTGTGTTGGATAATGTAAAGGATAATTTAATAAAAAATGAATGTGTTGATGTTGCT
GTTCCAAGTAAATGTTTAAAGTAAAACAAAACATATCGAGCCAAAAGAAAACAGATAGTGT
TCTGATTGTATCAGGCCGGTACCCGTGTGTGAGATTAACCGTTTATCGATGTGGGAATA
TTTGAACGACGCCGAACGCAAGAAGACTTCGACAACGGACCAGAACAACAGATCGTATT
TCGGGACAAAACCGGGCAGAGCAGGAGCGTGGCCGAGGGAGAGCTCCAGGAGACGA
AATCCAATATAAATAATTCAGGGACACAACTGCAATTGTTATAGACGATAGCAGCGAA
GCCGAAAACTTTGAAAATGAGGGGAGTTTTAATGAAGATTTACTGGCAACCACAGTATTG
GAAACATTGTAA

U86

>2284_U86

ATGGAGCCAGCAAAACCCTCTGGAAACAACATGGGATCCAATGATGAACGTATGCAAGAG
TACCGTCCCGATCCAATGATGGAAGAATCTATCCAACAGATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCAAATGAAGACTTG
AAAACTTTAGCTGAACATAATTCCAACAAGGAAATGATCATAACGACGTCAATCCAAAGC
GCCATGAAATTTATCCGGACTATACTGTGATGCGGACATCACACACACACAACCGTTGAGT
GATAATACGCATCAAGATCCTATATATTCTCAAGATTCCGGAATATTTTCAAAAACTATA
CGGGATCCACGGTTAGCAGCTCAGTTCCACAGACAATGCACATCCTCTGCTTCCAATCTA
CCGTCAAACGAAAGCGGTTTCGACCCAGGTGCGCTTTGCTTTCGGAACTACCAAGTCAATTC
GTACAGCCCAAGTACACGTACATAACCAAAATGCAAATCTACAGAACAATTTACCTCG
TTGCCGTATCAACCTTATCATGATCCATACAGAGATATAGAGAGTTCTTACAGAGAATCT
AGAAATACCAATAGAGGATATGACTATAACTTTAGACATCATTTCGTATAGGCCCAGAGGA
GGAAATGGAATAATAATTACTACAATCCAAACTCAAAGTATCAACAGCCATACAAAAGA
TGTTTTACCAGAACTTATAACAGACGCGGAAGAGGCCACAGATCATAACGATTGCTCAGAC
AGATCCGCGGATTTACCATATGAACATTACACGTATCCAAACTACGAACAACAGAATCCA
GACCCCGTATGAATAACTATAAAGATTTTACACAACCTAACAAAACAAGTTCAATTTTGA
GCCAATGACTATAGCATGGCATTTCACAGACTCCACACACGTACAATCTGACAATTAC
AATCATCCAACAAAAGCACAGACAATACCAGAACTACAAAAACTAAAAAACATGAGGCT
ACTAAGGATAACGAAACATCCAGAGAAAATCAAGTTTTAATCATCTAATGACGCGATCAGT
TTATCATAACAGACCTAGTCCACAAAACGAAATATTATTAATAAATAATATGACACTGAC
GTTATTCGGTTACCGAAGGAAGCGTTAACAGGCAATGGCTCAAATCGTGACGTGGATATT
CAGAAGTACAAAAAGCACATATCCGTTGTAGATCAGTCCAGAAGAAAAAAGAACGCGAGT
TCTCAAACAAAACATGATGAAAACCATGCCTCATCACGTTCTGATCTTAAAGAGAGA
AAATCTAATGAGCATGAAGATAAAGCCGTTACTAAAGCTAGAGATTTTTCAAAATTAGAT

CCCCTATTATCTCCTCTTCCGTTGACACCTGAACCCGCCATAGACTTTGCGGACCATACA
GATAAATTCCTATTTCGACACCAGAATTC AACCAAATAAAAACAAAATCTACACCGATCAAAA
ACTTCTCTACAAGACACAGTTC CATATCAAAACATACTCCAAGAGCCCCGACCAAAGAT
AATTCCTTACAAAAACATCATGACTCCAAGACAATTATCCTAAAATGAAACATTCTCCT
GGACGTACAACATCTAAAAAAAATACTACAAACTCCAACGGACACCAAACCTTTAAGGAA
GTATCCGTTAAGAATGTTTCAGGGAAAGCCACATCTACATCACCAAAATCAAAAAACACAC
CATTATTCCAGTAGCTCTGATGAAGAAGGTCAATACAAATCTCCGGTAAAAACTATTATC
CCATCACCTTCTCCATATTGCAAGTTAAAAAATCCTTCAATTATGGACAAAAACAGTGCT
AAAAACCACACTGCAAGCGCCGACAAAAATTTAACAGACAATTCACCTATTTCGGAGTAAT
TTAAATCCTACTGCTTTTAAACAAATCCAACAATAACAAAAGCATCACGAATAGCACTTCT
AACAGTGATGAGTGTACTGACAAGAAACCTA ACTGCAATTCACCAAATAATGAATCTAAA
GATCCTAACCGAACTTGTGGCAAAAACCTCTGATAAACATCTGAGTAAAAGTTGCACCATG
GCTTCTAAAAGGGCTCCCAGCAGAGCCTCAAGCAGAACCTCCAGCAGAGCTTCCAGCAGG
GCCTCAAGCAGAGCCTCCAGCAGAGCTTCCAGCAGGGATTCTAGCAGAGCCTCCAGCAGA
GCCTCCAGCAGGGATTCTAGCAGAGCCTCCAGTAGAGCCTCCAGCAGGGCCTCAAGCAGA
GCCTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGG
GCCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGAGCCTCCAGCAGG
GATTCCTAGCAGGGCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGA
GCCTCAAGTAAAGCCTCAAGAAAAGCCTCCAGCAGAGCCTCCAGCAGGGCTTCCAGCAGG
GCCTCAAGCAAAGCCTCCGGTAAAGCCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGTAGG
AATTCCTAGCAGGGCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGA
GCCTCCAGTAGGGATTCTAGCAGGGCTCAAGTAAAGCCTCAAGAAAAGCCTCCAGCAGA
GCCTCCAGCAGGGCTTCCAGCAGGGCTTCCAGCAGGGCTTCCAGCAGGGAGCTGAGGCAG
ATTTATTGCGATTCTAATAAAAGACAAACGCCTCCACATGACACCTCAATTAATACAAAA
TTTGAAATATCAGAAATAAAATTTGATGTGGAGAAGATCTAAATTTCTACAAAAATACT
GCAGCTCGACTCCAGTGTCTTAAATCACAATGATCAATTTTACAACCCCCGATTTCAGACCT
CATATAAGAACCAACAGAAAAAATCAGAATCCACCAATGACACCGATTTCAGAAAGTTCT
ATGTCTAGATGCAAATCTCACTGTGCAAATTCACCAGATTCACTGACCATTGTAAGAAGA
AAAAACACAAATCTGGATCATCAAGCATCTCTTCCAGTATTGAAGAAAATTCAGATCT
AATTCACATATAGTAACCGGAAAAGAGAAGTTTACACCATTTTACTATCAGTCTTCCAGA
ACTAGATCATCATCATCATCAACATCATCATCAGCTTCTTTATCATGTTCCAAATCC
ACATTA AAAACATGCCGCAAAAACCTCAATATAAAAGACAATAAACAAAATAAAATCTAAATCA
GATTCAAAACACAAAACCTACAAACATGTCTT CAGATTACGAATCGAATAGACATGCAGAT
GTGTTTAGAAAAC TCCCCGGAAGCAGGTGAAAAATTTCCCTACACAATTCTAGTCCATTT
AACACACATGAACAATCCAATCATTTCCGAAAACGCCATTGATGAAGAACAAAAAAGGCA
CCTAATATCACGACTTCTCACCTACATCAAAAAACAAAATGTCAAATTACACAACACAAAA
AAATGCAAAAAAAAACGCCCCCGCGATGATGATT CAGATTCCAGCATAAAAAATTTCTGC
AAGAAAAGAATATCCGGAGCGCAGAAAACAGAGTCTGAAGTAAGTAAAAGTACGATTTG
TGTTATAGAGATTATATTAGACTAAAAGAAAAGAAAAGTATCAGAAAAATTTAAAATTCAC
CGCGGGAGGGTGGCTACAAAAGATTTTCAAAAAC TATTTAGAAAATACTATGAGAGCATTT
GAATACAAACAGATCCCCAAAAACCTTGCAGTGAGAAAATCTAAAAGAAGCTGTATAC
AGCATCTGCTGTAATGGGCTCTCTAACAACGCAGCCATCATTATGTACTTTACAAGATCA
AAAAAGTAGCTCAAATCATAAAGATAATGCAAAAAGAATTAATGATTTCGACCCAACATC
ACTGTTTCAGAGGCATTTAAAATGAATCATGCACCGCCTAAGTATTATGACAAAGATGAA
ATTAAACGATTTATCCAAC TACAAAAACAAGGACCGCAAGAATTATGGGATAAAATTTGAA
AACAAACAACACATGATCTATTTACGAGACATTCGGACGTAAAGACTATGATAATTTAC
GCCGCTACGCCCATCGACTTTGTAGGAGCTGTTAAAACATGCAATAAGTATGCCAAAGAT
AACCTAAAGAAATTTGTCTTCCGGTGTGTTCAATCATCGACGGTGACAATCCTATCTCC
ATCTATAATCCTATCAGCAAAGATTTTAAATCGAAATTTAGTACACTTTCAAAATGTTAA

>5055_U86

ATGGAGCCAGCAAACCCCTCTGGAAACAACATGGGATCCAATGATGAACGTATGCAAGAC
TACCGTCCCGATCCAATGATGGAAGAATCTATCAAACAAATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCACATCAAGACTTG
AAAACTTTAGCTGAACATAATTTCAAACAAGGAAATGATCATAACGACGTCATCCAAAGC
GCCATGAAATTTACCGACTATACTGTGATGCGGACATCACACACACACAACCGTTGAGT
GATAATACGCATCAAGATCCTATATATTTCTCAAGATTCCGGAATATTTTCAAAAAC TATA

CGGGATCCACGGTTAGCAGCTCAGTTCCACAGACAATGCACATCCTCTGCTTCCAATCTA
CCGTCAAACGAAAGCGGTTTCGACCCAGGTGCGCTTTGCTTCGGAACCTACCAAGTCAATTC
GTACAGCCCAAGTACACGTACATAACAAAAATGCAAATCTACAGAACAATTTACCTCG
TTGCCGTATCAACCTAATCATGATCCATACAAAAGGTATAGAGAATTCCTACAGAGAATCT
AGAAATACCTATACAAGATATGACTCTAACTTTAGACATCATTCCAATAGGCCCAGAGGA
GAAATGAAAAATATAATTACTACAATCCAAACTCAAAATATCAACAGCCATACAAAAGA
TGTTTTACCAGAACTTATAACAGACGCGGAAGAGGCTACAGATCATACGATTTCTCAGAC
AGATCCGCGGATTTTTTCATATGATCAATACATGTATCCAACCTACGAACAACATAATCCA
GACCCCGTATGAATAACTGTAAAGATTTTACACAACCTAACAAACAAGTTCAATTTTGAA
GACTATGACTATAGCATGGCATTTCACAGACTCCGCACAAGTACAATCTGATAATTAC
AATCATCCAACAAAAGAACAGACAATACCAGAAACTACAAAAACTAAAAAACATGAGGCT
ACTAAGGATAACGAAACATCCACAGAAAATCAAATTTTAACTGCTCCTGACGCAATCAGT
TTATCATACAGACCTAGTTCAAAAAAATGGATATTATTAATAAATATATGACACTGAC
GTTATTCATTACCGAAGGAAGCGTTAACAGCTAATGGCTCAAATCGTGACGTGGATATT
CAGAAGTACAAAAAGCACATATCCGTTGTAGATCAGTCCAGAAGAAAAAAGAACGCAGT
TCTCAAACAAAACAAACATGATGAAAACCATGCCTCATCACGTTCTGATCTTAAAGAGATA
AAATCTAATGAGCATGAAGATAAAGCCGTTACTAAAGCTAGAGATTTTTCAAATTAGAT
CCCCTATTATCTCCTCGTCCGTTGACACCTGAACCCGCCATAGACTTTGCAGACCATACA
GATAAATCTGTTCGACACCAGAATTC AACCAAAGAAAAACAAAATCTACAGCGATCAAAA
ACTTCTCTACAAGACACAGTTCCCATATCAAAACATACTCCAAGAGCCCCGACCAAAGAT
AATCTTTACAAAAACATCATGACTCCAAAGACAATTATCCTAAAATGAAACATTCTCCT
GGACGTACAACATCTAAAAAATACTACAAACTCCAACGGACACCAAACCTTTAAGGAA
GTATCCGTTAAGAATGTTTCAGGGAAAGCCACATCTACACCACCAAATCAAAAAACACAC
CATTATCCCAGTAGCTCTGATGAAGAAGGTCAATACAAATCCCGGTAAAAACTATTATC
CCATCACCTTCTCCATATTCCAAGTTAAAAATCCTTCAATTATGGACAAAAACAGTGCT
AAAAACCACACTGCAAGCGCCGACAAAAATTTAACAGACAATTCACCTATTTCGGAGTAAT
TTAAATCCTACTGCTTTTAAACAATCTAACAGTAACAAAAGCATCACGAATAGCACTTCT
AACAGTGATGAATGTACTGACAAGAAACCTAACTGCGATTCCACGAAAAATGAATCTAAA
GATACTAACCGAACTTGTAGCAAAAACCTCTGATAAACATCTGAGTAAAAGTTGCACCATG
GCTTCTAAAAGGGCTCCCAGCAGAGCCTCAAGCAGAGCCTCCAGTAGGGATTCTAGCAGG
GCCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGAGATTCTAGCAGAGCCTCCAGCAGA
GCCTCCAGCAGGGATTCTAGCAGAGCCTCCAGTAGAGCCTCAAGCAGGGCCTCAAGCAAA
GCCTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGG
GCCTCAAGCAAAGCCTCAAGCAAAGCCTCCAGCAGGGATTCTAGCAGAGCCTCCAGCAGG
GATTCTAGCAGGGATTCTAGCAGGGCCTCAAGCAGGGCCTCAAGCAGGGATTCTAGCAGG
GCCTCAAGCAAAGCCTCAAGTAAAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGG
GCCTCAAGCAAAGCCTCAAGTAAAGCCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGG
GATTCTAGCAGGGCCTCAAGCAGGGCCTCAAGCAAAGCCTCCAGCAGGGATTCTAGCAGG
GCCTCAAGCAGGGCCTCCAGCAGGGATTCTAGCAGGGCCTCAAGCAAAGCCTCAAGTAAA
GCCTCCAGCAGAGCCTCCACTTCCAGCAGGAAGCTGAAGCAGATTTATCACGATTCTAAT
AAAATACAAACACCTCCACATGACACCTCAATTAATACAGAATCTGAAATATCAAAAATA
AAATTTGATGTGGACATGAACATAATTTCTACAAAAATACTGCAGCTCGACTTCAGTAC
TTTAACCACAATGATCAATTTTACAACCCCGATTTCAGACCTCATATAAGAACCAACAGA
AAAAATCAGAATCCACCAATGACACCGATTTCAGAAAGTTCTATGTCTAGATGCAAATCT
CAATGTCGAAATTCACCAGATTTCACCGAGAAATCTAAGAAGAAAAAACACAAATCTGGA
TCATCAAGTATCTCTTCCAGTATTGAAGAAAATTCAGATTTGATTACATACAGCAACC
GAAAAGATAAGTTTACAACATTTAACTATCAATCTTCCAGAACTAGATCATCATCAACA
TCATCATCAGCTTCTTTATCATGTTCCAAATCCACATTAATAAATATGGCGAAAAGTCAA
TATAAAGACAATGAACAAATAAAATCTAAATCAGATTCAAAACACAAAACCTACAAACATG
TCTTCAGATTACGAATCGAATAGACATGCAGATGTGTTTAGAAACTCCCGGAAGCAGGT
GAAAAATTTCCCTACACAATTTCTAGTCCATTTAACACACATGAACAATCCAATCATTC
GAAAACGCCATTGATGAAGAACAAAAAAGGCACCTAATGTACGACTTCTCACCTACAG
GAAGACCAAAGTCTCATATCACACAACACAAAAAATGCAAAAAAAGCGCCCCGCGAT
GATGATTTCAGATTCCAGTATAAACAATTTCTGCAAGAAAAAATATCCGGAGCGCAAAA
GCAGATTCTGAAATAAGTGAACCTGAAGATTTGTCTTATAGAGATTATGTCAGACTAAAA
GAAAGAAAAGTATCGGAAAAATTTAAATTCACCGCGGAAGGGTGGCTACAAAAGATTTT
CAAAACCTCTTTAGAAATACTATAAGAGCATTTGAATACAAACAGATCCCCAAAAACCT
TGCAATGAAAAACTCTAAAAGAAGCTATATACAGCATCTGCTGTAATGGGTACTCTAAC

AACGCAGCCATCATTATGTACTTTACAAGATCAAAAAAGTAGCTCAAATCATAAAGATA
ATGCAAAAAGAATTAATGATTTCGACCCAACATCACTGTTTTAGAGCCATTTAAAATGAAT
CATGCACCGCCTAAGTATTATGACAAAAGATGAAATTAACGATTTATACAACACAAAA
CAAGGACCGCAAGAATTATGGGATAAATTTGAAAACAACACAACACATGATCTATTTACG
AGACATTCGGACGTAACAACTATGATAATTTACGCCGCTACGCCCATCGACTTTGTAGGA
GCTGTTAAAACATGCAATAAGTATGCCAAAAGATAACCCTAAAGAAATTTGCTCTTCGGGTG
TGTTCAATCATCGACGGTGACAATCCTATCTCTATCTATAATCCTATCAGCAAAGATTTT
AAATCGAAATTTAGTACACTTTCAAATGTTAA

>5814_U86

ATGGAGCCAGCAAAACCCCTCTGGAACAACATGGGATCCAATGATGAACGTATGCAAGAG
TACCGTCCCGATCCAATGATGGAAGAATCTATCAAACAAATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCACATCAAGACTTG
AAAACTTTAGCTGAACATAATTCCAAACAAGGAAATGATCATAACGACGTCATCCAAAGC
GCCATGAAATTATCCGGACTATACTGTGATGCGGACATCACACACACAACCGTTGAGT
GATAATACGCATCAAGATCCTATATATTCTCAAGATTCCGGAATATTTTCAAACCTATA
CGGGATCCACGGTTAGCAGCTCAGTTCACAGACAATGCACATCCTCTGCTTCCAATCTA
CCGTCAAACGAAAGCGGTTTCGACCCAGGTGCGCTTTGCTTCGGAACCTACCAAGTCAATTC
GTACAGCCCAAGTACACGTACATAACCAAAATGCAAACTACAGAACAATTTACCTCG
TTGCCGTATCAACCTAATCATGATCCATACAAAGGTATAGAGAATTCCTACAGAGAATCT
AGAAATACCTATACAAGATATGACTCTAACTTTAGACATCATTCCAATAGGCCCAGAGGA
GAAATGAAAATATAATTACTACAATCCAAACTCAAATATCAACAGCCATACAAAAGA
TGTTTTACCAGAACTTATAACAGACGCGGAAGAGGCTACAGATCATACGATTTCTCAGAC
AGATCCGCGGATTTTTCATATGATCAATACATGTATCCAACCTACGAACAACATAATCCA
GACCCCGTATGAATAACTGTAAAGATTTTACACAACCTAACAAACAAGTTCAATTTTGAA
GACTATGACTATAGCATGGCATTTCACAGACTCCGCACACGTACAATCTGATAATTAC
AATCATCCAACAAAAGCACAGACAATACCAGAACTACAAAACCTAAAAACATGAGGCT
ACTAAGGATAACGAAACATCCACAGAAAATCAAATTTTAAACGTCTCCTGACGCAATCAGT
TTATCATAACAGACCTAGTCCAAAAAATGGATATTATTAATAAATAATATATGACACTGAC
GTTATTCATTACCGAAGGAAGCGTTAACAGCTAATGGCTCAAATCGTGACGTGGATATT
CAGAAGTACAAAAAGCACATATCCGTTGTAGATCAGTCCAGAAGAAAAAAGAACGCAGT
TCTCAAACAAACAAACATGATGAAAACCTTGCCCTCATCACGTTCTGATCTTAAAGAGATA
AAATCTAATGAGCATGAAGATAAAGCCGTTACTAAAGCTAGAGATTTTTCAAATTAGAT
CCCCTATTATCTCCTCTTCCGTTGACACCTGAACCCGCATAGACTTTGCAGACCATACA
GATAAATCTGTTCGACACCAGAATTC AACCAAAGAAAACAAAATCTACAGCGATCAAAA
ACTTCTCTACAAGACACAGTTCCCATATCAAACATACTCCAAGAGCCCCGACCAAAGAT
AATCTTACAAAAACATCATGACTCCAAAGACAATTATCCTAAAATGAAACATTCTCCT
GGACGTACAACATCTAAAAAATACTACAAACTCCAACGGACACCAAACCTTTAAGGAA
GTATCCGTTAAGAATGTTTCAGGGAAAGCCACATCTACACCACCAAATCAAACACAC
CATTATCCCAGTAGCTCTGATGAAGAAGGTCAATACAAATCCCGGTAAAAACTATTATC
CCATCACCTTCTCCATATTCCAAGTTAAAAATCCTTCAATTATGGACAAAAACAGTGCT
AAAAACCACACTGCAAGCGCCGACAAAAATTTAACAGACAATTCACCTATTTCGGAGTAAT
TTAAATCCTACTGCTTTTAAACAATCTAACAGTAACAAAAGCATCACGAATAGCACTTCT
AACAGTGATGAATGTACTGACAAGAAACCTAACTGCGATTCCACGAAAAATGAATCTAAA
GATACTAACCGAATTTGTAGCAAAAACCTCTGATAAACATCTGAGTAAAAGTTGCACCATG
GCTTCTAAAAGGGCTCCAGCAGAGCCTCAAGCAGAGCCTCCAGTAGAGCCTCCAGCAGG
GATTC TAGCAGGGCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGG
GCCTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGAGACTCCAGCAGAGCCTCCAGCAGG
GATTC TAGCAGGGCTCAAGCAAAGCCTCCAGCAGAGCCTCCAGCAGAGCCTCCAGCAGG
GATTC TAGCAGGGATTCAAGCAGGGCTCAAGCAAAGCCTCCAGCAGGGATTCTAGCAGG
GCCTCAAGCAGGGCTCAAGCAAAGCCTCAAGTAAAGCCTCCAGCAGAGCCTCCAGTAGG
AATTC TAGCAGGGCTCAAGCAGGGCTCAAGCAAAGCCTCCAGCAGGGATTCTAGCAGA
GCCTCCAGCAGAGCCTCCAGCAGGGATTCTAGCAGGGCTCAAGCAAAGCCTCAAGTAAA
GCCTCCAGCAACACTTCCAGCAGGAAGCTGAAGCAGATTTATCACGATTCTAATAAATA
CAAACACTCCACATGACACCTCAATTAATACAGAATCTGAAATATCAAATAAATTT

CGATGTGGACATGAACTAAATTTCTACAAAAATACTGCAGCTCGACTTCACTACTTTAAC
CACAATGATCAATTTTACAACCCCGATTTCAGACCTCATATAAGAACCAACAGAAAAAAA
TCAGAATCCACCAATGACACCGATTTCAGAAAAGTTCTATGTCTAGATGCAAATCTCAATGT
CGAAATTCACCAGATTTCACCGAGAAATCTAAGAAGAAAAAACACAAATCTGGATCATCA
AGTATCTCTTCCAGTATTGAAGAAAATTCAGATTTGATTCACATACAGCAACCCGAAAA
GATAAGTTTACAACATTTAACTATCAATCTTCCAGAAGTAGATCATCATCAACATCATCA
TCAGCTTCTTTATCATGTTCCAAATCCACATTAATAATATGGCGCAAAAAGTCAATATAAA
GACAATGAACAAATAAAATCTAAATCAGATTCAAAACACAAAACCTACAAACATGTCTTCA
GATTACGAATCGAATAGACGTGCAGATGTGTTTAGAAAACCTCCCGGAAGCAGGTGAAAAA
TTTCCCCTACACAATTTCTAGTCCATTTAACACACATGAACAATCCAATCATTCCGAAAAAC
GCCATTGATGAAGAACAAAAAAGGCACCTAATATCACGACTTCTCACCTACGGGAAGAC
CAAAGTCTCATATCACACAACACAAAAAATGCAAAAAAAGCGCCCCGCGATGATGAT
TCAGATTCCAGTATAAAAAATTTCTGCAAGAAAAAATATCCGGAGCGCAAAAAGCAGAT
TCTGAAATAAGTGAACCTGAAGATTTGTCTTATAGAGATTATGTCAGACTAAAAGAAAA
AAAGTATCGGAAAAATTTAAATTCACCGCGGAAGGGTGGCTACAAAAGATTTTCAAAAA
CTCTTTAGAAATACTATAAGAGCATTTGAATACAAACAGATCCCCAAAAACCTTGCAAT
GAAAAACTCTAAAAGAAGCTGTATACAGCATCTGCTGTAATGGGTACTCTAACAACGCA
GCCATCATTATGTACTTTACAAGATCAAAAAAAGTAGCTCAAATCATAAAGATAATGCAA
AAAGAATTAATGATTGACCCAACATCACTGTTTTAGAGCCATTTAAAATGAATCATGCA
CCGCCTAAGTATTATGACAAAGATGAAATTAACGATTTATACAACCTACAAAAACAAGGA
CCGCAAGAATTATGGGATAAATTTGAAAACAACACAACACATGATCTATTTACGAGACAT
TCGGACGTAAAAACTATGATAATTTACGCCGCTACGCCCATCGACTTTGTAGGAGCTGTT
AAAACATGCAATAAGTATGCCAAAGATAACCCTAAAGAAATTTGCTCTTCGGGTGTGTTCA
ATCATCGACGGTGACAATCCTATCTCTATCTATAATCCTATCAGCAAAGATTTTAAATCG
AAATTTAGTACACTTTCAAAATGTTAA

U90

>2284 U90

ATGGAGCCAGCAAACCCTCTGGAAACAACATGGGATCCAATGATGAACGTATGCAAGAG
TACCGTCCCGATCCAATGATGGAAGAATCTATCCAACAGATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCAAATGAAGACTTG
AAAACCTTAGCTGCTCAGTCCGCTGGGAATTGCATCCAAAGTATTGGTGCATCAGTGAAG
GCTGCCATGAAACAGGAACAGTCTGACATGGAAGACAAATTAATTAAGTGCCTGGCTTA
TTAACGCAGCAACAATCCATGTTTCATAGGACTAGGATTAGAACAATTATCACAGCTGATA
AACATTAATTTATTAAGTAGTGTCTCAACTAAATATGTTGAATCTTATTCAAAAATGTTG
CATGGAAAAGAGTTGGATTTTTTTAATTTGGTGCGAACCTAGATTCATTGTTTTTGCATGT
GATAAATTCGATGGGTTAGTAAAAAGGTGGCATCAGAATCCAGAAACTTGCTCCTGGAT
TTGAGAGCTAATATGAACAATGACATTTAATTAAGGCTGTCAAAGACATTTTTTCTAAAGCT
ACTGTGACATTAGATTGTGCAAACTTAACCAGGGCGCCACCATGTTGATGATGATGGCA
CACAACAAAGAGATGTCCAATCCTGACATCTCCAGTAAGGACTTTTGTGAAAAAATAAAT
GTACTTAAACAGACATTACTGGAAGGTAAAAATGAGATAGTAGAAAACAAATGCTAAAAAT
ATGCAAATTTTACAGACATTTGCTATAAAGCAGATGAATCAAATCTTCATGGATGGTTGT
GATAAAGCATTTCTTAAACTTAATGTGAATTTGAAAAATTTAATTAAGTGCCTGCAAAAAAT
CTTGCCAATGCTATTCTCCAGAGTATTATTATATGCAGCAATGAATTTTCATGGCAACAC
CTCAAACCTCTGAGACAGGGATTTAAAGCTACCATACTGACTATGATTACCCAAGCATGT
GAACGCTTAGAATCAAGTTATGATGATACTGGTTTTAATTAACCTTTAGCTCCCCTTCAA
ATAATGGAAGGCTACATTAATATGGATAAAAAATAGACAGTCTTCCATTTGTGATGGAAAC
ACAGACCCATCAGACTCAATGATATTAGACTTAGCTGATTTTGTGATGACCATGGAAGATAT
TCAGAGGAATCTTCTATCGAATCTATCCATGAAGATGATGACAATAAAATGTATCCATGC
ACTCCTTCTCCAGAAGTGCCAGGGAAATCCAAATATGTAGGAACATTCAGTAAAAACAGC
AGACAGTCAGGGGACGAACAAACAAACCCTAACTGTGTAGGTACAGCCTCAGTGACAGAT
CTGGGCGGCCCTTGATAATTTGAACTCCATCAGCGGCCTCCAGAGTTGTAATAAACATGTTG
CTAGAAAGACTACTTTGACACTCAATGTGATTTCAGTAGTTGAAGGGACAGAACAACATGGT
TTCCATGACAACCTTTAATATCAGAAATGCTTATGTTTGGTTATGAAACTGATCACTCG
GCACCCTATGAGAGTGAAGTGATAACAATGATGAAATGATTACATTGCAAACCTCTGAT

TCAGCTGCGAGAAC TAATAACATTCATATGAATAGCACTGATGAGAATACACCATTTAGC
AATTCCTATATGTTCCACCACCAGAGGTCACCCCTTCTAAGAAAAATGTCAAACCTAAGTCT
ATGACTCAGGGAAGCAAACCTAAAAAGAGGGTAGCTAAAAGGAAACATGTTTTCAAGTAAA
TCACCCAAACTTAAAAAGATTAAGTCAGATCAGTTACCTAAGGCTGCAGACGTAATTGTT
ATAAGTTCAGAGAGTGAAGATGAAGAGGATGGTGACAATATTATTGGTAACTCAATCTTA
ATAAAAGCTATTAATCAGAATCTGATTCTGAATCGAGCAGCGAGTCAGATGACTGTACA
ACTGAAGATAAACAGTTACATTTAAGTGATTATGATAAAGTTACCAATAATGGTGATTGT
CCAAGTAAAGGTTTTCCAACCTCCAGTGTTTTACCATCCCAATAAGGTCAATGCAGGGTATC
CATGATATCAGATCCAAATTTATTCCCAAAAAAATTGGATTTGGTATATGAAGAAAACT
CACCAAGTAGACAATTGTCTTATTCACAATTCTGAAAACATAAATACAAAAGATGACAGT
GATGGGACAGAAGCCAAACACTGCTTTTATGAATCACTTTGTCCCCATCAAACCGGATGAT
GAAGATTATGATAAGGACAATGTCTCATATATTTATAATAAAATACAGAATAGCAAAAATA
GACTCGGGAGTGATAACACCAACTAAGAGACTTATCACAGATATGATGATGGATAATTTT
ATGGATCTAAATGATATCATTAACAGGGCATAACCAAACATTGCCAGGATCTTTGCAAT
AAATATAATGTTGTCAACCCCTACTACATGTGAGAAAGATTTAAATGTGACAAATTTCCAA
ACTTTTGCAACAACCGCAACTCAAGTGTTTGACCCACCGGTAACAGGAAATAATTCCTCC
ATACTAAATATTATCAATGACGCAACAAGTCAAAATGATGAGAACAGATGCACTGAAGGT
ACAAGTAATTC TAATGAAAAATGTACCAATATAAGTGATTGCAACAGTGATGGTACAGAG
GCCTTTAAATTAGATGGCTACCCCTCTGATTATGATCCTTTTCGTAGAAAATGCTCAAATC
TATTAG

>5055_u90

ATGGAGCCAGCAAACCCCTCTGGAAACAACATGGGATCCAATGATGAACGTATGCAAGAC
TACCGTCCCGATCCAATGATGGAAGAATCTATCAAACAAATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCACATCAAGACTTG
AAAACTTTAGCTGCTCAGTCCGCTGGGAATTGCATCCAAAGTATTGGTGCATCAGTGAAG
GCTGCCATGAAACAGGAACAGTCTGACATGGAAGACAAATTAATTAAGTGCGCTGGCTTA
TTAACGCAGCAACAATCCATGTTTCATAGGACTAGGATTAGAACAATTATCACAGTTGATA
AACATTAATTTATTAAGTAGTGCTTCAACTAAATATGTTGAATCTTATTCAAAAATGTTG
CATGGAAAAGAGTTGGATTTTTTTTAATTTGGTGCGAACCTAGATTCATTGTTTTTGCATGT
GATAAATTCGATGGGTTAGTGAAAAAGGTGGCATTAGAATCCAGAACTTGCTCCTGGAT
TTGAGAGCTAATATGAACAATGACATTATTAAGGCCGTCAAAGACATTTTTTTCTAAAGCT
ACTGTGACATTAGATTGTGAGAACTTAACCAGGGCGCCACCATGTTGATGATGATGGCA
CACAACAAAGAGATGTCCAATCCTGACATCTCCAGTAAGGACTTTTTGTGAAAAAATAAAT
GCACTTAAACAGACATTACTGGAAGGTAAAAATGAGATAGTAGAAACAAATGGTAAAAAT
ATGCAAAATTTTACAGACATTTGCTATAAAGCAGATGAATCAAATCTTCATGGATGGTTGT
GATAAAGCATTTCCTTAAACTTAATGTGAATTTGTAATAATTTAATTAATGCTGCGCCAAAAAT
CTTGCCAATGCTATTTTCCAGAGTATTGTTATATGCAGCAATGAATTTTCATGGCAACAC
CTTAAACTTCTGAGACAGGGATTTAAAGCTACCATGCTGAATATGATTACCCAAACATGT
GAACACTTAGAATCAAGTTATGATGATACTGGTTTTAATTAACCTTTAGCTCCCTTCAA
ATAATGGAAGGCTACATTAATATGAATAAAAAATAGACAGTCTTCCATTTGTGATGGAAC
ACAGACCCATCAGACTCAATGATATTAGACTTAGCTGATTTTGTGACCATGGAAGATAT
TCAGAGGAATCTTCTATCGAATCTATCCATGAAGATGATGACAATAAAATGTATCCATAC
ACTCTTTCTCCAGAAGTGCCAGGGAAATCCAAATATGTAGGAACATTCCTGAAAACAGC
AGACAGTCAGGGGACGAAAAACAAACCCCTAACTGTATAGGTACAGCCTCAGTGACAGAT
TTGGGCGGCCCTGATAACTTGAACCTCATCAGCGGCCTCCAGAGTTGTAAAAACATGTTG
CTACAAAGACTACTTGACACTCAATGTGATTGAGTAGTTGAAGGGACAGAACAAGATGGT
TTACATGACAACCCTTTAAATATCAGAAATGTTAATGTTTGGTTATGAAACTGATCACTCG
GCACCCATGAGAGTGAAAGCGATAACAATGATGAAATTTGATTACATTGCAAACCTCTGAT
TCAGCTGCGAGAAC TAATAACATTCATATGAATAGCACTGATGAGAATACACCATTTAGC
AATTCCTATATGTTCCACCACCAGAGGTCACCCCTTCTAAGAAAAATGTCAAACCTAAGTCT
ATGACTCCGGGAAGCAAACCTAAAAAGAGGGTAGCTAAAAGGAAACATGTTTTCAAGTAAA
TCACCCAAACATAAAAAAGATTAAGACAGATCAGTTACCTAAGGCTGCAGACGTAATTGTA
ATAAGTTCAGAGAGTGAAGATGAAGAGGATGGTGACAATATTATTGGTAACTCAATCTTA
ATAAAAGCTATTAATCAGAATTAGATTCTGAATCGAGCAGCGAGTCAGATGACTGTACA
TCTGATCATAAACAGTTACATTTGAGTGATTATGATAAAGTTACCAATAATGGTCATTGT
CCAAGTAATGGTTTTCCAACCTCCAGTGTTTTACCATCCCAATAAGGTCAATGCAGGGTACC

CATGGTATCAAATCCAAATTTATTCCCAAAAAAATTGGATTTGGTATATGAAGAAAAC
CACCAAGTAGACAATTGTCCATTTCACAGTTCTGAAGACGTGAATGTAAAAGATGACAGT
GATGGGACAGAAGCCAAACACTGCTTTATAAATCACTTTGTCCCCATCAAACAAATGAT
GAAGACTATGATAAGGACAATGTCTCATATATTTATAATAAAATACAGAATAGCAAAAATA
GACTCGGGAGTCATAACACCAACTAAGAGACTTATCATAGATATGGTGATGGATAATTTT
ATGGATCTAAATGATATCATTTAAACAGGGCATAACCAAACATTGCCAGGATCTTTGCAAT
AAATATAATGTTGTCAACCCCTACTACATGTGAGAAAAGATTTAAATGTGGCAAATTTCCAA
ACTTTTGCAACAACCGCAACTCAAGTGTTTTGACCCACCGGTAACAGGAAATAATTCCTCC
ATACTAAATATTATCAATGACACAACAAGTCAAAATGATGAGAACAGATGCACTGAAGGT
GCAAGTAATGCTAATGAAAAATATACCAATATAAGTGATTGCAACAGTGATGGTACAGAG
GCCTTTAAATTAGATGGCTACCCCTCTGATTATGATCCGTTTCGTAGAAAATGCTCAAATC
TATTAG

>5814_U90

ATGGAGCCAGCAAACCCCTCTGGAAACAACATGGGATCCAATGATGAACGTATGCAAGAG
TACCGTCCCGATCCAATGATGGAAGAATCTATCAAACAAATATTGGAAGAGAGTCTCATG
TGTGATACATCCTTCGATGACCTGATTCTTCCAGGCCTTGAAAGCTTTGGTCTTATTATA
CCAGAGTCTTCCAACAACATAGAGTCCAATAATGTAGAAGAAGGATCACATCAAGACTTG
AAAACTTTAGCTGCTCAGTCCGCTGGGAATTGCATCCAAAGTATTGGTGCATCAGTGAAG
GCTGCCATGAAACAGGAACAGTCTGACATGGAAGACAAATTAATTAAGTGCCTGGCTTA
TTAACGCAGCAACAATCCATGTTTCATAGGACTAGGATTAGAACAATTATCACAGTTGATA
AACATTAATTTATTAAGTAGTGCTTCAACTAAATATGTTGAATCTTATTCAAAAATGTTG
CATGGAAAAGAGTTGGATTTTTTTAATTGGTGCGAACCTAGATTCATTGTTTTTGCATGT
GATAAATTCGATGGGTTAGTAAAAAGGTGGCATTAGAATCCAGAACTTGCTCCTGGAT
TTGAGAGCTAATATGAACAATGACATTATTAAGGCCGTCAAAGACATTTTTTCTAAAGCT
ACTGTGACATTAGATTGTGAGAACTTAACCAGGGCGCCACCATGTTGATGATGATGGCA
CACAACAAAGAGATGTCCAATCCTGACATCTCCAGTAAGGACTTTTGTGAAAAAATAAAT
GCACTTAAACAGACATTACTGGAAGGTAAAAATGAGATAGTAGAAAACAAATACTAAAAAT
ATGCAAATTTTACAGACATTTGCTATAAAGCAGATGAATCAAATCTTCATGGATGGTTGT
GATAAAGCATTCTTTAAACTTAATGTGAATTTGTAATAAATTTAATTAAGTGCCTGGCTTA
CTTGCCAATGCTATTCTCCAGAGTATTATTATATGCAGCAATGAATTTTCATGGCAACAC
CTTAAACTTCTGAGACAGGGATTTAAAGCTACCGTGCTGAATATGATTACCCAGGCATGT
GAACGCTTAGAATCAAGTTATGATGATACTGGTTTTAATTAACCTTTAATTCCTTCAA
ATAATGGAAGGCCACATTAATATGAATAAAAAATAGACAGTCTTCCATTTGTGATGGAAC
ACAGACCCATCAGACTCAATGATATTAGATTTAGCTGAATTTGATGATCATGGAAGATAT
TCAGAGGAATCTTCTATCGAATCTATCCATGAAGATGATGACAATAAAATGTATCCATGC
ACTCCTTCTCCAGAAGTGCCAGGGAAATCCAAATATGTAGGAACATTCCTGAAAACAGC
AGACAGTCAGGGGACGAACAAACAAACCCCTAACTGTGTAGGTACAGCCTCAGTGACAGAT
CTGGGCGGCCCTTGATAATTTGAACTCCATCAGCGGCCTCCAGAGTTGTAAAAACATGTTG
CTAGAAAAGACTACTTGACACTCAATGTGATTTCAGTAGTTGAAGGGACAGAACAACATGGT
TTCCATGACAACCTTTAATATCAGAAATGCTTATGTTTGGTTATGAAACTGATCACTCG
GCACCCATGAGAGTGAAAGTGATAACAATGATGAAATTGATTACATTGCAAACCTTGAT
TCAGCTGCGAGAACTAATAACATTCATATGAATAGCACTGATGAGAATACACCATTTAGC
AATTCATATATGTTACCACCAGAGGTCAACCCCTTCTAAGAAAAATGTCAAACCTAAGTCT
ATGACTCAGGGAAGCAAACCTAAAAAGAGGGTAGCTAAAAGGAAACATGTTTCAAGTAAA
TCACCCAAACTTAAAAAGATTAAGTCAGATCAGTTACCTAAGGCTGCAGACGTAATTGTT
ATAAGTTCAGAGAGTGAAGATGAAGAGGATGGTGACAATATTATTGGTAACCTCAATCTTA
ATAAAAGCTATTAAATCAGAATCTGATTCGAATCGAGCAGCGAGTCAGATGACTGTACA
ACTGAAGATAAACAGTTACATTTAAGTGATTATGATAAAGTTACCAATAATGGTGATTGT
CCAAGTAAAGTTTTTCCAACCTCAGTGTTTTACCATCCCAATAAGGTCAATGCAGGGTATC
CATGATATCAGATCCAAATTTATTCCCAAAAAAATTGGATTTGGTATATGAAGAAAAC
CACCAAGTAGACAATTGTCCATTTCACAATTCGAAAACATAAATACAAAAGATGACAGT
GATGGGACAGAAGCCAAACACTGCTTTATGAATCACTTTGTCCCCATCAAACCGGATGAT
GAAGATTATGATAAGGACAATGTCTCATATATTTATAATAAAATACAGAATAGCAAAAATA
GACTCGGGAGTGATAACACCAACTAAGAGACTTATCACAGATATGATGATGGATAATTTT
ATGGATCTAAATGATATCATTTAAACAGGGCATAACCAAACATTGCCAGGATCTTTGCAAT
AAATATAATGTTGTCAACCCCTACTACATGTGAGAAAAGATTTAAATGTGACAAATTTCCAA
ACTTTTGCAACAACCGCAACTCAAGTGTTTTGACCCACCGGTAACAGGAAATAATTCCTCC

ATACTAAATATTATCAATGACGCAACAAGTCAAAATGATGAGAACAGATGCACTGAAGGT
ACAAGTAATTTCTAATGAAAAATGTACCAATATAAGTGATTGCAACAGTGATGGTACAGAG
GCCTTTAAATTAGATGGCTACCCCTTCTGATTATGATCCTTTTCGTAGAAAATGCTCAAATC
TATTA

U91

>2284_U91

ATGGGGAAAAATGTCATCTACTGGAAATGGAAAACTAATCTAAAGATACTGGCATGTATG
CTATTGATATTTCTAATAGCAACAATATCTTTGTTAATACTGGAAATTATATCGGGCCAA
AGATATTCCAATGATGATTCCGAAGGTGTGACTGCGGCGCTGAAGCATGTAAGCACACCT
ACTACCAACTGCACTGAAACCCTACTCCTGATTTTCGTAACGAGCCAAGCCACGGGAAAC
AAAGAATCAATGAAAAAAAATGAAGGGGAACCTCCAGTTTGGATTTCAGGCTTTAACTACA
ACATTATCAATTATCCTATTGGTTTGCATAATAATGGCATGCATTATATGTTCCCGGACC
ACGGAAGAAGAGAAGAGCGGAATGCAATCTTCTGCTTCTTCAGTAGAGACTTTACAGTCT
TTAAATGAGGCTATTTTACCGAAAGGTGAAATGAATGTGTAA

>5055_U91

ATGGGGAAAAATGTCATCTACTGGAAATGGAAAACTAATCTAAAGATACTGGCATGTCTG
CTATTGATATTTCTAATAGTAACAATATTTTTGTTAATACTGGAAATTATATCGGGCCAA
AGATATTCCAATGATGATTCCGAAGGTGTGACTGCGGCGCTGAAGCATGTAAGCACACCT
ACTACCAACTGCACTGAAACCCTACTCCTGATTTTCGTAACGAGCCAAGCCACGGGAAAC
AAAGAATCAATGAAAAAAAATGAAGGGGAACCTCCAGTTTGGATTTCAGGCTTTAACTACA
ACATTATCAATTATCCTATTGGTTTGCATAATAATGGCATGCATTATATGTTCCCGGACC
ACGGAAGAAGAGAAGAGCGGAATGCAATCTTCTGCTTCTTCAGTAGAGACTTTACAGTCT
TTAAATGAGGCTATTTTACCGAAAGGTGAAATGAATGTGTAA

>5814_U91

ATGGGGAAAAATGTCATCTACTGGAACCTGGAACCTGGAAAAACTAATCTAAAGATACTGGCA
TGTCTGCTATTGATATTTCTAATAGCAACAATATTTTTGTTAATACTGGAAATTATACCG
AGCCAAAGATATTCCAATGATGATTCCGAAGGTGTGACTGCGGCGCTGAAGCATGTAAGC
ACACCTACTACCAACTGCACTGAAACCCTACTCCTGATTTCCGTAACGAGCCAAGCCACG
GAAAACAAAGAATCAATGAAAAAAAATGAAGGGGAACCTCCAGTTTGGATTTCAGGCTTTA
ACTACAACATTATCAATTACCCATTGGTTTGCATAATAATGGCATGCATTATATGTTCC
CGGACCACGGAAGAAGAGAAGAGCGGAATGCAATCTTCTGCTTCTTCAGTAGAGACTTTA
CAGTCTTTAAATGAGGCTATTTTACCGAAAGGTGAAATGAATGTGTAA

U95

>2284_U95

ATGTCGCAAATCCGGACGATTTCTTTTGGCAACAAATTCTCAATATGGATCCGACGGAA
CTTTTGTGCGGACGATGCAATCTCGTCCACGTGTGACGAAAACACTGCAGCTGGAACCTGAT
ATTATGCAGCCTTTGTGCGTTGATATGTCTGTCCAACCGACGACATCGGGGTGTTATACC
GGAACGATGATGTCAGACTCGTATTCCCTCCAATGACACACCGGATCAAAACTGCCAATCA
CATCCGACTTTGACGGACCTTTGCGCCGGAAGACATGAACAATCAAATCAATGTTCAACCC
ACTCAGATGACCTTTCAACCCATTTCCCCACCGATACAGGATCAAAATTATGGCTATAGC
AATAATATGATTAATCCCATCAAACCTGCATCAATTATCAAATTACATGGTCATTTCGATA
GGAGAAATGACATCTTCAGACCCCTTCTCTCATGTTAATGCTCAACACTTTGTCCAACAG
TCTAATCCTAAATCTTTCTAAATATGCATCCGTGCATGACCGCTAACTCACAAGAACCC
GGGGAATGTCATTCTGTCCATAATCAATCTTCTGGAAGCGGGTCTAATTCTTATGACATT
ACTACGGCACCCCGGATGAATGGATATTGACTACATCCGGCGGAGAATCCTGGATTCTG
AAACGGAACACCCCTAATCCTCCAATAGCAGAATAATTTAATTTTCAACAACGCACAA
CCGATTTTCCATATCCAACCTCAAGTATCTGGGTTTAGCAACTGTTCCCTATCAGGGGGT
GGTCTACAGTCACGTGAGTATGTAATTTGGCAGGTTACACTCCCGTGAGCAACATGCAG
AATATGAAAACTCATGGGCGACACAAC TAGGAGCATCCATTGACTGCATGACTGCGTCA
CCAAC TGGTATAGATGCGAATGTCAACTCTGCTTTCCAGCTGGTGGGGGTGGCATCAAAT
AATTGGGGTGCTAATGTTTCAATGGGAGGCATGGGTAATTATATAGAGAATAACAGTCCA

TGGAACCAATACTATAAAATCGGAAATGGCTAATAATTCATAAAATGTCAACGAGTCGAAT
GGTTATGGCCATGTGGGGTGTGTTAATGACGCCATTTCCGATAAAACAAGGTGGCACCGCT
AACGTCCCATCTTCTCTATTGAATCCGGATCACCAAGATTGGATGAGGATGACTGGTACG
AGCACACATATGTTAAACAATATAAAATGCTAAAACAGAAATGGATTTTCCGGAAAATGAT
GGCAATGTCCACGGAGCCACCGGTGTAGATTTATCATCAACCCTGTCCAATGGTCAGCAG
CATAGTGCATTTGAAGGAAATGGAACATTTCCCGTGGTAAATACCGGCAATCCAAGAACG
ACGTTGAACTGTGCTCCTCAGTTGTTGTGGCTTGGAGGAAACCCGACAATTGCACATCCA
ATGAATGGGAATAATGAAACCCAGATAAAAAATGTATGCAAGCCTACCCCCAGCACAGGT
GTAATAAAAAAATTAACCTTTGACTGTGTTGACCGAGGGGAAGGAATAGGAAATTTTCCC
AACTGGCGTCTTTTGTAGCTATGAGAGAAAAATGTCAAACACGGTCAATCCTTCTTGT
GGAACTTTTCTAGCACAGTTGGAGGGATTACGCCAGCAGAATTCATTTGAAAAAATATA
TCCCTTACTGATGTACAGTTTTCTGAAGAAGATGATGTGCTTTCCAGTGCATCCTCGGTC
AGCTGTAATAGCAACTGCGACATGAAAATCGGAGTTTTCGCAGCAAGGTACCACTGTGTCT
AATTTACAACAAGGGTTTAGGCAACAAATGAACATGAACTCATCCGTGGTTAAAAATGGAC
GGGAGCTATAAAACACAAGAGATGTCAAATGACTGCGCGACAGACGTACCGGATAATGCA
GGCATCATTCAGAAAAATAACGTATATACTCTCAAATTGAGAATCCAGAGGATGGCCGG
GTACGGGAGGGGGGGTGTTCAGACGACAAAAATCAACCTCCGCGGAAAAGTGCAGCTATT
CATAACATGAAGTCCGAAGGCGTGACTTGCAGTATGTGCATGACGGCTGCGGATTCCACA
CGGCAGGATGCGTCTGGAGGTTCAAGCAGCGGAACCAAGAATGAATATTACGATGATGAA
TCGGAAATTAACCGGACTGTGCGATACAGACAGCGACAACGAAGTCCAATGCTGTCAAGAA
GTAAAAAAGATCGGTCCCAAGACGTACTCCTCAGAAAATTTTAATCCTGAGTATCGCCAA
GCCAAGAGATTGCTTGCCGACATTCATACAGAAGATGGATTCCAGACCCAGTGCATATG
GAAGAAAACGACGGTCCGTTCAATCCAATAGTGACTAGACCAGCCACTGTGTGTATGAAC
GGTAGACGCCGAAGGACTTTCTTCCGGCAATGTGTTACCGTTTTAGTCTCTGAGTAAA
TTAACTTACTTCAAAGAATTACTGCAGAGCTATGTGTCTAAAAATAACAACAGTTATTTA
AGTCTCAGTTGGCCTCCCAAACACGGGGTGTATGTGCATGTGAGAGAAGAAATTAGGGTAT
GAACATATACCTACCTTGAAGGACAAGTTTTCCACTGCCATGTGGGTGGATGATGGTCTTT
GGAGTTGTGCGGTGACAGAACCGCCCGCGGCTCTGAATAAAAAACATGGTGATTCTCCTGTGT
GAAAATAGATGGGTGCTTCTGCATAATTAATCAAACCTCTACACACGAGCTGTTTTTAGCC
GCCTCGGATTTGAAACAGTTTATGGAGGAGGGTCTGTCTAGATGTGATCCTATATATGAA
CAAGTAACAGTGCCTGACGGTGTGGCCATGGAGAATCTCTTCGAAACTTCTTGAGAAAC
AGTAGAACGTTTCAGGATCTTATGGATCAAAGGGACAACATGCACGGATGTAACGGACC
TTTAACGGAATGCCGTGGTAGACTAGGTGACAGAGTCATACATCTCTGCAATCCTGATTCA
GTTGATTCCATACCAGCAGACGAGGCCGTGATGTGAAGGAAGACCCTTGTATTTCTTT
GCATATGTTACCATTTCAAATCAAACCCCGCTACTAACGCTACCGTGTTAATAGCCGCG
GATAAAGATTTAAGAATATATGGCTATCATAAAGGTCGCCCCAGGATACGGTACTTGTGT
AAGAACGTACAACTTTCTTTAAAGCCGGGGCTAGAAAAGTTCTATTTAGATTTTCAGATT
ACACCTAAACGCTTGCTTGCCGTCAACGAAGAGGAGTATCTGAATACTCTACAAAACGCT
CCTTGCTGCTCTTAAACCATCTGTGTTTAGAAAAGATTTACTCTCAGGAAGGTAAATAA

>5055_U95

ATGTCGCAAGTCCGGACGATTTCTTTTTGGCAACAAATCCTCAATATGGATCCGACGGAA
CTTTTGTGCGGACGATGCAATCTCGTCCACGTGTGACGAAAACACTGCAGCTGGAAGTATGAT
ATACAGCCTCTGTGCGTTGATATGTCTGTCCAACCGACGACATCGGGGTGTCATACCGGA
GCGATGACGACAGATTTCGTATTTTTCCAATGATACACCGGATCAAAACTGCCAATCACAT
TCGACTTTGATGGACCTTTTCGCTGGAAGACATGAACAATCAAATCAATGTTCAACCCAAT
CAGATGACCTTTCAATCCATTTCCCCACCGATGCAGGGTGAATAATATGTCTATAGCAAT
AATATGATTAATCCCATCAGACCTGCATCAATTATCAAATTACATGGTCATTCGATAGGA
GAAATGACAGTTACAGACCTTCTCTCATGTTAATGCTCAACACTTTGTCCAACAGTCT
AATCCTAAATTTCTGTCTAAATATGCATCCGTGCATGACCGCTAACTCACAAGGACCCGGG
GAATGTCAGTCTGTCCATAATCAATCTTCTGGAAGCGGGTCTAATTTCTTATGACATTAAT
ACGGCACCCCGGATGAATGGATATTGACTACATCCATCGGAGAATCCTGGATTCTGAAA
CGGAACACCCCTAATCCTCAAATAGCAGAATAATTTCAATTTCAACAACGCACAACCG
ATTTTCCATACCCAACCTCAAGTATCTGGGTTTAGCAACTGTTCCCTATCAGGGGGGTGGT
CTACAGTACAGTGTGATCTAAATGTGTGAGGTCTCCCTCCCGTGCACAACATGCAGAAT
ATGAAAAACTCATGGGCGACACAACCTAGGCGCATCCATTAATGCTGACTGCGTCACCA
ACTGGTATAGATACGAATGTCAACTCTGCTTTCCAGCCGGTGGGGTGGTATCAAATAAT
TGGGGTGTAAATGTTTCAATGGGAGGCATGGGTAATTATATAGAGAATAACAGTCCATCG

AACCAATACTATAAATCGGAAATGGCTAATAATTCATAAATGTCAACGAGTCAATGGT
TATGGCCATGTGGGGTGTGTTAATGACGCCATTTCCGATAAAACAAGGTGGCACCGCTAAC
GTCCCATCTTCTCTATTGAATCCGGATCACCAAGATTGGATGAGGATGACTGGTACGAGC
ACACATATGTTAAACAATATAAATGCTATAACAGAAATGGAAAATATGATTTTTCCGGAA
AATGATGGCAATGTCCACGGAGCCACCGGTGTAGATTTATCATCAACCCTGTCCAATGGT
CAGCAGCATAGCGCATTTGAAGGAAATGGAACATTTCCCGTGGTAAATACCGGCAATCCA
AGAACGACGTTGAACTGTGCTCCTCAGTTGTCTTGGCTTGGAGGAAACCCGACAATTGCA
CATCCAATGAATGGGAATAATAAAGCCCTATATAAAAATGTATGCAAGCCTACCCCCAGC
ACAGGTGTAATAAAAAAATTAACATTTGACTGTGTTGACCGAGGGGAAGGAATAGGAAAT
TTTCCCAAACCTGGCGTCTTTTGTAGCTATGAGAGAAAAATGTCAAACAAGGTCAATCCT
TCTTGTGGAACTTTTCTAGCACAGTTGGAGGGATTACGCCAGCAGAATTCATTTGAAAAA
AATATATCCCTTGTGATCTACAGTTTTCTGAAGATGATGATGTGCTTTCCAGTGCATCC
TCGGTCAGCTGTAATAGCAACTGCGACATGAAAATCGGAGTTTCGCAGCAAGGTACCACT
GTGTCTAATTTACAACAAGGGTTTAGGCAACAAATCAACATGAACTCATCCGTGGTTAAA
ATGGACGGGAGCTATAAAACACAAGAGATGTCAAATGACTGCGCGACAGACGTCTCGGAT
AATGCAGGCACCATTCAACAAGTTAAACGTACACACTCTCAAATTGAGAATCCAGAGGAT
GGCCGGGTACGGGAGGAAGAGTGTTCAGACGTCACAGATCCACTTCCGCGAAAAAGTGCG
CGTATTCATAACATGAAGTCCGAAGGCGTGACTTGCGGTATGTGCATGACGGCTGCGGAT
TCCACACGGCAGGATGCGTCTGGAGGTTTCGAGCAGCGGAACCAAGAATGAATATTACGAT
GATGATTCGGAATTAACCGGACTGTTCGGATACAGACAGCGACAACGAAGTCCAATGCTGT
CAAGAAGTAAAAAGATCGGTCCCAAGACGTACTCCTCAGAAAATTTTAATCCTGAGTAT
CGCCAAGCCAAGAGATTGCTTGCAGACATTCATACAGAAGATGGATTCCAGACACAGTC
AATATGGAAGAAAACGACGGTCCGTTCAATCCAATAGTGACTAGACCAGCCACTGTGTGT
ATGAACGGTAGACGCCGAAGGACTTTCTTCCGGCAATGTGTTACCGTTTTAGTCTCTG
AGTAAATTAACTTACTTCAAAGAATTACTGCAGAGCTATGTGTCTAAAAATAACAATAGT
TATTTAAGTATCAGTTGGCCTCCCAAACACGGGGTGTATGTGCATGTGAGAGAAGAAATTA
GGGTATGAACATATACCTACCTTGAAGGACAAGTTTTCCACTGCCATGTGGGTGGATGATG
GTCCTTGGAGTTGTTCGGTGCAGAACCGCCCGGGCTCTGAATAAAAAACATGGTGATTCTC
CTGTGTGAAAAATAAATGGGTGCTTCTGCATAAATTAATCAAACCTCTACACACGAGCTGTTT
TTAGCCGCTCGGATTTGAAACAGTTTATTGAGGAGGGTCTGTCTAGATGTGATCCTATA
TATGAACAAGTAACAGTGCCTTACGGTGTGGCCATGGGAACTCTCTTCGAAACTTCTTG
AGAAACAGTAGAACGTTTCAGGATCTTATGGATCAAAGGGACAACATGCACGGATGTAAC
TGGACCTTTAACGGAATGCCGTGGTAGACTAGGTGACAGAGTCATACATCTCTGCAATCCT
GATTCGGTTGATTCCATACCAGCAGACGAGGCCGTGATGTGAAGGAAGACCCTTGTAT
TTCTTTGCATATGTTACCCTTTCAAATCAAACCCCGCTACTAACGCTACCGTGTTAATA
GCCGCGGATAAAGATTTAAGAATATATGGCTATCATAAAGGTCGCCCCAGGATACGGTAC
TTGTGTAAGAACGTACAACTTTCTTTAAAGCCGGGGCTAGAAAAGTTCTATTTAGATTTT
CAGATTACACCTAAACGCTTGTCTGCGTCAACGAAGAGGAGTATCTGAATACTCTACAA
AACGCTCCTTGTCTGCTCTTAAACCATCTGTGTTTAGAAAAGATTTACTCTCAGGAAGGT
AAATAA

>5814_U95

ATGTCCTCAAATCTGGACGATTCCTTTTGGCAACAAATTCTCAATATGGATCCGACGGAA
CTTTTGTGCGGACGATGCAATCTCGTCCACGTGTGACGAAAACACTGCAGCTGGAAGTATG
TTTATGCAGCCTTCTGGCGTTGATATGTCTGTCCAACCGACGACATCGGGGTGTCATACC
GGAATGATGACGACAGATTTCGAGATTTGCAATGACACACGGAATCAAAACTGCCAATCA
CATTTGACTTTGATGGACCTTTCACTGGAAGGCATGAACAATCAAATCAATGTTCAACCC
ACTCAGATGACCTTTCAACCCATTTCCCCACCGATGCAGGATCAAAATTTATGGCTATAGC
AATAATATGAATAATCCTATCAGACCTGCATCAATTATGAAATTACATGGTCAATTCGATA
GGAAAAATGACATCTTCAGACCTTCCCTCTCATGTTAATGCTCAACACTTTGTCCAACAG
TCTAATCCTAAATCTTTCTAAATATGCATCCGTGCATGACCGCTAACTCACAAGAACCC
GGGGAATGTCATTTTGTCCATAATCAATCTTCTGGAAGCGGGTCTAATTTCTTATGACATT
ACTACGGCACCCGGCGGTGAATGGATATTTACTACACCCGGCGGAGAATCCTGGATTCTG
AAACGGAACACCCCTAATCCTCAAATAGCAGAACTAATTTATTTTTCAACAACGCACAA
CCGATTTTCCATACCCAACCTCAAGTATCTGGGTTTAGCAACTGTTCTTATCAGGGGGGT
GGTCTACAGTCACGTGAGTATGTAAATGTGTGAGGTTCCGCTCCCGTGCACAACATGCAG
AATATAAAAACTCATGGGCGACACAACCTAGGCGCATCCATTAATGATGACTGCGTCA
CCAACGGTGTAGATGCGAATGTCAACTCTGCTTTCCAGCTGGTGGGGGTGGCATCAAAT

AATTGGGGTGCTAATGTTTCAATGGGAGGCATGGGTAATTATATAGACAATAACAGTCCA
TGCAACCAATACTATAAATCGGAAATGGCTAATAATTCCATAAATATCAACGAGTCAAT
GTTTATGGCCATGTGGGGTGTGTTAATGACGCCATTTCCGATAAACAAGGTGGCACCCT
AACGTCCCATCTTCTCTATTGAATCCGAATCACCAAGATTGGATGAGGATGACTGGTACG
AGCACACATATGTTAAACAATATAAATGCTAAAACAGAAATGGATTTTCCGGAAAATGAT
GGCAATGTCCACGGAGCCACCGGTGTAGATTTATCATCAACCCTGTCCAATGGTCAGCAG
CATAGTGCATTTGAAGGAAATGGAACATTTCCCGTGGTAAATACCGGCAATCCAAGAACG
ACGTTGAACTGTGCTCCTCAGTTGTCTTGGCTTGGAGGAAACCCGACAATTGCACATCCA
ATGAATGGGAATAATGAAACCCAGATAAAAAATGTATGCAAGCCTACCCCCAGCACAGGT
GTAATAAAAAAATTAACCTTTGACTGTGTTGACCGAGAGGAAGGAATAGGAAATTTTCCC
AACTGGCGTCTTTTGTAGCTATGAGAGAAAAATGTCAAACACGGTCAATCCTTCTTGT
GGAATTTCTTAACACAGTTGGAGGGATTACGCCAGCAGAATTCATTTGAAAAAATATA
TCCCTTGCTGATCTACAGTTTCTGAAGAAGATGATGTGCTTTCCAGTGCATCCTCGGTC
AGCTGTAATAGCAACTGCGACATGAAAATCGGAGTTTTCGCAGCAAGGTACCACTGTGTCT
AATTTACAACAAGGGTTTAGGCAACAAATGAACATGAACTCATCCGTGGTTAAAATGGAC
GGGAGCTATAAAACACAAGAGATGTCAAATGACTGCGCGACAGACGTACCGGATAATGCA
GGCATCATTCAAGAAAATAAACGTATATACTCTCAAATTGAGAATCCAGAGGATGGCCGC
GTACGGGAGGGGAGTGTTCAGACGTCACAAATCATCCTCCGCGAAAAGTGCAGGATT
CATAACATGAAGTCCGAAGGCGTAACCTGCGGTATGTGCATGACGGCTGCGGATTCCACA
CGGCAGGATGCGTCTGGAGGTTTCGAGCAGCGGAACCAAGAATGAATATTACGATGATGAT
TCGGAAATTAACCGGACTGTTCGGATACAGACAGCGACAACGAAGTCCAATGCTGTCAAGAA
GTAAAAAAGATCGGTCCCAAGACGTACTCCTCAGAAAATTTTAATCCTGAGTATCGCCAA
GCCAAGAGATTGCTTGCCGACATTCATACAGAAGATGGATTCCAGACCCAGTCGATATA
GAAGAAAACGACGGTCCGTTCAATCCAATAGTACTAGACCAGCCACTGTGTGTATGAAC
GGTAGACGCCGAAGGACTTTCTTCCGGCAATGTGTTACCGTTTTAGTCTCTGAGTAAA
TTAACTTACTTCAAAGAATTACTGCAGAGCTATGTGTCTAAAAATAACAACAGTTATTTA
AGTCTCAGTTGGCTCCCAAACACGGGGTGTATGTATGTCTAGAGAAGAAATTGGGGTAT
GAATATATACCTACCTTGAGAGACACGTTTTCCAGTGCATGTGGGTGGATGATGGTCTTT
GGAGTTGTGCGGTGCAGAACCGCCCGCGGCTCTGAATAAAAAACATGGTGAATCTCCTGTGT
GAAAATAAATGGGTGCTTCTGCATAATTAATCAAACTCTACACACGAGCTGTTTTTAGCC
GCCTCGGATTTGAAACAGTTTATGGAGGAGGGTCTGTCTAGATGTGATCCTATATATGAA
CAAGTAACAGTGCCTGACGGTGTGGCCATGGAGAATCTCTTCGAAACTTCTTGAGAAAC
AGTAGAACGTTTCAGGATCTTATGGATCAAAGGGACAACATGCACGGATGTAACGGACC
TTTAACGGAATGCCGTGGTAGACTAGGTGACAGAGTCATACATCTCTGCAATCCTGATTCA
GTTGATTCCATACCAGCAGACGAGGCCGTGATGTGAAGGAAGACCCTTGTATTTCTTT
GCATATGTTACCATTTCAAATCAAACCCCGCTACTAACGCTACCGTGTTAATAGCCCGG
GATAAAGATTTAAGAATATATGGCTATCATAAAGGTCGCCCCAGGATACGGTACTTGTGT
AAGAACGTACAACTTTCTTTAAAGCCGGGGCTAGAAAGTTCTATTTAGATTTTCAGATT
ACACCTAAACGCTTGTCTGCGGTCAACGAAGAGGAGTATCTGAATACTCTACAAAACGCT
CCTTGTCTGCTCTTAAACCATCTGTGTTTAGAAAGATTTACTCTCAGGAAGGTAAATAA

U100

>2284_U100

ATGGCAACCGCAAGGCTGGGCGCTATGAAACCGCCGAGATCTTGTGCGCTGATATTTTTG
TGGGCGTTTTTCGATGGCGACGGCCCCGACGAATGCTACGGCCACCGACGTGCGGGAACC
GTGAAATCCACACCGCCGCCAGAGGACAAAGGCAACTATACCGCAAAATACTATGACAAA
AACATTTATTTTAACATTTACGAAGGTAGAAATTTCTACGCCGAGGAGACGGACGCTATGG
GAAATAATCTCTAAATTTTCTACCTCCGAAATGTTAAGTTTAAGAAAAGTTAAAGCTTTC
GTCCCCGTAGATGATAACCCGACTACTACGTTAGAAGATATAGCAGACATCCTGAACTAC
GCCGTTTGCATGACAATAGCTGCGGTTGTACCATAGAGACACAAGCACGCATCATGTTTC
GGAGATATCATAATCTGCGTTCCTCTCTCGGCCGACAGCAAAGGAGTGAGAAATTTCAA
GATAAAATAATGCCATGCGTCTATCACAGATTCTCTCTAGCAGCCTCGGCCTTCACTG
TCTCTGCTTTATGGCGTTTTCGGGAGCAACTATAATAGTTTGGCCTATATGAGGCGCCTG
AAACCACTGACTGCGATGACGGCGATAGCTTTCTGTCCAATGACTACTAAATTAGAGCTA
CGTCAAAACTACAAAATGAAAGAGACTCTATGCGAACTGATCGTCAGTATCGAGATTTTG
AAAATACGAAACAATGGCGGACAGACTATGAAGACGTTAACCTCGTTTTGCGATCGTTTCA

AAAGACAACGACGGGCAAGATTGGGAAACCTGTACGCGATTTGCGCCCGTTAATATAGAG
GACATATTGAGGTATAAACGTGCCGCTAATGACACGTGTTGTCTGTCACCGAGACGTGCAG
CATGGTCGTTCTACGTTAGAATCTTCTAACTTATGGACTCAGACTCAATATTTTGAACCG
TGGCAAGACATCGTTGACGTATACGTCCCTATAAATGATACTCATTGTCCAAACGATAGC
TACGTGCTGTTTGAACACTGCAAGGTTTTCGAATGGTGTTCACGTTTAAATAAAAAACGAG
ACGAAGAATTATTTATCTTCCGTACTAGGGTTTAGAAATGCACTGTTTGAACCTGAGGAA
CTGATGGAAACCATCGCAATGCGACTTGCCTCACAAATATTGTCAATGGTTCGGACAGCAA
GGAACAACCTATACGCGACATAGATCCTGCCATCGTAAGCGCATTGTGGCACTCTTTGCCA
GAAAACCTGACAACGACTAATATAAAAATACGATATAGCCAGTCCGACACACATGGCTCCC
GCATTATGCACCATATTTGTACAGACCGGAACTTCAAACAACGCTTTAGAAATGCGGGA
CTACTGATGGTTAACAACATCTTTCACAGTCCAAGGGAGATATACTACACAAAATATGTTT
GAAAGAAAAGAATATGTATATAAACACCTCGGACAAGCTTTATGTCAGGACGGTGAATTT
CTGTTTCAAATGAAGGCCAGAAGTTTTGTGCGACCTCTCACGGACAATAGAACAATTGTA
TATACCATGCAAGACCAAGTTCAAACAACCTTGTCCGTAACGTGGATGGATTTCAATCTA
GTCATTAGTGACTACGGAAGAGATGTGATAAAACAATTTAACCAAAAAGTGCTATGTTGGCA
CGCAAAAACGGGCAAGATATCTACAGATGGAAACGTTTATATCGGATTTATTCCGTCAT
GAATGTCATCAAAAATAATTATTATGTGTTAGACAAAAAACTACAAATGTTTTATCCGACT
ACTCATAGTAACGAATTACTTTTTTATCCCTCGGAAGCTACGTTGCCCTCACCTTGGAAAG
GAACCTCCGTTTTCTCACCTTGGCCGGAACCTACGTTTCCCTCGCGTTGGTATTGGCTC
TTATTAATTTACACAAATTACTGA

>5055_U100

ATGGCAACCGCAAGGCTGGGCGCTATGAAACCGCCGAGATCTTGTGCGCTGATATTTTTG
TGGGCGTTTTTCGATGGCGACGGCCCGACGAATGCTACGGCCACCGACGTGCGGGAACC
GTGAAATCCACACCGCCGAGAGGACAAAGGCAACTATACCGCAAAATACTATGACAAA
AACATTTATTTTAAATTTACGAAGGTAGAAATTTCTACGCCGAGGAGACGGACGCTATGG
GAAATAATCTCTAAATTTTCTACCTCCGAAATGTTAAGTTTAAAGAAAAGTTAAAGCTTTC
GTCCCGTAGATGATAACCCGACTACTACGTTAGAAGATATAGCAGACATCCTGAACTAC
GCCGTTTTCGATGACAATAGCTGCGGTTGTACCATAGAGACACAAGCACGCATCATGTTT
GGAGATATCATAATCTGCGTTCCCTCTCTCGGCCGACAGCAAAGGAGTGAGAAAATTTCAA
GATAAAAATAATGCCATGCGGTTATCACAGATTCTCTCTAGCAGCCTCGGCCTTACCTG
TCTCTGCTTTTATGGCGTTTTCGGGAGCAACTATAATAGTTTGGCCTATATGAGGCGCCTG
AAACCACCTGACTGCGATGACGGCGATAGCTTTCTGTCCAATGACTACTAAATTAGAGCTA
CGTCAAAACTACAAAATGAAAGAGACTCTATGCGAACTGATCGTCAGTATCGAGATTTT
AAAATACGAAACAATGGCGGACAGACTATGAAGACGTTAACCTCGTTTTGCGATCGTTT
AAAGACAACGACGGGCAAGATTGGGAAACCTGTACGCGATTTGCGCCCGTTAATATAGAG
GACATATTGAGGTATAAACGTGCCGCTAATGACACGTGTTGTCTGTCACCGAGACGTGCAG
CATGGTCGTTCTACGTTAGAATCTTCTAACTTATGGACTCAGACTCAATATTTTGAACCG
TGGCAAGACATCGTTGACGTATACGTCCCTATAAATGATACTCATTGTCCAAACGATAGC
TACGTGCTGTTTGAACACTGCAAGGTTTTCGAATGGTGTTCACGTTTAAATAAAAAACGAG
ACGAAGAATTATTTATCTTCCGTACTAGGGTTTAGAAATGCACTGTTTGAACCTGAGGAA
CTGATGGAAACCATCGCAATGCGACTTGCCTCACAAATATTGTCAATGGTTCGGACAGCAA
GGAACAACCTATACGCGACATAGATCCTGCCATCGTAAGCGCATTGTGGCACTCTTTGCCA
GAAAACCTGACAACGACTAATATAAAAATACGATATAGCCAGTCCGACACACATGGCTCCC
GCATTATGCACCATATTTGTACAGACCGGAACTTCAAACAACGCTTTAGAAATGCGGGA
CTACTGATGGTTAACAACATCTTTCACAGTCCAAGGGAGATATACTACACAAAATATGTTT
GAAAGAAAAGAATATGTATATAAACACCTCGGACAAGCTTTATGTCAGGACGGTGAATTT
CTGTTTCAAATGAAGGCCAGAAGTTTTGTGCGACCTCTCACGGACAATAGAACAATTGTA
TATACCATGCAAGACCAAGTTCAAACAACCTTGTCCGTAACGTGGATGGATTTCAATCTA
GTCATTAGTGACTACGGAAGAGATGTGATAAAACAATTTAACCAAAAAGTGCTATGTTGGCA
CGCAAAAACGGGCAAGATATCTACAGATGGAAACGTTTATATCGGATTTATTCCGTCAT
GAATGTCATCAAGATAATTATTATGTGTTAGACAAAAAACTACAAATGTTTTATCCGACT
ACTCATAGTAACGAATTACTTTTTTATCCCTCGGAAGCTACGTTGCCCTCACCTTGGAAAG
GAACCTCCGTTTTCTCACCTTGGCCGGAACCTACGTTTCCCTCGCGTTGGTATTGGCTC
TTATTAATTTACACAAATTACTGA

>5814_U100

ATGGCAACCGCAAGGCTGGGCGCTATGAAACCGCCGAGATCTTGTGCGCTGATATTTTTG

TGGGCGTTTTTCGATGGCGACGGCCCCGACGAATGCTACGGCCCACCGACGTGCGGGAACC
GTGAAATCCACACCCGCCAGAGGACAAAGGCAACTATACCGCAAATACTATGACAAA
AACATTTATTTTAAACATTTACGAAGGTAGAAATTTCTACGCCGAGGAGACGGACGCTATGG
GAAATAATCTCTAAATTTTCTACCTCCGAAATGTTAAGTTTAAAGAAAAGTTAAAGCTTTC
GTCCCGTAGATGATAACCCGACTACTACGTTAGAAGATATAGCAGACATCCTGAACTAC
GCCGTTTTCGATGACAATAGCTGCGGTTGTACCATAGAGACACAAGCACGCATCATGTTTC
GGAGATATCATAATCTGCGTTCCTCTCTCGGCCGACAGCAAAGGAGTGAGAAATTTCAAA
GATAAAATAATGCCATATGGGTCTATCACAGATTCTCTCTAGCAGCCTCGGCCTTCACCTG
TCTCTGCTTTATGGCGCTTTCGGGAGCAACTATAATAGTTTGGCCTATATGAGGCGCCTG
AAACCACTGACTGCGATGACGGCGATAGCTTTCTGTCCAATGACTACTAAATTAGAGCTA
CGTCAAAACTACAAAATGAAAGAGACTCTATGCGAACTGATCGTCAGTATCGAGATTTTG
AAAATACGAAACAATGGCGGACAGACTATGAAGACGTTAACCTCGTTTTGCGATCGTTTCGA
AAAGACAACGACGGGCAAGATTGGGAAACCTGTACGCGATTTGCGCCCGTTAATATAGAG
GACATATTGAGGTATAAACGTGCCGCTAATGACACGTGTTGTGTCACCGAGACGTGCAG
CATGGTCGTTCTACGTTAGAATCTTCTAACTTATGGACTCAGACTCAATATTTTGAACCG
TGGCAAGACATCGTTGACGTATACGTCCCTATAAATGATACTCATTGTCCAAACGATAGC
TACGTGCTGTTTGAACACTGCAAGGTTTTCGAATGGTGTTCACGTTTAAATAAAAACGAG
ACGAAGAATTATTTATCTTCCGTACTAGGGTTTAGAAATGCACTGTTTGAACCTGAGGAA
CTGATGGAAACCATCGCAATGCGACTTGCCTCACAAATATTGTCAATGGTTCGGACAGCAA
GGAACAACCTATACGCGACATAGATCCTGCCATCGTAAGCGCATTGTGGCACTCTTTGCCA
GAAAACCTGACAACGACTAATATAAAAATACGATATAGCCAGTCCGACACACATGGCTCCC
GCATTTATGCACCATATTTGTACAGACCGGAACTTCAAAAACAACGCTTTAGAAATGCGGGA
CTACTGATGGTTAACAACATCTTTCACAGTCCAAGGGAGATATACTACACAAAATATGTTTC
GAAAGAAAAGAATATGTATATAAACACCTCGGACAAGCTTTATGTCAGGACGGTGAAATT
CTGTTTCAAAATGAAGGCCAGAAGTTTTGTGCGACCTCTCACGGACAATAGAACAATTGTA
TATACCATGCAAGACCAAGTTCAAAAACCTTGTCCGTAACGTGGATGGATTTCAATCTA
GTCATTAGTGACTACGGAAGAGATGTGATAAACAATTTAACCAAAAAGTGCTATGTTGGCA
CGCAAAAACGGGCAAGATATCTACAGATGGAAACGTTTATATCGGATTTATTCCGTCAT
GAATGTCATCAAAAATAATTTATTTATGTGTTAGACAAAAAACTACAAATGTTTTATCCGACT
ACTCATAGTAACGAATTACTTTTTTTATCCCTCGGAAGCTACGTTGCCCTCACCTTGGAAG
GAACCTCCGTTTTCTCACCTTGGCCGGAACCTACGTTTCCCTCGCGTTGGTATTGGCTC
TTATTAATTTACACAAATTACTGA

7.7 Alignment of the Origin of Lytic Replication of CI-HHV-6A

The origin of lytic replication [Turner et al., 2002] of ciHHV-6A isolates 2284, 5055 and 5814 compared to HHV-6A and HHV-6B reference strains U1102, AJ, GS, Z29 and HST.

Start of U41 Start ori-lyt(mori)67610

```

HHV-6A U1102   AACGGGAGCAGAAACTACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTTGATCACAGA 67641
HHV-6A GS     AACGGGAGCAGAAACTACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTTGATCACAAA
HHV-6A AJ     AACGGGAGCAGAAACTACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTTGATCACAAA
CI-HHV-6A 5055 AACGGGAGCAGAAACTACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTTGATCACAAA
CI-HHV-6A 2284 AACGGGAGCAGAAACTACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTTGATCACAAA
CI-HHV-6A 5814 AACGGGAGCAGAAACTACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTTGATCACAAA
HHV-6B Z29   AACAGGAGCAGAAACCACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTCGATCACAAA
HHV-6B HST   AACAGGAGCAGAAACCACCGTTTCGTTTTTCATCAGCCATCTTTGTGGATTCGATCACAAA
***          *****

```

```

HHV-6A U1102   AACAGAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
HHV-6A GS     AACAAAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
HHV-6A AJ     AACAAAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
CI-HHV-6A 5055 AACAAAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
CI-HHV-6A 2284 AACAAAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
CI-HHV-6A 5814 AACAAAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
HHV-6B Z29   AACAAAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
HHV-6B HST   AACAAAGATAATGGGGTTTTGTGGTGAATCCTTATATATATATGTTTGACGTAACATAAC
****          *****

```

```

HHV-6A U1102   ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAATTATACCGTTTTCTATATGAGGT
HHV-6A AJ     ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGGATTATACCGTTTTCTATATGAGGT
HHV-6A GS     ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGGATTATACCGTTTTCTATATGAGGT
CI-HHV-6A 5055 ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGGATTATACCGTTTTCTATATGAGGT
CI-HHV-6A 2284 ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAATTATACCGTTTTCTATATGAGGT
CI-HHV-6A 5814 ACGCGTCATCAAACATAAGAGTAAACCACAAGTTGAATTATACCGTTTTCTATATGAGGT
HHV-6B Z29   ACGCGTCATCAAACATAAAAAGTAAACCACAAGTTGAAATATACCGTTTTCTATATGAGTT
HHV-6B HST   ACGCGTCATCAAACATAAAAAGTAAACCACAAGTTGAAATATACCGTTTTCTATATGAGTT
*****          ***** * *****

```

```

HHV-6A U1102   TTACGGTCAAAAAGAAAAACGATTTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA
HHV-6A GS     TTACGGACAAAAGAAAAACGATTTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA
HHV-6A AJ     TTACGGACAAAAGAAAAACGATTTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA
CI-HHV-6A 5055 TTACGGACAAAAGAAAAACGATTTTTTTTATGCAAATATTTTCCACGCAGATGATATGACA
CI-HHV-6A 2284 TTACGGACAAAAGAAAAACGATTTTTTTTATGCAAAGATTTTCCACGCAGATGATATGACA
CI-HHV-6A 5814 TTACGGACAAAAGAAAAACGATTTTTTTTATGCAAAGATTTTCCACGCAGATGATATGACA
HHV-6B Z29   TTACGGACAAAAGAAAAACGATTTTCTTATGCAAATATTTTCCACGCAGATGATATGACA
HHV-6B HST   TTACGGACAAAAGAAAAACGATTTTCTTATGCAAATATTTTCCACGCAGATGATATGACA
*****          *****

```

OBP2

```

HHV-6A U1102   CGCCCCTTAATTTAAATTTATGCAAATCGTCGTCCACCTCAGGTACAATAGTATATTATA 67822
HHV-6A GS     CGCCCCTTAATTTAAATTTATGCAAATCGTCGTCCACCCAGGTACAATAGTATATTATA
HHV-6A AJ     CGCCCCTTAATTTAAATTTATGCAAATCGTCGTCCACCTCAGGTACAATAGTATATTATA
CI-HHV-6A 5055 CGCCCCTTAATTTAAATTTATGCAAATCGTCGTCCACCTCAGGTACAATAGTATATTATA
CI-HHV-6A 2284 CGCCCCTTAATTTAAATTTATGCAAATCGTCGTCCACCTCAGGTACAATAGTATATTATA
CI-HHV-6A 5814 CGCCCCTTAATTTAAATTTATGCAAATCGTCGTCCACCTCAGGTACAATAGTATATTATA
HHV-6B Z29   CGCCCCTTAATTTAAATTTATGTAATCGTCGTCCACCTCAGGTACAATAGTATATTATA
HHV-6B HST   CGCCCCTTAATTTAAATTTATGTAATCGTCGTCCACCTCAGGTACAATAGTATATTATA
*****          *****

```

OBP1

```

HHV-6A U1102   TATATATTATTTTAAATAAACTTATTGAGGACGGGAGAACGAGGGCGTGGCGTTTACGTCA 67882
HHV-6A GS     TATATATTATTTTAAATAAACTTATTGAGGACGGGAGAACGAGGGCGTGGCGTTTACGTCA
HHV-6A AJ     TATATATTATTTTAAATAAACTTATTGAGGACGGGAGAACGAGGGCGTGGCGTTTACGTCA
CI-HHV-6A 5055 TATATATTATTTTAAATAAACTTATTGAGGACGGGAGAACGAGGGCGTGGCGTTTACGTCA
CI-HHV-6A 2284 TATATATTATTTTAAATAAACTTATTGAGGACGGGAGAACGAGGGCGTGGCGTTTACGTCA
CI-HHV-6A 5814 TATATATTATTTTAAATAAACTTATTGAGGACGGGAGAACGAGGGCGTGGCGTTTACGTCA
HHV-6B Z29   TATATAGTTTTTAAATAAACTTATTGAGGACGGGAGAACGAAGGCGTGGCGTTTACGTCA

```

HHV-6B HST TATATAGTTTTTTAATAAACTTATTGAGGACGGGAGAACGAAGCGTGCGCTTTACGTCA
***** * *****
HHV-6A U1102 TAGCCTAATTATGCATTTCTCAGAACAGGATTTAAGAAGGCTGCGAGCGGCCACTGTTCA
HHV-6A GS TAGTCTAATTATGCATTTCTCAGAACAGGATTTAAAAAGGCTGCGAGCGGCTGACTGTTCA
HHV-6A AJ TAGTCTAATTATGCATTTCTCAGAACAGGATTTAAAAAGGCTGCGAGCGGCTGACTGTTCA
CI-HHV-6A 5055 TAGTCTAATTATGCATTTCTCAGAACAGGATTTAAAAAGGCTGCGAGCGGCTGACTGTTCA
CI-HHV-6A 2284 TAGCCTAATTATGCATTTCTAGAACAGGATTTAAAAAGGCTGCGAGCGGCCACTGTTCA
CI-HHV-6A 5814 TAGCCTAATTATGCATTTCTAGAACAGGATTTAAAAAGGCTGCGAGCGGCCACTGTTCA
HHV-6B Z29 TAGCCTAATTATACATTTCTCAGAACAGGATTTAAAAAGGCTGCGAGCGGCCGCTGTTCA
HHV-6B HST TAGCCTAATTATACATTTCTCAGAACAGGATTTAAAAAGGCTGCGAGCGGCCGCTGTTCA
*** *****

HHV-6A U1102 GAGGGACGCTGGGGTACGGCTTGATACGTTTGATTGAAAATGATTCTTCGTGTGCTATTT
HHV-6A GS GAGGGACGCTGGGGTACGGCTTGAGACGTTTGATTGAAAATGATTCTTCGTGTGCTATTT
HHV-6A AJ GAGGGACGCTGGGGTACGGCTTGAGACGTTTGATTGAAAATGATTCTTCGTGTGCTATTT
CI-HHV-6A 5055 GAGGGACGCTGGGGTACGGCTTGAGACGTTTGATTGAAAATGATTCTTCGTGTGCTATTT
CI-HHV-6A 2284 GAGGGACGCTGGGGTACGGCTTGAGACGTTTGATTGAAAATGATTCTTCGTGTGCTATTT
CI-HHV-6A 5814 GAGGGACGCTGGGGTACGGCTTGAGACGTTTGATTGAAAATGATTCTTCGTGTGCTATTT
HHV-6B Z29 GAGGGACGCTGGGGTACGACTTGAGACGTTTGACTGAAAATGATCCTTCGTGTACTATTT
HHV-6B HST GAGGGACGCTGGGGTACGACTTGAGACGTTTGACTGAAAATGATCCTTCGTGTACTATTT

HHV-6A U1102 TCTGCAAAAAAATTAATTCGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
HHV-6A GS TCTGCAAAAAAATTAATTCGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
HHV-6A AJ TCTGCAAAAAAATTAATTCGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
CI-HHV-6A 5055 TCTGCAAAAAAATTAATTCGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
CI-HHV-6A 2284 TCTGCAAAAAAATTAATTCGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
CI-HHV-6A 5814 TCTGCAAAAAAATTAATTCGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
HHV-6B Z29 TCTACAAAAAATTAATTTGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
HHV-6B HST TCTACAAAAAATTAATTTGCCGGCGACAGTAAACTTTTCAGCGGAATTTCAAAAAATTA
*** *****

IDR1

HHV-6A U1102 TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAACGTGAAAACATATTTACG 68112
HHV-6A GS TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAACGTGAAAACATATTTACG
HHV-6A AJ TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAACGTGAAAACATATTTACG
CI-HHV-6A 5055 TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAACGTGAAAACATATTTACG
CI-HHV-6A 2284 TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAACGTGAAAACATATTTACG
CI-HHV-6A 5814 TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAACGTGAAAACATATTTACG
HHV-6B Z29 TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAAGGTGAAAACATATTTACG
HHV-6B HST TTCCACATGTAATTTAAGCATTTTAAAACGTATAAACTCACAAGGTGAAAACATATTTACG

HHV-6A U1102 AATACAGTAGTTTTCCGTGATATTTTT-CGAAATTAAAAAATTTTAAATCGGGTAAATGAT
HHV-6A GS AATACAGTAGTTTTCCGTGATATTTTTTCGAAATTAAAAAATTTTAAATCGGGTAAATGAT
HHV-6A AJ AATACAGTAGTTTTCCGTGATATTTTTTCGAAATTAAAAAATTTTAAATCGGGTAAATGAT
CI-HHV-6A 5055 AATACAGTAGTTTTCCGTGATATTTTTTCGAAATTAAAAAATTTTAAATCGGGTAAATGAT
CI-HHV-6A 2284 AATACAGTAGTTTTCCGTGATATTTTTTCGAAATTAAAAAATTTTAAATCGGGTAAATGAT
CI-HHV-6A 5814 AATACAGTAGTTTTCCGTGATATTTTTTCGAAATTAAAAAATTTTAAATCGGGTAAATGAT
HHV-6B Z29 AATACAGTAGTTTTCCGTGATATTTTTTAAAAATTAATAAATTTTAAATCGTGAAGTAAT
HHV-6B HST AATACAGTAGTTTTCCGTGATATTTTTTAAAAATTAATAAATTTTAAATCGTGAAGTAAT

U1102 indel-1

HHV-6.U1102 AAAGCATATTATTAGATTTCTCACGTTAACAAAGCAAGTTTTTTTGGAGTTTTTCGGTATA 68242
HHV-6A GS AAAGCATACTA-----AGTTTTTT-GAGGTTTTTCGGTATA
HHV-6A AJ AAAGCATACTA-----AGTTTTTT-GAGGTTTTTCGGTATA
CI-HHV-6A 5055 AAAGCATACTA-----AGTTTTTT-GAGGTTTTTCGGTATA
CI-HHV-6A 2284 AAAGCATATTATTAGATTTCTCACGTTAACAAAGCAAGTTTTTT-GAGGTTTTTCGGTATA
CI-HHV-6A 5814 AAAGCATATTATTAGATTTCTCACGTTAACAAAGCAAGTTTTTT-GAGGTTTTTCGGTATA
HHV-6B Z29 AAAGCATACTATTAGATTTCTCACGTTAACAGAGCAAGCTCTTC-GAGGTTTTTCAGTAGA
HHV-6B HST AAAGCATACTATTAGATTTCTCACGTTAACAGAGCAAGCTCTTC-GAGGTTTTTCAGTAGA
***** ** * * * *****

IDR2

HHV-6A U1102 ATTTTAAACATTTACTTCCACATGTAATTTAAGCATTTTAAAACGTATAACTCACAACGTG 68302

HHV-6A GS ATGTTAACATTTAGTTCCACATGTAATTTAAGCGTTTTAAAAACGTATAGCTCACAACGTG
HHV-6A AJ ATGTTAACATTTAGTTCCACATGTAATTTAAGCGTTTTAAAAACGTATAGGTCACAACGTG
CI-HHV-6A 5055 ATGTTAACATTTAGTTCCACATGTAATTTAAGCGTTTTAAAAACGTATAGCTCACAACGTG
CI-HHV-6A 2284 ATTTTAAACATTTAGTTCCACATGTAATTTAAGCATTAAAAACGTATAACTCACAACGTG
CI-HHV-6A 5814 ATTTTAAACATTTAGTTCCACATGTAATTTAAGCATTAAAAACGTATAACTCACAACGTG
HHV-6B Z29 ATTTCAATATTTA-----
HHV-6B HST ATTTCAATATTTA-----
* * * * *

HHV-6A U1102 AAAACATTTACGAATACAGTAGTTTTAGTGATGTTTTTCGAAATTCAAAAAGTTTAAAT
HHV-6A GS AAAACATTTACGAATACAGTAGTTTTTCGTGATATTTTTTCGAAATTCAAAAATTTAAGT
HHV-6A AJ AAAACATTTACGAATACAGTAGTTTTTCGTGATATTTTTTCGAAATTCAAAAATTTAAT
CI-HHV-6A 5055 AAAACATTTACGAATACAGTAGTTTTTCGTGATATTTTTTCGAAATTAATAATTTTAAAT
CI-HHV-6A 2284 AAAACATTTACGAATACAGTAGTTTTTCGTGATATTTTTTCGAAATTCAAAAAGTTTAAAT
CI-HHV-6A 5814 AAAACATTTACGAATACAGTAGTTTTTCGTGATATTTTTTCGAAATTCAAAAAGTTTAAAT
HHV-6B Z29 -----
HHV-6B HST -----

I DR 3

HHV-6A U1102 CGGGTAAATGATAAAGCATATTATTAGATTACACATGTAATTTAAGCATTTTAAAAACGTA 68422
HHV-6A GS CGGGTAAATGATAAAGCATATTATTAGATTCCACATGTAATTTAAGCATTTTAAAAACGTA
HHV-6A AJ CGGGTAAATGATAAAGCATATTATTAGATTCCACATGTAATTTAAGCATTTTAAAAACGTA
CI-HHV-6A 5055 CGGGTAAATGATAAAGCATATTATTAGATTCCACATGTAATTTAAGCATTTTAAAAACGTA
CI-HHV-6A 2284 CGGGTAAATGATAAAGCATGTTATTAGATTACACATGTAATTTAAGCATTTTAAAAACGTA
CI-HHV-6A 5814 CGGGTAAATGATAAAGCATGTTATTAGATTACACATGTAATTTAAGCATTTTAAAAACGTA
HHV-6B Z29 -----ATTCCACATGTAATTTAAGCATTTTAAAAATGTA
HHV-6B HST -----ATTCCACATGTAATTTAAGCATTTTAAAAATGTA
* * * * *

HHV-6A U1102 GAATTCACAAAGTGACAAAACATT---AAATACAGTAGTTTTCACGGTATTTTTGAAAT
HHV-6A GS GAATTCACAAAGTGACAAAACATTCACAAATACAGTAGTTTTCACGGTATTTTTCGGGAT
HHV-6A AJ GAATTCACAAAGTGACAAAACATTCACAAATACAGTAGTTTTCACGGTATTTTTCGGGAT
CI-HHV-6A 5055 GAATTCACAAAGTGACAAAACATTCACAAATACAGTAGTTTTCACGGTATTTTTCGGGAT
CI-HHV-6A 2284 GAATTCACAAAGTGACAAAACATTGACGAATACAGTAGTTTTCACGGTATTTTTGGAAT
CI-HHV-6A 5814 GAATTCACAAAGTGACAAAACATTGACGAATACAGTAGTTTTCACGGTATTTTTGGAAT
HHV-6B Z29 GAATTCACAAAGTGACAAAACATTCACAAATACAGTAGTTTTCACGGTATTTTT-GAAAT
HHV-6B HST GAATTCACAAAGTGACAAAACATTCACAAATACAGTAGTTTTCACGGTATTTTT-GAAAT
* * * * *

HHV-6A U1102 TAA---ATTTTAAATCGGGTACATGATAAAGCATGCTGTTAAATTCACGTTTTTAGGGCA
HHV-6A GS TAA---ATTTTAAATCGGGTAAATGACAAAGCATGCTGTTAAATTCACGTTTTTAGGGCA
HHV-6A AJ TAA---ATTTTAAATCGGGTAAATGACAAAGCATGCTGTTAAATTCACGTTTTTAGGGCA
CI-HHV-6A 5055 TAA---ATTTTAAATCGGGTAAATGACAAAGCATGCTGTTAAATTCACGTTTTTAGGGCA
CI-HHV-6A 2284 TAA---ATTTTAAATCGGGTAAATGATAAAGCATGCTGTTAAATTCACATTTTTAGGGCA
CI-HHV-6A 5814 TAA---ATTTTAAATCGGGTAAATGATAAAGCATGCTGTTAAATTCACATTTTTAGGGCA
HHV-6B Z29 TAATAAATTTTAAATCGGGTAAATGGTAAAGCATGCTGTTAAATTCACGTTAACAGAGCA
HHV-6B HST TAATAAATTTTAAATCGGGTAAATGGTAAAGCATGCTGTTAAATTCACGTTAAAAGAGCA
* * * * *

HHV-6A U1102 AGTTCGCGAGGTTTGCTTTGTGATTATCAGTGAATGTATTCAGCGTTTCAATTTTAGAT
HHV-6A GS AGTTCGCGAGGTTTGCTCTGTGATTATCAGTGAAT--ATTCAGCGTTTCAATTTTAGAT
HHV-6A AJ AGTTCGCGAGGTTTGCTCTGTGATTATCAGTGAAT--ATTCAGCGTTTCAATTTTAGAT
CI-HHV-6A 5055 AGTTCGCGAGGTTTGCTCTGTGATTATCAGTGAAT--ATTCAGCGTTTCAATTTTAGAT
CI-HHV-6A 2284 AGTTCGCGAGGTTTGCTCTGTGATTATCAGTGAATGTATTCAGCGTTTCAATTTTAGAT
CI-HHV-6A 5814 AGTTCGCGAGGTTTGCTCTGTGATTATCAGTGAATGTATTCAGCGTTTCAATTTTAGAT
HHV-6B Z29 AGTTCGCGAGGTTTGCTATGTGTTTATCAGTGAATGTATTTAGTATTTCAATTTTAGAT
HHV-6B HST AGTTCGCGAGGTTTGCTATGTGTTTATCAGTGAATGTATTTAGTATTTCAATTTTAGAT
* * * * *

U1102 indel-2 End ori-lyt(mori)68712

HHV-6A U1102 TATGGTTCGGTAAAAATATAGGTCTTTGTAAATCATTGATTTAGCTTTTATATGTATT 68714
HHV-6A GS TATGGTTCGGTAAAAATATAGGTCTTTG-----
HHV-6A AJ TATGGTTCGGTAAAAATATAGGTCTTTG-----
CI-HHV-6A 5055 TATGGTTCGGTAAAAATATAGGTCTTTG-----

CI-HHV-6A 2284 TATGGTTCGGTAAAAATATAGGTCTTTGTAAAAATCATTGATTTAGCTTTTATATGTATT
CI-HHV-6A 5814 TATGGTTCGGTAAAAATATAGGTCTTTGTAAAAATCATTGATTTAGCTTTTATATGTATT
HHV-6B Z29 CATGGTTCGGTAAAGATATAAGTCCGTGAAAAATTTTTGGTTTAGTTTTCATATCTACT
HHV-6B HST CATGGTTCGGTAAAGATATAAGTCCGTGAAAAATTTTTGGTTTAGTTTTCATATCTACT
***** **

7.8 Optimisation of CCR5 Surface Expression

Steps were taken to optimise transient surface expression through investigation of several factors affecting electroporation and subsequent recombinant gene expression. Optimal gene expression is typically achieved with electroporation during early log phase. Therefore the growth rate of the cell lines was assessed under the defined equipment and laboratory conditions from Vaidehi et al., 2009, figure 7.4. Additionally the duration of time following transfection and amount of receptor construct plasmid DNA transfected were also considered with regards to their effects on construct cell surface expression, figures 7.5 and 7.6. Due to reagent constraints these were tested with expression of the human chemokine receptor CCR5 and an anti-CCR5 antibody. These defined a set of conditions which would be utilised to allow for maximal surface expression of the receptor constructs, reproducibility between assays but also take into consideration practical requirements of total cell numbers required for the downstream applications. Cells were subsequently electroporated when cultured cell density was $\sim 0.9 \times 10^6$ cells/ml, to ensure they were in early log phase. Electroporation occurred with $2 \mu\text{g}$ of plasmid DNA/ 1×10^6 cells, then cells were cultured for a further 48hrs before downstream analysis

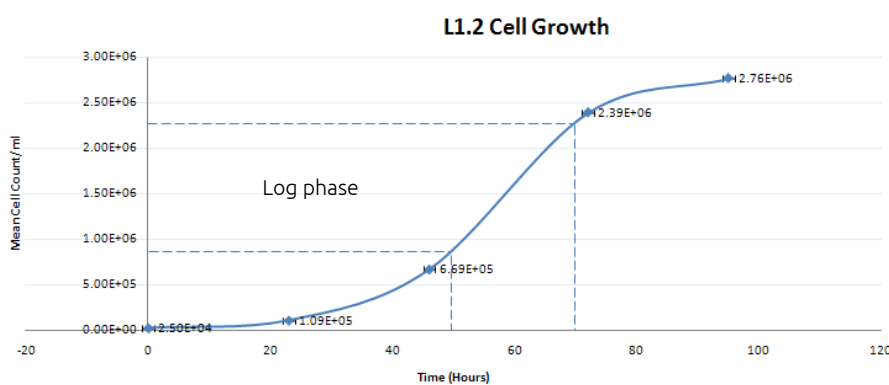


Figure 7.4 L1.2 cell growth time curve. Log phase indicated by dashed lines.

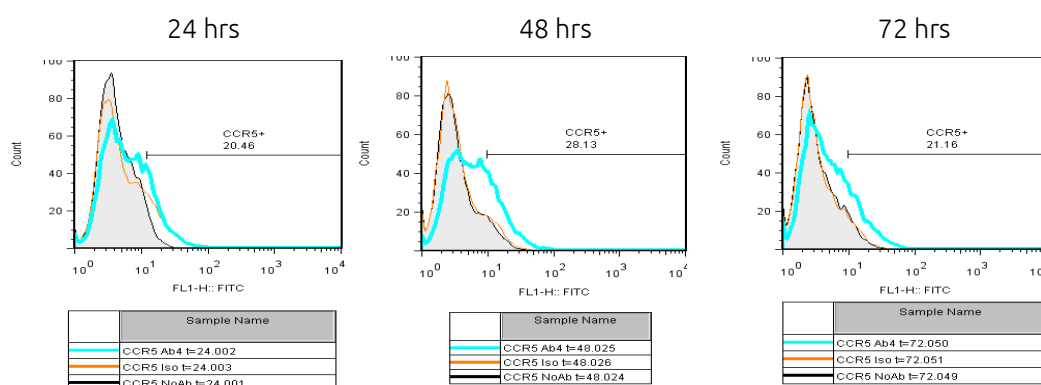


Figure 7.5 Surface expression of CCR5 at time points following transfection. Shaded with black outline indicates no stain. Orange indicated isotype control staining. Blue indicates anti-CCR5 staining. Times indicate number of hours post transfection.

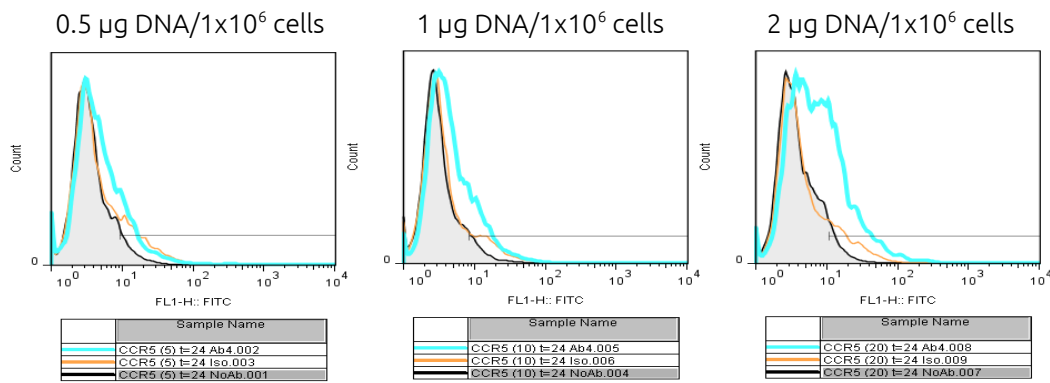


Figure 7.6 Effect of transfected DNA on surface expression of CCR5 at 48 hours post transfection. Shaded with black outline indicates no stain. Orange indicated isotype control staining. Blue indicates anti-CCR5 staining. Quantities represent the total amount of CCR5 plasmid DNA transfected. 2 µg DNA/1x10⁶ cells was the maximum tested due to concerns about altering the electroporation buffer, due to the low overall volume.

7.9 Generation of Stably Expressing Cell Lines

Cells were transfected, as described in section 2.12.4, concurrently with an experimental vector or the pEGFP vector (Clontech, Mountain View, California, USA). Transfected cells were then grown, as described in section 2.12.1, but additionally placed under selection through the addition of 400 µg/ml Geneticin (G418) (Life Technologies, Paisley, UK). Cells were grown for ~3 weeks in selective media until complete selection, as determined by 100% live cell EGFP expression. Cell stocks of experimentally transfected cells were then prepared as described in section 2.12.2.

7.10 Actin Polymerisation Assay

Actin polymerisation assays were performed based on those described previously [Burger et al., 2005; Udi et al., 2013]. 1.5×10^6 cells/ml or receptor transfected cells were centrifuged and resuspended in RPMI 1640/0.5% BSA (both Sigma Aldrich, Gillingham, UK). Cells were then stimulated with relevant chemokine, at 0, 15, 60, and 120 second time points 400 μ l of the stimulated cells was mixed with 100 μ l of a dying solution and incubated for 20 minutes at room temperature. The dying solution consisted of 50 μ g/ml fluorescein isothiocyanate (FITC)-labeled phalloidin, 1 mg/ml L- α -lysophosphatidylcholine and 37% formaldehyde in phosphate-buffered saline (all from Sigma Aldrich, Gillingham, UK). Cells were then washed twice by centrifugation at 400 g for 4 minutes in a Mistral 3000i centrifuge (MSE, Lower Sydenham, UK) and resuspension in PBS 1640/0.1% BSA (both Sigma Aldrich, Gillingham, UK). Intracellular F-actin was analysed by flow cytometry by comparison to a baseline of mean fluorescence prior to chemokine addition.

References

- Ablashi DV, Salahuddin SZ, Josephs SF, Imam F, Lusso P, Gallo RC, Hung C, Lemp J & Markham PD (1987) HBLV (or HHV-6) in human cell lines. *Nature*. **329**: 207
- Ablashi DV, Balachandran N, Josephs SF, Hung CL, Krueger GR, Kramarsky B, Salahuddin SZ & Gallo RC (1991) Genomic polymorphism, growth properties, and immunologic variations in human herpesvirus-6 isolates. *Virology*. **184**: 545-552.
- Ablashi DV, Agut H, Berneman Z, Campadelli-Fiume G, Carrigan D, Ceccerini-Nelli L, Chandran B, Chou S, Collandre H, Cone R, Dambaugh T, Dewhurst S, DiLuca D, Foa-Tomasi L, Fleckenstein B, Frenkel N, Gallo R, Gompels UA, Hall C, Jones M, Lawrence G, Martin M, Montagnier L, Neipel F, Nicholas J, Pellet P, Razzaque A, Torrelli G, Thomson B, Salahuddin S, Wyatt L & Yamanishi K (1993) Human herpesvirus-6 strain groups: a nomenclature. *Arch Virol*. **129**: 363-366.
- Ablashi DV, Lapps W, Kaplan M, Whitman JE, Richert JR & Pearson GR (1998a) Human herpesvirus-6 (HHV-6) infection in multiple sclerosis: a preliminary report. *Mult Scler*. **4**: 490-496.
- Ablashi DV, Marsh S, Kaplan M, Whitman JE & Pearson GR (1998b) HHV-6 infection in HIV-infected asymptomatic and AIDS patients. *Intervirology*. **41**: 1-9.
- Ablashi DV, Devin CL, Yoshikawa T, Lautenschlager I, Luppi M, Kuhl U & Komaroff AL (2010) Review part 3: human herpesvirus-6 in multiple non-neurological diseases. *J Med Virol*. **82**: 1903-1910.
- Ablashi D, Agut H, Alvarez-Lafuente R, Clark DA, Dewhurst S, Diluca D, Flamand L, Frenkel N, Gallo R, Gompels UA, Hollsberg P, Jacobson S, Luppi M, Lusso P, Malnati M, Medveczky P, Mori Y, Pellet PE, Pritchett JC, Yamanishi K & Yoshikawa (2014) Classification of HHV-6A and HHV-6B as distinct viruses. *Arch Virol*. **159**: 863-870.
- Achour A, Malet I, Deback C, Bonnafeous P, Boutolleau D, Gautheret-Dejean A & Agut H (2009) Length variability of telomeric repeat sequences of human herpesvirus 6 DNA. *J Virol Methods*. **159**: 127-130.
- Acott PD, Lee SH, Bitter-Suermann H, Lawen JG & Crocker JF (1996) Infection concomitant with pediatric renal allograft rejection. *Transplantation*. **62**: 689-691.
- Agut H, Collandre H, Aubin JT, Guetard D, Favier V, Ingrand D, Montagnier L & Huraux JM (1989) In vitro sensitivity of human herpesvirus-6 to antiviral drugs. *Res Virol*. **140**: 219-228.
- Agut H (2011) Deciphering the clinical impact of acute human herpesvirus 6 (HHV-6) infections. *J Clin Virol*. **52**: 164-171.
- Ahlgvist J, Fotheringham J, Akhyani N, Yao K, Fogdell-Hahn A & Jacobson S (2005) Differential tropism of human herpesvirus 6 (HHV-6) variants and induction of latency by HHV-6A in oligodendrocytes. *J Neurovirol*. **11**: 384-394.
- Ai LS, Lee SF, Chen SS & Liao F (2004) Molecular characterization of CCR6: involvement of

multiple domains in ligand binding and receptor signaling. *J Biomed Sci.* **11**: 818-828.

Akashi K, Eizuru Y, Sumiyoshi Y, Minematsu T, Hara S, Harada M, Kikuchi M, Niho Y & Minamishima Y (1993) Brief report: severe infectious mononucleosis-like syndrome and primary human herpesvirus 6 infection in an adult. *N Engl J Med.* **329**: 168-171.

Akhyani N, Berti R, Brennan MB, Soldan SS, Eaton JM, McFarland HF & Jacobson S (2000) Tissue distribution and variant characterization of human herpesvirus (HHV)-6: increased prevalence of HHV-6A in patients with multiple sclerosis. *J Infect Dis.* **182**: 1321-1325.

Akkapaiboon P, Mori Y, Sadaoka T, Yonemoto S & Yamanishi K (2004) Intracellular processing of human herpesvirus 6 glycoproteins Q1 and Q2 into tetrameric complexes expressed on the viral envelope. *J Virol.* **78**: 7969-7983.

Allen SJ, Crown SE & Handel TM (2007) Chemokine: receptor structure, interactions and antagonism. *Annu Rev Immunol.* **25**: 787-820.

Alvarez-Lafuente R, De las Heras V, Bartolome M, Picazo JJ & Arroyo R (2004) Relapsing-remitting multiple sclerosis and human herpesvirus 6 active infection. *Arch Neurol.* **61**: 1523-1527.

Ambrosini A, Paul S, Hu S & Riethman H (2007) Human subtelomeric duplicon structure and organization. *Genome Biol.* **8**: R151.

Amon W & Farrell PJ (2005) Reactivation of Epstein-Barr virus from latency. *Rev Med Virol.* **15**: 149-156.

Andrews SJ & Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet.* **15**: 193-204.

Arbuckle JH, Medveczky MM, Luka J, Hadley SH, Luegmayer A, Ablashi D, Lund TC, Tolar J, De Meirleir K, Montoya JG, Komaroff AL, Ambros PF & Medveczky PG (2010) The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vitro and in vivo. *Proc Natl Acad Sci U S A.* **107**: 5563-5568.

Arbuckle JH & Medveczky PG (2011) The molecular biology of human herpesvirus-6 latency and telomere integration. *Microbes Infect.* **13**: 731-741.

Arbuckle JH, Pantry SN, Medveczky MM, Prichett J, Loomis KS, Ablashi D & Medveczky PG (2013) Mapping the telomere integrated genome of human herpesvirus 6A and 6B. *Virology.* **442**: 3-11.

Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, Holdorf M, Weissman JS & Ganem D (2014) KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.* **10**: e1003847.

Arvin AM & Gildea D (2013) Varicella-Zoster Virus. In Knipe D & Howley PM (Eds.) *Fields Virology* (6th edition), Lippincott, Williams and Wilkins, Philadelphia.

- Asano Y, Yoshikawa T, Suga S, Kobayashi I, Nakashima T, Yazaki T, Kajita Y & Ozaki T (1994) Clinical features of infants with primary human herpesvirus 6 infection (exanthem subitum, roseola infantum). *Pediatrics*. **93**: 104-108.
- Ashrafpoor G, Andreoletti L, Bruneval P, Macron L, Azarine A, Lepillier A, Danchin N, Mousseaux E & Redheuil A (2013) Fulminant human herpesvirus 6 myocarditis in an immunocompetent adult. Role of cardiac magnetic resonance in a multidisciplinary approach. *Circulation*. **128**: 445-447.
- Assefa S, Keane TM, Otto TD, Newbold C & Berriman M (2009) ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*. **25**: 1968-1969.
- Aubin JT, Collandre H, Candotti D, Ingrand D, Rouzioux C, Burgard M, Richard S, Huraux JM & Agut H (1991) Several groups among human herpesvirus 6 strains can be distinguished by Southern blotting and polymerase chain reaction. *J Clin Microbiol*. **29**: 367-372.
- Aubry V, Mure F, Mariame B, Deschamps T, Wyrwicz LS, Manet E & Gruffat H (2014) Epstein-Barr virus late gene transcription depends on the assembly of a virus-specific preinitiation complex. *J Virol*. **88**: 12825-12838.
- Auger GA, Pease JE, Shen X, Xanthou G & Barker MD (2002) Alanine scanning mutagenesis of CCR3 reveals that the three intracellular loops are essential for functional receptor expression. *Eur J Immunol*. **32**: 1052-1058.
- Avalos BR, Bartynski KJ, Elder PJ, Kotur MS, Burton WG & Wilkie NM (1994) The active monomeric form of macrophage inflammatory protein-1 alpha interacts with high- and low-affinity classes of receptors on human hematopoietic cells. *Blood*. **84**: 1790-1801.
- Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E & Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*. **318**: 798-801.
- Baaten BJ, Staines KA, Smith LP, Skinner H, Davison TF & Butter C (2009) Early replication in pulmonary B cells after infection with Marek's disease herpesvirus by the respiratory route. *Viral Immunol*. **22**: 431-444.
- Bachelier F, Ben-Baruch A, Burkhardt AM, Combadiere C, Farber JM, Graham GJ, Horuk R, Sparre-Ulrich AH, Locati M, Luster AD, Mantovani A, Matsushima K, Murphy PM, Nibbs R, Nomiyama H, Power CA, Proudfoot AE, Rosenkilde MM, Rot A, Sozzani S, Thelen M, Yoshie O & Zlotnik A (2013) International union of pharmacology. LXXXIX. Update on the extended family of chemokine receptors and introducing a new nomenclature for atypical chemokine receptors. *Pharmacol Rev*. **66**: 1-79.
- Bacon K, Baggiolini M, Broxmeyer H, Horuk R, Lindley I, Mantovani A, Matsushima K, Murphy P, Nomiyama H, Oppenheim J, Rot A, Schall T, Tsang M, Thorpe R, Van Damme J, Wadhwa M, Yoshie O, Zlotnik A & Zoon K (2002) Chemokine/chemokine receptor nomenclature. *J Interferon Cytokine Res*. **22**: 1067-1068.
- Baird DM, Jeffreys AJ & Royle NJ (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human

Xp/Yp telomere. *EMBO J.* **14**: 5433-5443.

Bajetto A, Bonavia R, Barbero S, Piccioli P, Costa A, Florio T & Schettini G (1999) Glial and neuronal cells express functional chemokine receptor CXCR4 and its natural ligand stromal cell-derived factor-1. *J Neurochem.* **73**: 2348-2357.

Baker M (2010) Structural biology: The gatekeepers revealed. *Nature.* **465**: 823-826.

Balada E, Vilardell-Tarres M & Ordi-Ros J (2010) Implication of human endogenous retroviruses in the development of autoimmune diseases. *Int Rev Immunol.* **29**: 351-370.

Bannert N, Craig S, Farzan M, Sogah D, Santo NV, Choe H & Sodroski J (2001) Sialylated O-glycans and sulfated tyrosines in the NH₂-terminal domain of CC chemokine receptor 5 contribute to high affinity binding of chemokines. *J Exp Med.* **194**: 1661-1673.

Barzon L, Lavezzo E, Costanzi G, Franchin E, Toppo S & Palu G (2013) Next-generation sequencing technologies in diagnostic virology. *J Clin Virol.* **58**: 346-350.

Bates M, Monze M, Bima H, Kapambwe M, Clark D, Kasolo FC & Gompels UA (2009) Predominant human herpesvirus 6 variant A infant infections in an HIV-1 endemic region of Sub-Saharan Africa. *J Med Virol.* **81**: 779-789.

Baur JA, Zou Y, Shay JW & Wright WE (2001) Telomere position effect in human cells. *Science.* **292**: 2075-2077.

Bazan JF, Bacon KB, Hardiman G, Wang W, Soo K, Rossi D, Greaves DR, Zlotnik A & Schall TJ (1997) A new class of membrane-bound chemokine with a CXC3C motif. *Nature.* **385**: 640-644.

Beerenwinkel N & Zagordi O (2011) Ultra-deep sequencing for the analysis of viral populations. *Curr Opin Virol.* **1**: 413-418.

Beisser PS, Laurent L, Virelizier JL & Michelson S (2001) Human cytomegalovirus chemokine receptor gene US28 is transcribed in latently infected THP-1 monocytes. *J Virol.* **75**: 5949-5957.

Belyi VA, Levine AJ & Skalka AM (2010) Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog.* **6**: e1001030.

Benjamini Y & Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**: e72.

Berahovich RD, Miao Z, Wang Y, Premack B, Howard MC & Schall TJ (2005) Proteolytic activation of alternative CCR1 ligands in inflammation. *J Immunol.* **174**: 7341-7351.

Bergallo M, Galliano I, Montanari P, Gambarino S, Mareschi K, Ferro F, Fagioli F, Tovo PA & Ravanini P (2015) CMV induces HERV-K and HERV-W expression in kidney transplant recipients. *J Clin Virol.* **68**: 28-31.

- Berneman ZN, Ablashi DV, Li G, Eger-Fletcher M, Reitz MS, Hung CL, Brus I, Komaroff AL & Gallo RC (1992) Human herpesvirus 7 is a T-lymphotropic virus and is related to, but significantly different from, human herpesvirus 6 and human cytomegalovirus. *Proc Natl Acad Sci U S A.* **89**: 10552-10556.
- Berns KI, Pinkerton TC, Thomas GF & Hoggan MD (1975) Detection of adeno-associated virus (AAV)-specific nucleotide sequences in DNA isolated from latently infected Detroit 6 cells. *Virology.* **68**: 556-560.
- Berthet C, Raj K, Saudan P & Beard P (2005) How adeno-associated virus Rep78 protein arrests cells completely in S phase. *Proc Natl Acad Sci U S A.* **102**: 13634-13639.
- Besemer J, Lomsadze A & Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**: 2607-2618.
- Bhattacharjee B, Artiles AB, Renzette N, Calvo-Calle MJ, Stern LJ & Kowalik TF (2014) Whole genome, population sequencing of HHV-6 reveals the presence of variability in laboratory grown viruses. *Unpublished.*
- Biancotto A, Grivel JC, Lisco A, Vanpouille C, Markham PD, Gallo RC, Margolis LB & Lusso P (2009) Evolution of SIV toward RANTES resistance in macaques rapidly progressing to AIDS upon coinfection with HHV-6A. *Retrovirology.* **6**: 61-71.
- Bigalke B, Klingel K, May AE, Kandolf R & Gawaz MG (2007) Human herpesvirus 6 subtype A-associated myocarditis with 'apical ballooning'. *Can J Cardiol.* **23**: 393-395.
- Billstrom MA, Johnson GL, Avdi NJ & Worthen GS (1998) Intracellular signaling by the chemokine receptor US28 during human cytomegalovirus infection. *J Virol.* **72**: 5535-5544.
- Birnbaum T, Padovan CS, Sporer B, Rupprecht TA, Ausserer H, Jaeger G & Pfister HW (2005) Severe meningoencephalitis caused by human herpesvirus 6 B in an immunocompetent woman treated with ganciclovir. *Clin Infect Dis.* **40**: 887-889.
- Birnboim HC & Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513-1523.
- Blanpain C, Doranz BJ, Vakili J, Rucker J, Govaerts C, Baik SS, Lorthioir O, Migeotte I, Libert F, Baleux F, Vassart G, Doms RW & Parmentier M (1999) Multiple charged and aromatic residues in CCR5 amino-terminal domain are involved in high affinity binding of both chemokines and HIV-1 Env protein. *J Biol Chem.* **274**: 34719-34727.
- Blanpain C, Doranz BJ, Bondue A, Govaerts C, De Leener A, Vassart G, Doms RW, Proudfoot A & Parmentier M (2003) The core domain of chemokines binds CCR5 extracellular domains while their amino terminus interacts with the transmembrane helix bundle. *J Biol Chem.* **278**: 5179-5187.
- Bodaghi B, Jones TR, Zipeto D, Vita C, Sun L, Laurent L, Arenzana-Seisdedos F, Virelizier JL & Michelson S (1998) Chemokine sequestration by viral chemoreceptors as a novel viral escape strategy: withdrawal of chemokines from the environment of cytomegalovirus-

infected cells. *J Exp Med.* **188**: 855-866.

Bogner E, Radsak K & Stinski MF (1998) The gene product of human cytomegalovirus open reading frame UL56 binds the pac motif and has specific nuclease activity. *J Virol.* **72**: 2259-2264.

Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* **30**: 2114-2120.

Boller K, Schonfeld K, Lischer S, Fischer N, Hoffmann A, Kurth R & Tonjes RR (2008) Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *J Gen Virol.* **89**: 567-572.

Bondue A, Jao SC, Blanpain C, Parmentier M & LiWang PJ (2002) Characterization of the role of the N-loop of MIP-1 beta in CCR5 binding. *Biochemistry.* **41**: 13548-13555.

Bongers G, Maussang D, Muniz LR, Noriega VM, Fraile-Ramos A, Barker N, Marchesi F, Thirunarayanan N, Vischer HF, Qin L, Mayer L, Harpaz N, Leurs R, Furtado GC, Clevers H, Tortorella D, Smit MJ & Lira SA (2010) The cytomegalovirus-encoded chemokine receptor US28 promotes intestinal neoplasia in transgenic mice. *J Clin Invest.* **120**: 3969-3978.

Boomker JM, de Jong EK, de Leij LF & Harmsen MC (2006a) Chemokine scavenging by the human cytomegalovirus chemokine decoy receptor US28 does not inhibit monocyte adherence to activated endothelium. *Antiviral Res.* **69**: 124-127.

Boomker JM, The TH, de Leij LF & Harmsen MC (2006b) The human cytomegalovirus-encoded receptor US28 increases the activity of the major immediate-early promoter/enhancer. *Virus Res.* **118**: 196-200.

Borenstein R & Frenkel N (2009) Cloning human herpes virus 6A genome into bacterial artificial chromosomes and study of DNA replication intermediates. *Proc Natl Acad Sci U S A.* **106**: 19138-19143.

Borenstein R, Zeigerman H & Frenkel N (2010) The DR1 and DR6 first exons of human herpesvirus 6A are not required for virus replication in culture and are deleted in virus stocks that replication well in T-cell lines. *J Virol.* **84**: 2648-2656.

Borst EM, Kleine-Albers J, Gabaev I, Babic M, Wagner K, Binz A, Degenhardt I, Kalesse M, Jonjic S, Bauerfeind R & Messerle M (2013) The human cytomegalovirus UL51 protein is essential for viral genome cleavage-packaging and interacts with the terminase subunits pUL56 and pUL89. *J Virol.* **87**: 1720-1732.

Boutolleau D, Duros C, Bonnafous , Caiola D, Karras A, Castro ND, Ouachee M, Narcy P, Gueudin M, Agut H & Gautheret-Dejean A (2006) Identification of human herpesvirus 6 variants A and B by primer-specific real-time PCR may help to revisit their respective role in pathology. *J Clin Virol.* **35**: 257-263.

Bowman EP, Campbell JJ, Druey KM, Scheschonka A, Kehrl JH & Butcher EC (1998) Regulation of chemotactic and proadhesive responses to chemoattractant receptors by RGS (regulator of G-protein signaling) family members. *J Biol Chem.* **273**: 28040-28048.

- Bradel-Tretheway BG, Zhen Z & Dewhurst S (2003) Effects of codon-optimisation on protein expression by the human herpesvirus 6 and 7 U51 open reading frame. *J Virol Methods*. **111**: 145-156.
- Brelot A, Heveker N, Montes M & Alizon M (2000) Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem*. **275**: 23736-23744.
- Bresler M, Sheehan S, Chan AH & Song YS (2012) Telescopier: de novo assembly of highly repetitive regions. *Bioinformatics*. **28**: i311-i317.
- Britt-Compton B, Rowson J, Locke M, Mackenzie I, Kipling D & Baird DM (2006) Structural stability and chromosome-specific telomere length is governed by cis-acting determinants in humans. *Hum Mol Genet*. **15**: 725-733.
- Brown JC & Newcomb WW (2011) Herpesvirus capsid assembly: insight from structural analysis. *Curr Opin Virol*. **1**: 142-149.
- Brudek T, Christensen T, Hansen HJ, Bobecka J & Moller-Larson A (2004) Simultaneous presence of endogenous retrovirus and herpes virus antigens has profound effect on cell-mediated immune responses: implications for multiple sclerosis. *AIDS Res Hum Retroviruses*. **20**: 415-423.
- Buerger I, Reefschlaeger J, Bender W, Eckenberg P, Popp A, Weber O, Graeper S, Klenk HD, Ruebsamen-Waigmann H & Hallenberger S (2001) A novel non-nucleoside inhibitor specifically targets cytomegalovirus DNA maturation via UL89 and UL56 gene products. *J Virol*. **75**: 9077-9086.
- Bultmann BD, Klingel K, Nabauer M, Wallwiener D & Kandolf R (2005) High prevalence of viral genomes and inflammation in peripartum cardiomyopathy. *Am J Obstet Gynecol*. **193**: 363-365.
- Burg JS, Ingram JR, Venkatakrishnan AJ, Jude KM, Dukkipati A, Feinberg EN, Angelini A, Waghay D, Dror RO, Ploegh HL & Garcia KC (2015) Structural biology. Structural basis for chemokine recognition and activation of a viral G protein-coupled receptor. *Science*. **347**: 1113-1117.
- Burger M, Hartmann T, Krome M, Rawluk J, Tamamura H, Fujii N, Kipps TJ & Burger JA (2005) Small peptide inhibitors of the CXCR4 chemokine receptor (CD184) antagonize the activation, migration, and antiapoptotic responses of CXCL12 in chronic lymphocytic leukemia B cells. *Blood*. **106**: 1824-1830.
- Cahill MJ, Koser CU, Ross NE & Archer JA (2010) Read length and repeat resolution: exploring prokaryote genomes using next-generation sequencing technologies. *PLoS One*. **5**: e11518.
- Calcedo R, Vandenberghe LH, Gao G, Lin J & Wilson JM (2009) Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. *J Infect Dis*. **199**: 381-390.

- Calnek BW (2001) Pathogenesis of Marek's disease virus infection. *Curr Top Microbiol Immunol.* **255**: 25-55.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics.* **10**: 421.
- Campanella GS, Lee EM, Sun J & Luster AD (2003) CXCR3 and heparin binding sites of the chemokine IP-10 (CXCL10). *J Biol Chem.* **278**: 17066-17074.
- Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J & Cavalli-Sforza LL (2002) A human genome diversity cell line panel. *Science.* **296**: 261-262.
- Cardinali G, Gentile M, Cirone M, Zompetta C, Frati L, Faggioni A & Torrisi MR (1998) Viral glycoproteins accumulate in newly formed annulate lamellae following infection of lymphoid cells by human herpesvirus 6. *J Virol.* **72**: 9738-9746.
- Cardone J, Le Friec G & Kemper C (2011) CD46 in innate and adaptive immunity: an update. *Clin Exp Immunol.* **164**: 301-311.
- Caruso A, Favilli F, Rotola A, Comar M, Horejsh D, Alessandri G, Grassi M, Di Luca D & Fiorentini S (2003) Human herpesvirus-6 modulates RANTES production in primary human endothelial cell cultures. *J Med Virol.* **70**: 451-458.
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J (2005) ACT: the artemis comparison tool. *Bioinformatics.* **21**: 3422-3423.
- Carver TJ, Harris SR, Berriman M, Parkhill J & McQuillan JA (2012) Artemis: an integrated platform for visualisation and analysis of high-throughput sequence-based experimental data. *Bioinformatics.* **28**: 464-469.
- Casarosa P, Bakker RA, Verzijl D, Navis M, Timmerman H, Leurs R & Smit MJ (2001) Constitutive signaling of the human cytomegalovirus-encoded chemokine receptor. *J Biol Chem.* **276**: 11333-11337.
- Caselli E, Bracci A, Galvan M, Boni M, Rotola A, Bergamini C, Cermelli C, Dal Monte P, Gompels UA, Cassai E & Di Luca D (2006) Human herpesvirus (HHV-6) U94/REP protein inhibits betaherpesvirus replication. *Virology.* **346**: 402-414.
- Caserta MT, McDermott MP, Dewhurst S, Schnabel K, Carnahan JA, Gilbert L, Lathan G, Lofthus GK & Hall CB (2004) Human herpesvirus 6 (HHV6) DNA persistence and reactivation in healthy children. *J Pediatr.* **145**: 478-484.
- Caserta MT, Hall CB, Schnabel K, Lofthus G, Marino A, Shelley L, Yoo C, Carnahan J, Anderson L & Wang H (2010) Diagnostic assays for active infection with human herpesvirus 6 (HHV-6). *J Clin Virol.* **48**: 55-57.

- Caserta MT, Hall CB, Canfield RL, Davidson P, Lofthaus G, Schnabel K, Carnahan J, Shelley L & Wang H (2014) Early developmental outcomes of children with congenital HHV-6 infection. *Pediatrics*. **134**: 1111-1118.
- Catusse J, Parry CM, Dewin DR & Gompels UA (2007) Inhibition of HIV-1 infection by viral chemokine U83A via high-affinity CCR5 interactions that block human chemokine-induced leukocyte chemotaxis and receptor internalization. *Blood*. **109**: 3633-3639.
- Catusse J, Spinks J, Mattick C, Dyer A, Laing K, Fitzsimons C, Smit MJ and Gompels UA (2008) Immunomodulation by herpesvirus U51A chemokine receptor via CCL5 and FOG-2 down-regulation plus XCR1 and CCR7 mimicry in human leukocytes. *Eur J Immunol*. **38**: 763-777.
- Catusse J, Clark DJ & Gompels UA (2009) CCR5 signalling, but not DARC or D6 regulatory, chemokine receptors are targeted by herpesvirus U83A chemokine which delays receptor internalisation via diversion to a caveolin-linked pathway. *J Inflamm (Lond)*. **6**: 22.
- Chakera A, Seeber RM, John AE, Eidne KA & Greaves DR (2008) The duffy antigen/receptor for chemokines exists in an oligomeric form in living cells and functionally antagonizes CCR5 signaling through hetero-oligomerisation. *Mol Pharmacol*. **73**: 1362-1370.
- Chakravarty L, Rogers L, Quach T, Breckenridge S & Kolattukudy PE (1998) Lysine 58 and histidine 66 at the C-terminal alpha-helix of monocyte chemoattractant protein-1 are essential for glycosaminoglycan binding. *J Biol Chem*. **273**: 29641-29647.
- Challoner PB, Smith KT, Parker JD, MacLeod DL, Coulter SN, Rose TM, Schultz ER, Bennett JL, Garber RL & Chang M (1995) Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proc Natl Acad Sci U S A*. **92**: 7440-7444.
- Chan PK, Ng HK, Hui M & Cheng AF (2001) Prevalence and distribution of human herpesvirus 6 variants A and B in adult human brain. *J Med Virol*. **64**: 42-46.
- Chang YL, Parker ME, Nuovo G & Miller JB (2009) Human herpesvirus 6-related fulminant myocarditis and hepatitis in an immunocompetent adult with fatal outcome. *Hum Pathol*. **40**: 740-745.
- Chapenko S, Millers A, Nora Z, Logina I, Kukaine R & Murovska M (2003) Correlation between HHV-6 reactivation and multiple sclerosis disease activity. *J Med Virol*. **69**: 111-117.
- Chee MS, Satchwell SC, Preddie E, Weston KM & Barrell BG (1990) Human cytomegalovirus encodes three G protein-coupled receptor homologues. *Nature*. **344**: 774-777.
- Chelli M & Alizon M (2001) Determinants of the trans-dominant negative effect of truncated forms of the CCR5 chemokine receptor. *J Biol Chem*. **276**: 46975-46982.
- Chen Y, Green SR, Almazan F & Quehenberger O (2006) The amino terminus and third extracellular loop of CX3CR1 contain determinants critical for distinct receptor functions. *Mol Pharmacol*. **69**: 857-865.

- Cheung AK, Hoggan MD, Hauswirth WW & Berns KI (1980) Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. *J Virol.* **33**: 739-748.
- Chevigne A, Fievez V, Schmit JC & Deroo S (2011) Engineering and screening the N-terminus of chemokines for drug discovery. *Biochem Pharmacol.* **82**: 1438-1456.
- Choo KB, Pan CC & Han SH (1987) Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. *Virology.* **161**: 259-261.
- Chou S & Marousek GI (1992) Homology of the envelope glycoprotein B of human herpesvirus-6 and cytomegalovirus. **191**: 523-528.
- Chung CT, Niemela SL & Miller RH (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A.* **86**: 2172-2175.
- Cirone M, Campadelli-Fiume G, Foa-Tomasi L, Torrisi MR & Faggioni A (1994) Human herpesvirus 6 envelope glycoproteins B and H-L complex are undetectable on the plasma membrane of infected lymphocytes. *AIDS Res Hum Retroviruses.* **10**: 175-179.
- Clark DA & Griffiths PD (2003) Human herpesvirus 6: relevance of infection in the immunocompromised host. *Br J Haematol.* **120**: 384-395.
- Clark DA, Tsao EHF, Leong HN, Ward KN, Nacheva EP & Griffiths PD (2006a) Reply to Boutolleau et al. and Luppi et al. *J Infect Dis.* **194**: 1021-1023.
- Clark DA, Nacheva EP, Leong HN, Brazma D, Li YT, Tsao EH, Buyck HC, Atkinson CE, Lawson HM, Potter MN & Griffiths PD (2006b) Transmission of integrated human herpesvirus 6 through stem cell transplantation: implications for laboratory diagnosis. *J Infect Dis.* **193**: 912-916.
- Clark DA & Ward KN (2008) Importance of chromosomally integrated HHV-6A and HHV-6B in the diagnosis of active HHV-6 infection. *Herpes.* **15**: 28-32.
- Clark DJ (2011) Identification of specificity determinants of human herpesvirus 6 chemokines and relations to immunotherapy. Ph.D. Thesis. London School of Hygiene & Tropical Medicine. U.K.
- Clark DJ, Catusse J, Stacey A, Borrow P & Gompels UA (2013) Activation of CCR2+ human proinflammatory monocytes by human herpesvirus-6B chemokine N-terminal peptide. *J Gen Virol.* **94**: 1624-1635.
- Clubb RT, Omichinski JG, Clore GM & Gronenborn AM (1994) Mapping the binding surface of IL-8 complexed with an N-terminal fragment of the type 1 human interleukin-8 receptor. *FEBS Lett.* **338**: 93-97.
- Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC & Lusso P (1995) Identification of

RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science*. **270**: 1811-1815.

Cole C, Barber JD & Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*. **36**: W197-201.

Coleman CB, Nealy MS & Tibbetts SA (2010) Immature and transitional B cells are latency reservoirs for a gammaherpesvirus. *J Virol*. **84**: 13045-13052.

Collot S, Petit B, Bordessoule D, Alain S, Touati M, Denis F & Ranger-Rogez S (2002) Real-time PCR for quantification of human herpesvirus 6 DNA from lymph nodes and saliva. *J Clin Microbiol*. **40**: 2445-2451.

Colvin RA, Campanella GS, Manice LA & Luster AD (2006) CXCR3 requires tyrosine sulfation for ligand binding and a second extracellular loop arginine residue for ligand-induced chemotaxis. *Mol Cell Biol*. **26**: 5838-5849.

Combadiere C, Ahuja SK & Murphy PM (1995) Cloning and functional expression of a human eosinophil CC chemokine receptor. *J Biol Chem*. **270**: 16491-16494.

Cone RW, Huang ML, Hackman RC & Corey L (1996) Coinfection with human herpesvirus 6 variants A and B in lung tissue. *J Clin Microbiol*. **34**: 877-881.

Contreras-Galindo R, Kaplan MH, Dube D, Gonzalez-Hernandez MJ, Chan S, Meng F, Dai M, Omenn GS, Gitlin SD & Markovitz DM (2015) Human Endogenous Retroviruses Type-K (HERV-K) Virus Particles Package and Transmit HERV-K-Related Sequences. *J Virol*. doi: JVI.00544-15.

Copeland KF (2005) Modulation of HIV-1 transcription by cytokines and chemokines. *Mini Rev Med Chem*. **5**: 1093-1101.

Corbin-Lickfett KA, Chen IH, Cocco MJ & Sandri-Goldin RM (2009) The HSV-1 ICP27 RGG box specifically binds flexible, GC-rich sequences but not G-quartet structures. *Nucleic Acids Res*. **37**: 7290-7301.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW & Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J Am Chem Soc*. **117**: 5179-5197.

Croft M (2010) Control of immunity by the TNFR-related molecule OX40 (CD134). *Annu Rev Immunol*. **28**: 57-78.

Crump MP, Gong JH, Loetcher P, Rajarathnam K, Amara A, Arenzana-Seisdedos F, Virelizier JL, Baggiolini M, Sykes BD & Clark-Lewis I (1997) Solution structure and basis for functional activity of stromal cell-derived factor-1: dissociation of CXCR4 activation from binding and inhibition of HIV-1. *EMBO J*. **16**: 6996-7007.

Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Dargan DJ, Thomson M, Griffiths PD, Wilkinson GW, Schulz TF & Davison AJ (2010) Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J Gen Virol*.

91: 605-615.

Cuomo L, Trivedi P, Cardillo MR, gagliardi FM, Vecchione A, Caruso R, Calogero A, Frati L, Faggioni A & Ragona G (2001) Human herpesvirus 6 infection in neoplastic and normal brain tissue. *J Med Virol.* **63**: 45-51.

Cusanelli E, Romero CA & Chartrand P (2013) Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol Cell.* **51**: 780-791.

Dagna L, Pritchett JC & Lusso P (2013) Immunomodulation and immunosuppression by human herpesvirus 6A and 6B. *Future Virol.* **8**: 273-287.

Daibata M, Taguchi T, Taguchi H & Miyoshi I (1998a) Integration of human herpesvirus 6 in a Burkitt's lymphoma cell line. *Br J Haematol.* **102**: 1307-1313.

Daibata M, Taguchi T, Nemoto Y, Taguchi H & Miyoshi I (1999b) Inheritance of chromosomally integrated human herpesvirus 6 DNA. *Blood.* **94**: 1545-1549.

Daibata M, Taguchi T, Sawada T, Taguchi H & Miyoshi I (1998c) Chromosomal transmission of human herpesvirus 6 DNA in acute lymphoblastic leukaemia. *Lancet.* **352**: 543-544.

Damania B & Cesarman E (2013) Kaposi's Sarcoma-Associated Herpesvirus. In Knipe D & Howley PM (Eds.) *Fields Virology* (6th edition), Lippincott, Williams and Wilkins, Philadelphia.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R & 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics.* **27**: 2156-2158.

Dargan DJ, Douglas E, Cunningham C, Jamieson F, Stanton RJ, Baluchova K, McSharry BP, Tomasec P, Emery VC, Percivalle E, Sarasini A, Gerna G, Wilkinson GW & Davison AJ (2010) Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J Gen Virol.* **91**: 1535-1546.

Davison AJ (2007) Comparative analysis of the genomes. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (Eds.) *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Cambridge: Cambridge University Press, pp. 10-26.

Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, Pellet PE, Roizman B, Studdert MJ, Thiry E (2009) The order Herpesvirales. *Arch Virol.* **154**: 171-177.

Daya S, Cortez N & Berns KI (2009) Adeno-associated virus site-specific integration is mediated by proteins of the nonhomologous end-joining pathway. *J Virol.* **83**: 11655-11664.

De Bolle L, Andrei G, Snoeck R, Zhang Y, Van Lommel A, Otto M, Bousseau A, Roy C, De Clerq E & Naesens L (2004) Potent, selective and cell-mediated inhibition of human herpesvirus 6 at an early stage of viral replication by the non-nucleoside compound

CMV423. *Biochem Pharmacol.* **67**: 325-336.

De Bolle L, Van Loon J, De Clercq E & Naesens L (2005a) Quantitative analysis of human herpesvirus 6 cell tropism. *J Med Virol.* **75**: 76-85.

De Bolle L, Naesens L & De Clercq E (2005b) Update on human herpesvirus 6 biology, clinical features, and therapy. *Clin Microbiol Rev.* **18**: 217-245.

De Clercq E, Naesens L, De Bolle L, Schols D, Zhang Y & Neyts J (2001) Antiviral agents active against human herpesviruses HHV-6, HHV-7 and HHV-8. *Rev Med Virol.* **11**: 381-395.

De Clercq E (2003) Clinical potential of the acyclic nucleoside phosphonates cidofovir, adefovir, and tenofovir in treatment of DNA virus and retrovirus infections. *Clin Microbiol Rev.* **16**: 569-596.

De Clercq E & Naesens L (2006) In search of effective anti-HHV-6 agents. *J Clin Virol.* **37**: S82-S86.

De Filippis L, Foglieni C, Silva S, Vescovi AL, Lusso P & Malnati MS (2006) Differentiated human neural stem cells: a new ex vivo model to study HHV-6 infection of the central nervous system. *J Clin Virol.* **37**: S27-32.

Deiss LP, Chou J & Frenkel (1986) Functional domains within the α sequence involved in the cleavage-packaging of herpes simplex virus DNA. *J Virol.* **59**: 605-618.

Delecluse HJ & Hammerschmidt W (1993) Status of Marek's disease virus in established lymphoma cell lines: herpesvirus integration is common. *J Virol.* **67**: 82-92.

Delecluse HJ, Schuller S & Hammerschmidt W (1993) Latent Marek's disease virus can be activated from its chromosomally integrated state in herpesvirus-transformed lymphoma cells. *EMBO J.* **12**: 3277-3286.

Deng H & Dewhurst S (1998) Functional identification and analysis of cis-acting sequences which mediate genome cleavage and packaging in human herpesvirus 6. *J Virol.* **72**: 320-329.

Depledge DP, Palser AL, Watson SJ, Lai IY, Gray ER, Grant P, Kanda RK, Leproust E, Kellam P & Breuer J (2011) Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One.* **6**: e27805.

Depledge DP, Gray ER, Kundu S, Cooray S, Poulsen A, Aaby P & Breuer J (2014) Evolution of cocirculating varicella-zoster virus genotypes during a chickenpox outbreak in Guinea-Bissau. *J Virol.* **88**: 13936-13946.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis SB, Gabriel SB, Altshuler D & Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* **43**: 491-498.

Desmetz C, Lin YL, Mettling C, Portalès P, Rabesandratana H, Clot J & Corbeau P (2006)

- The strength of the chemotactic response to a CCR5 binding chemokine is determined by the level of cell surface CCR5 density. *Immunology*. **119**: 551-561.
- Dewannieux M, Blaise S & Heidmann T (2005) Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses. *J Virol*. **79**: 15573-15577.
- Dewin DR, Catusse J & Gompels UA (2006) Identification and characterization of U83A viral chemokine, a broad and potent beta-chemokine agonist for human CCRs with unique selectivity and inhibition by spliced isoform. *J Immunol*. **176**: 544-556.
- Dewhurst S, McIntyre K, Schnabel K & Hall CB (1993) Human herpesvirus 6 (HHV-6) variant B accounts for the majority of symptomatic primary HHV-6 infections in a population of U.S. infants. *J Clin Microbiol*. **31**: 416-418.
- Dhepakson P, Mori Y, Jiang YB, Huang HL, Akkapaiboon P, Okuno T & Yamanishi K (2002) Human herpesvirus-6 rep/U94 gene product has single-stranded DNA-binding activity. *J Gen Virol*. **83**: 847-854.
- Dieffenbach CW, Lowe TM & Dveksler GS (1993) General concepts for PCR primer design. *PCR Methods Appl*. **3**: S30-S37.
- Di Luca D, Katsafanas G, Schirmer EC, Balachandran N & Frenkel N (1990) The replication of viral and cellular DNA in human herpesvirus 6-infected cells. *Virology*. **175**: 199-210.
- Di Luca D, Secchiero P, Bovenzi P, Rotola A, Caputo A, Monini P & Cassai E (1991) Reciprocal in vitro interactions between human herpesvirus-6 and HIV-1 Tat. *AIDS*. **5**: 1095-1098.
- Di Luca D, Dolcetti R, Mirandola P, De Re V, Secchiero P, Carbone A, Boiocchi M & Cassai E (1994) Human herpesvirus 6: a survey of presence and variant distribution in normal peripheral lymphocytes and lymphoproliferative disorders. *J Infect Dis*. **170**: 211-215.
- Di Luca D, Mirandola P, Ravaioli T, Dolcetti R, Frigatti A, Bovenzi P, Sighinolfi L, Monini P & Cassai E (1995) Human herpesvirus 6 and 7 in salivary glands and shedding in saliva of healthy and human immunodeficiency virus positive individuals. *J Med Virol*. **45**: 462-468.
- Di Luca D, Mirandola P, Ravaioli T, Bigoni B & Cassai E (1996) Distribution of HHV-6 variants in human tissues. *Infect Agents Dis*. **5**: 203-214.
- Dohner K, Wolfstein A, Prank U, Echeverri C, Dujardin D, Vallee R & Sodeik B (2002) Function of dynein and dynactin in herpes simplex virus capsid transport. *Mol Biol Cell*. **13**: 2795-2809.
- Dohner K & Sodeik B (2005) The role of the cytoskeleton during viral infection. *Curr Top Microbiol Immunol*. **285**: 67-108.
- Dolken L, Pfeffer S & Koszinowski UH (2009) Cytomegalovirus microRNAs. *Virus Genes*. **38**: 355-364.
- Dominguez G, Dambaugh TR, Stamey FR, Dewhurst S, Inoue N & Pellet PE (1999) Human

herpesvirus 6B genome sequencing: coding content and comparison with human herpesvirus 6A. *J Virol.* **73**: 8040-8052.

Donaldson CD, Clark DA, Kidd IM, Breuer J & Depledge DD (2013) Genome sequence of human herpesvirus 7 strain UCL-1. *Genome Announc.* **1**: e00830-13.

Donati D, Martinelli E, Cassiani-Ingoni R, Ahlgvist J, Hou J, Major EO & Jacobson S (2005) Variant-specific tropism of human herpesvirus 6 in human astrocytes. *J Virol.* **79**: 9439-9448.

Downes GB & Gautam N (1999) The G protein subunit gene families. *Genomics.* **62**: 544-552.

Downing RG, Sewankambo N, Serwadda D, Honess R, Crawford D, Jarrett R & Griffin BE (1987) Isolation of human lymphotropic herpesviruses from Uganda. *Lancet.* **2**: 390.

Drobyski WR, Eberle M, Majewski D & Baxter-Lowe LA (1993) Prevalence of human herpesvirus 6 variant A and B infections in bone marrow transplant recipients as determined by polymerase chain reaction and sequence-specific oligonucleotide probe hybridization. *J Clin Microbiol.* **31**: 1515-1520.

Drobyski WR, Knox KK, Majewski D & Carrigan DR (1994) Brief report: fatal encephalitis due to variant B human herpesvirus-6 infection in a bone marrow-transplant recipient. *N Engl J Med.* **330**: 1356-1360.

Droese J, Mokros T, Hermosilla R, Schulein R, Lipp M, Hopken UE & Rehm A (2004) HCMV-encoded chemokine receptor US28 employs multiple routes for internalisation. *Biochem Biophys Res Commun.* **322**: 42-49.

Duchesnes CE, Murphy PM, Williams TJ & Pease JE (2006) Alanine scanning mutagenesis of the chemokine receptor CCR3 reveals distinct extracellular residues involved in recognition of the eotaxin family of chemokines. *Mol Immunol.* **43**: 1221-1231.

Duh EJ, Maury WJ, Folks TM, Fauci AS & Rabson AB (1989) Tumor necrosis factor alpha activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF-kappa B sites in the long terminal repeat. *Proc Natl Acad Sci U S A.* **86**: 5974-5978.

Dulery R, Salleron J, Dewilde A, Rossignol J, Boyle EM, Gy J, de Berranger E, Coiteux V, Jouet JP, Duhamel A & Yakoub-Agha I (2012) Early human herpesvirus type 6 reactivation after allogenic stem cell transplantation: a large-scale clinical study. *Biol Blood Marrow Transplant.* **18**: 1080-1089.

Duma L, Haussinger D, Rogowski M, Lusso P & Grzesiek S (2007) Recognition of RANTES by extracellular parts of the CCR5 receptor. *J Mol Biol.* **365**: 1063-1075.

Dunn W, Chou C, Li H, Hai R, Patterson D, Stolc V, Zhu H & Liu F (2003) Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A.* **100**: 14223-14228.

Dutheil N, Smith SC, Agundez L, Vincent-Mistiaen ZI, Escalante CR, Linden RM &

- Henckaerts E (2014) Adeno-associated virus Rep represses the human integration site promoter by two pathways that are similar to those required for the regulation of the viral p5 promoter. *J Virol.* **88**: 8227-8241.
- Dwinell MB, Eckmann L, Leopard JD, Varki NM & Kagnoff MF (1999) Chemokine receptor expression by human intestinal epithelial cells. *Gastroenterology.* **117**: 359-367.
- Dykes C, Chan H, Krenitsky DM & Dewhurst S (1997) Stringent structural and sequence requirements of the human herpesvirus 6B lytic-phase origin of DNA replication. *J Gen Virol.* **78**: 1125-1129.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792-1797.
- Efferth T, Romero MR, Wolf DG, Stamminger T, Marin JJ & Marschall M (2008) The antiviral activities of artemisinin and artesunate. *Clin Infect Dis.* **47**: 804-811.
- Efstathiou S & Preston CM (2005) Towards an understanding of the molecular basis of herpes simplex virus latency. *Virus Res.* **111**: 108-119.
- Eisenberg RJ, Atanasiu D, Cairns TM, Gallagher JR, Krummenacher C & Cohen GH (2012) Herpes virus fusion and entry: a story with many characters. *Viruses.* **4**: 800-832.
- Ekblom R & Wolf JB (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl.* **7**: 1026-1042.
- El-Asmar L, Springael JY, Ballet S, Andrieu EU, Vassart G & Parmentier M (2005) Evidence for negative binding cooperativity within CCR5-CCR2b heterodimers. *Mol Pharmacol.* **67**: 460-469.
- Enders G, Biber M, Meyer G & Helftenbein E (1990) Prevalence of antibodies to human herpesvirus 6 in different age groups, in children with exanthem subitum, other acute exanthematous childhood diseases, Kawasaki syndrome, and acute infections with other herpesviruses and HIV. *Infection.* **18**: 12-15.
- Endo A, Watanabe K, Ohye T, Suzuki K, Matsubara T, Shimizu N, Kurahashi H, Yoshikawa T, Katano H, Inoue N, Imai N, Takagi M, Morio T & Mizutani S (2014) Molecular and virological evidence of viral activation from chromosomally integrated human herpesvirus 6A in a patient with X-linked severe combined immunodeficiency. *Clin Infect Dis.* **59**: 545-548.
- Ensoli B, Lusso P, Schachter F, Josephs SF, Rappaport J, Negro F, Gallo RC & Wong-Staal F (1989) Human herpesvirus-6 increases HIV-1 expression in co-infected T cells via nuclear factors binding to the HIV-1 enhancer. *EMBO J.* **8**: 3019-3027.
- Epperson ML, Lee CA & Fremont DH (2012) Subversion of cytokine networks by virally encoded decoy receptors. *Immunol Rev.* **250**: 199-215.
- Farr TJ, Harnett GB, Pietroboni GR & Bucens MR (1990) The distribution of antibodies to HHV-6 compared with other herpesviruses in young children. *Epidemiol Infect.* **105**: 603-07.

- Farzan M, Mirzabekov T, Kolchinsky P, Wyatt R, Cayabyab M, Gerard NP, Gerard C, Sodroski J & Choe H (1999) Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell*. **96**: 667-676.
- Farzan M, babcock GJ, Vasilieva N, Wright PL, Kiprilov E, Mirzabekov T & Choe H (2002) The role of post translational modifications of the CXCR4 amino terminus in stromal derived factor 1 alpha association and HIV-1 entry. *J Biol Chem*. **277**: 29484-29489.
- Faten N, Agnes GD, Nadia BF, Nabil AB, Monia Z, Abderrahim K, Henri A, Salma F & Mahjoub A (2012) Quantitative analysis of human herpesvirus-6 genome in blood and bone marrow samples from Tunisian patients with acute leukemia: a follow-up study. *Infect Agent Cancer*. **7**: 31.
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*. **17**: 368-376.
- Feng D, Chen J, Yue Y, Zhu H, Xue J & Jia WW (2006) A 16bp Rep binding element is sufficient for mediating Rep-dependent integration into AAVS1. *J Mol Biol*. **358**: 38-45.
- Feng H, Sun Z, Farzan MR & Feng P (2010) Sulfotyrosines of the Kaposi's sarcoma-associated herpesvirus G protein-coupled receptor promote tumorigenesis through autocrine activation. *J Virol*. **84**: 3351-3361.
- Feng H, Shuda M, Chang Y & Moore PS (2008) Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*. **319**: 1096-10100.
- Fernandez EJ & Lolis E (2002) Structure, function, and inhibition of chemokines. *Annu Rev Pharmacol Toxicol*. **42**: 469-499.
- Fitzsimons CP, Gompels UA, Verzijl D, Vischer HF, Mattick C, Leurs R & Smit MJ (2006) Chemokine-directed trafficking of receptor stimulus to different g proteins: selective inducible and constitutive signaling by human herpesvirus 6-encoded chemokine receptor U51. *Mol Pharmacol*. **69**: 888-898.
- Flamand L, Gosselin J, D'Addario M, Hiscott J, Ablashi DV, Gallo RC & Menezes J (1991) Human herpesvirus 6 induces interleukin-1 beta and tumor necrosis factor alpha, but not interleukin-6, in peripheral blood mononuclear cell cultures. *J Virol*. **65**: 5105-5110.
- Flamand L, Gosselin J, Stefanescu I, Ablashi D & Menezes J (1995) Immunosuppressive effect of human herpesvirus 6 on T-cell functions: suppression of interleukin-2 synthesis and cell proliferation. *Blood*. **85**: 1263-1271.
- Flamand L, Stefanescu I & Menezes J (1996) Human herpesvirus-6 enhances natural killer cell cytotoxicity via IL-15. *J Clin Invest*. **97**: 1373-1381.
- Flamand L, Romerio F, Reitz MS & Gallo RC (1998) CD4 promoter transactivation by human herpesvirus 6. *J Virol*. **72**: 8797-8805.
- Flamand L, Komaroff AL, Arbuckle JH, Medveczky PG & Ablashi DV (2010) Review, part 1: human herpesvirus-6 basic biology, diagnostic testing, and antiviral efficacy. *J Med Virol*.

82: 1560-1568.

Flynn G, Maru S, Loughlin J, Romero IA & Male D (2003) Regulation of chemokine receptor expression in human microglia and astrocytes. *J Neuroimmunol.* **136:** 84-93.

Fong AM, Alam SM, Imai T, Haribabu B & Patel DD (2002) CX3CR1 tyrosine sulfation enhances fractalkine-induced cell adhesion. *J Biol Chem.* **277:** 19418-19423.

Fox JD, Briggs M, Ward PA & Tedder RS (1990) Human herpesvirus 6 in salivary glands. *Lancet.* **336:** 590-593.

Fragnet L, Blasco MA, Klapper W & Rasschaert D (2003) The RNA subunit of telomerase is encoded by Marek's disease virus. *J Virol.* **77:** 5985-5996.

Fraile-Ramos A, Kledal TN, Pelchen-Matthews A, Bowers K, Schwartz TW & Marsh M (2001) The human cytomegalovirus US28 protein is located in endocytic vesicles and undergoes constitutive endocytosis and recycling. *Mol Biol Cell.* **12:** 1737-1749.

Frederick MJ & Clayman GL (2001) Chemokines in cancer. *Expert Rev Mol Med.* **3:** 1-18.

Fredriksson R, Lagerstrom MC, Lundin LG & Schioth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol.* **63:** 1256-1272.

French C, Menegazzi P, Nicholson L, Macaulay H, Di Luca D & Gompels UA (1999) Novel, nonconsensus cellular splicing regulates expression of a gene encoding a chemokine-like protein that shows high variation and is specific for human herpesvirus 6. *Virology.* **262:** 139-151.

Furukawa M, Yasukawa M, Yakushijin Y & Fujita S (1994) Distinct effects of human herpesvirus 6 and human herpesvirus 7 on surface molecule expression and function of CD4+ T cells. *J Immunol.* **152:** 5768-5775.

Gao JL & Murphy (1994) Human cytomegalovirus open reading frame US28 encodes a functional beta chemokine receptor. *J Biol Chem.* **269:** 28539-28542.

Garber DA, Beverley SM & Coen DM (1993) Demonstration of circularization of herpes simplex virus DNA following infection using pulsed field gel electrophoresis. *Virology.* **197:** 459-462.

Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, Dopazo J, Meyer TF, Conesa A (2012) Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics.* **28:** 2678-2679.

Garzino-Demo A, Chen M, Lusso P, Berneman Z & DiPaolo JA (1996) Enhancement of TAT-induced transactivation of the HIV-1 LTR by two genomic fragments of HHV-6. *J Med Virol.* **50:** 20-24.

Geoffroy MC & Salvetti A (2005) Helper functions required for wild type and recombinant adeno-associated virus growth. *Curr Gene Ther.* **5:** 265-271.

Geourjon C & Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* **11**: 681-684.

Giallonardo FD, Topfer A, Rey M, Prabhakaran S, Duport Y, Leemann C, Schmutz S, Campbell NK, Joos B, Lecca MR, Patrignani A, Daumer M, Beisel C, Rusert P, Trkola A, Günthard HF, Roth V, Beerewinkel N & Metzner KJ. (2014) Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. *Nucleic Acids Res.* **42**: e115.

Giesen K, Radsak K & Bogner E (2000a) The potential terminase subunit of human cytomegalovirus, pUL56, is translocated into the nucleus by its own nuclear localization signal and interacts with importin alpha. *J Gen Virol.* **81**: 2231-2244.

Giesen K, Radsak K & Bogner E (2000b) Targeting of the gene product encoded by ORF UL56 of human cytomegalovirus into viral replication centers. *FEBS Lett.* **471**: 215-218.

Giraud C, Winocour E & Berns KI (1994) Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. *Proc Natl Acad Sci U S A.* **91**: 10039-10043.

Glosson NL & Hudson AW (2007) Human herpesvirus-6A and -6B encode viral immunoevasins that downregulate class I MHC molecules. *Virology.* **365**: 125-135.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES & Nusbaum C (2011) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol.* **27**: 182-189.

Goel P, Tailor P, Chande AG, Basu A & Mukhopadhyaya R (2013) An infectious HHV-6B isolate from a healthy adult with chromosomally integrated virus and a reporter based relative viral titer assay. *Virus Res.* **173**: 280-285.

Goldner T, Hewlett G, Ettischer N, Ruebsamen-Schaeff H, Zimmermann H & Lischka P (2011) The novel anticytomegalovirus compound AIC246 (Letermovir) inhibits human cytomegalovirus replication through a specific antiviral mechanism that involves the viral terminase. *J Virol.* **85**: 10884-10893.

Gompels UA & Macaulay HA (1995) Characterisation of human telomeric repeat sequences from human herpesvirus 6 and relationship to replication. *J Gen Virol.* **76**: 451-458.

Gompels UA, Nicholas J, Lawrence G, Jones M, Thomson BJ, Martin ME, Efsthathiou S, Craxton M & Macaulay HA (1995) The DNA sequence of human herpesvirus-6: structure coding content, and genome evolution. *Virology.* **209**: 29-51.

Gompels UA, Larke N, Sanz-Ramos M, Bates M, Musonda K, Manno D, Siame J, Monze M, Filteau S & CIGNIS Study Group (2012) Human cytomegalovirus infant infection adversely affects growth and development in maternally HIV-exposed and unexposed infants in Zambia. *Clin Infect Dis.* **54**: 434-442.

- Gompels UA & Kasolo FC (2006) HHV-6 genome: similar and different. In: Kreuger G & Ablashi DV (Eds.) Perspectives in Medical Virology Volume 12. Human herpesvirus-6: general virology, epidemiology and clinical pathology. London: Elsevier, pp. 23-46.
- Gong JH & Clark-Lewis I (1995) Antagonists of monocyte chemoattractant protein 1 identified by modification of functionally critical NH₂-terminal residues. *J Exp Med.* **181**: 631-640.
- Goodman AD, Mock DJ, Powers JM, Baker JV, Blumberg BM (2003) Human herpesvirus 6 genome and antigen in acute multiple sclerosis lesions. *J Infect Dis.* **187**: 1365-1376.
- Goodrum F, Caviness K & Zagallo P (2012) Human cytomegalovirus persistence. *Cell Microbiol.* **14**: 644-655.
- Gozansky EK, Louis JM, Caffrey M & Clore GM (2005) Mapping the binding of the N-terminal extracellular tail of the CXCR4 receptor to stromal cell-derived factor-1alpha. *J Mol Biol.* **345**: 651-658.
- Graham GJ, Locati M, Mantovani A, Rot A & Thelen M (2012) The biochemistry and biology of the atypical chemokine receptors. *Immunol Lett.* **145**: 30-38.
- Gravel A, Ablashi D & Flamand L (2013a) Complete genome sequence of early passaged human herpesvirus 6A (GS strain) isolated from North America. *Genome Announc.* **1**: e00012-13.
- Gravel A, Hall CB & Flamand L (2013b) Sequence analysis of transplacentally acquired human herpesvirus 6 DNA is consistent with transmission of a chromosomally integrated reactivated virus. *J Infect Dis.* **207**: 1585-1589.
- Gravel A, Dubuc I, Morissette G, Sedlak RH, Jerome KR & Flamand L (2015) Inherited chromosomally integrated human herpesvirus 6 as a predisposing risk factor for the development of angina pectoris. *Proc Natl Acad Sci U S A.* pii: 201502741.
- Greco A, Fester N, Engel AT & Kaufer BB (2014) Role of the short telomeric repeat region in Marek's disease virus replication, genomic integration, and lymphomagenesis. *J Virol.* **88**: 14138-14147.
- Greenstone HL, Santoro F, Lusso P & Berger EA (2002) Human herpesvirus 6 and measles virus employ distinct CD46 domains for receptor function. *J Biol Chem.* **277**: 39112-39118.
- Griffiths PD, Ait-Khaled M, Bearcroft CP, Clark DA, Quaglia A, Davies SE, Burroughs AK, Rolles K, Kidd IM, Knight SN, Noibi SM, Cope AV, Phillips AN & Emery VC (1999) Human herpesvirus 6 and 7 as potential pathogens after liver transplant: prospective comparison with the effect of cytomegalovirus. *J Med Virol.* **59**: 496-501.
- Grinde B (2013) Herpesviruses: latency and reactivation – viral strategies and host response. *J Oral Microbiol.* **5**: doi 103402.
- Grivel JC, Ito Y, Faga G, Santoro F, Shaheen F, Malnati MS, Fitzgerald W, Lusso P & Margolis L (2001) Suppression of CCR5- but not CXCR4-tropic HIV-1 in lymphoid tissue by human

herpesvirus 6. *Nat Med.* **7**: 1232-1235.

Grivel JC, Santoro F, Chen S, Faga G, Malnati MS, Ito Y, Margolis L & Lusso P (2003) Pathogenic effects of human herpesvirus 6 in human lymphoid tissue ex vivo. *J Virol.* **77**: 8280-8289.

Grundy JE (1990) Virologic and pathogenic aspects of cytomegalovirus infection. *Rev Infect Dis.* **12**: S711-S719.

Guo H, Shen S, Wang L & Deng H (2010) Role of tegument proteins in herpesvirus assembly and egress. *Protein Cell.* **1**: 987-998.

Gupta SK, Lysko PG, Pillarisetti K, Ohlstein E & Stadel JM (1998) Chemokine receptors in human endothelial cells. Functional expression of CXCR4 and its transcriptional regulation by inflammatory cytokines. *J Biol Chem.* **273**: 4282-4287.

Haggani AA & Tilton JC (2013) Entry inhibitors and their use in the treatment of HIV-1 infection. *Antiviral Res.* **98**: 158-170.

Hakacova N, Klingel K, Kandolf R, Engdahl E, Fogdell-Hahn A & Higgins T (2013) First therapeutic use of Artesunate in treatment of human herpesvirus 6B myocarditis in a child. *J Clin Virol.* **57**: 157-160.

Hall CB, Long CE, Schnabel KC, Caserta MT, McIntyre KM, Costanzo MA, Knott A, Dewhurst S, Insel RA & Epstein LG (1994) Human herpesvirus-6 infection in children. A prospective study of complications and reactivation. *N Engl J Med.* **331**: 432-438.

Hall CB, Caserta MT, Schnabel KC, Long C, Epstein LG, Insel RA, Dewhurst S (1998) Persistence of human herpesvirus 6 according to site and variant: possible greater neurotropism of variant A. *Clin Infect Dis.* **26**: 132-137.

Hall CB, Caserta MT, Schnabel KC, Boettrich C, McDermott MP, Lofthus GK, Carnahan JA & Dewhurst S (2004) Congenital infections with human herpesvirus 6 (HHV6) and human herpesvirus 7 (HHV7). *J Pediatr.* **145**: 472-477.

Hall CB, Caserta MT, Schnabel KC, McDermott MP, Lofthus GK, Carnahan LM, Gilbert LM & Dewhurst S (2006) Characteristics and acquisition of human herpesvirus (HHV) 7 infection in relation to infection with HHV-6. *J Infect Dis.* **193**: 1063-1069.

Hall CB, Caserta MT, Schnabel KC, Shelley LM, Marino AS, Carnahan JA, Yoo C, Lofthus GK & McDermott MP (2008) Chromosomal integration of human herpesvirus 6 is the major mode of congenital human herpesvirus 6 infection. *Pediatrics.* **122**: 513-520.

Hall CB, Caserta MT, Schnabel KC, Shelley LM, Carnahan JA, Marino AS, Yoo C & Lofthus GK (2010) Transplacental congenital human herpesvirus 6 infection caused by maternally chromosomally integrated virus. *J Infect Dis.* **201**: 505-507.

Hamatake M, Aoki T, Futahashi Y, Urano E, Yamamoto N & Komano J (2009) Ligand-independent higher order multimerization of CXCR4, a G-protein-coupled chemokine receptor involved in targeted metastasis. *Cancer Sci.* **100**: 95-102.

- Hammond ME, Shyamala V, Siani MA, Gallegos CA, Feucht PH, Abbot J, Lapointe GR, Moghadam M, Khoja H, Zakel J & Temkamp-Olson P (1996) Receptor recognition and specificity of interleukin-8 is determined by residues that cluster near a surface accessible hydrophobic pocket. *J Biol Chem.* **271**: 8228-8235.
- Han KH, Green SR, Tangirala RK, Tanaka S & Quehenberger O (1999) Role of the first extracellular loop in the functional activation of CCR2. The first extracellular loop contains distinct domains necessary for both agonist binding and transmembrane signaling. *J Biol Chem.* **274**: 32055-32062.
- Handel TM & Domaille PJ (1996) Heteronuclear (1H, 13C, 15N) NMR assignments and solution structure of the monocyte chemoattractant protein-1 (MCP-1) dimer. *Biochemistry.* **35**: 6569-6584.
- Handel TM, Johnson Z, Crown SE, Lau EK & Proudfoot AE (2005) Regulation of protein function by glycosaminoglycans – as exemplified by chemokines. *Annu Rev Biochem.* **74**: 385-410.
- Handel TM, Johnson Z, Rodrigues DH, Dos Santos AC, Cirillo R, Muzio V, Riva S, Mack M, Deruaz M, Borlat F, Vitte PA, Wells TN, Teixeira MM & Proudfoot AE (2008) An engineered monomer of CCL2 has anti-inflammatory properties emphasizing the importance of oligomerization for chemokine activity in vivo. *J Leukocyte Biol.* **84**: 1101-1108.
- Harberts E, Yao K, Wohler JE, Maric D, Ohayon J, Henkin R & Jacobson S (2011) Human herpesvirus-6 entry into the central nervous system through the olfactory pathway. *Proc Natl Acad Sci U S A.* **108**: 13734-13739.
- Hebert CA, Chuntharapai A, Smith M, Colby T, Kim J & Horuk R (1993) Partial functional mapping of the human interleukin-8 type A receptor. Identification of a major ligand binding domain. *J Biol Chem.* **268**: 18549-18553.
- Heckman KL & Pease LR (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc.* **2**: 924-932.
- Heming JD, Huffman JB, Jones LM & Homa FL (2014) Isolation and characterization of the herpes simplex virus 1 terminase complex. *J Virol.* **88**: 225-236.
- Hemmerich S, Paavola C, Bloom A, Bhakta S, Freedman R, Grunberger D, Krstenansky J, Lee S, McCarley D, Mulkins M, Wong B, Pease J, Mizoue L, Mirzadegan T, Polsky I, Thompson K, Handel TM & Jarnagin K (1999) Identification of residues in the monocyte chemotactic protein-1 that contact the MCP-1 receptor, CCR2. *Biochemistry.* **38**: 13013-13025.
- Henckaerts E, Dutheil N, Zeltner N, Kattman S, Kohlbrenner E, Ward P, Clément N, Rebollo P, Kennedy M, Keller GM & Linden RM (2009) Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proc Natl Acad Sci U S A.* **106**: 7571-7576.
- Hendrix RW (2013) Bacteriophages. In Knipe D & Howley PM (Eds.) *Fields Virology* (6th

edition), Lippincott, Williams and Wilkins, Philadelphia.

Hernanz-Falcon P, Rodriguez-Frade JM, Serrano A, Juan D, del Sol A, Soriano SF, Roncal F, Gomez L, Valencia A, Martinez AC & Mellado M (2004) Identification of amino acid residues crucial for chemokine receptor dimerization. *Nat Immunol.* **5**: 216-223.

Hileman RE, Fromm JR, Weiler JM & Linhardt RJ (1998) Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays.* **20**: 156-167.

Hill JA, Sedlak RH, Zerr DM, Huang ML, Yeung C, Myerson D, Jerome KR & Boeckh MJ (2015) Prevalence of chromosomally integrated human herpesvirus 6 in patients with human herpesvirus 6 central nervous system dysfunction. *Biol Blood Marrow Transplant.* **21**: 371-373.

Hirata Y, Kondo K & Yamanishi K (2001) Human herpesvirus 6 downregulates major histocompatibility complex class I in dendritic cells. *J Med Virol.* **65**: 576-583.

Hjorto GM, Kiilerich-Pedersen K, Selmeczi D, Kledal TN & Larsen NB (2013) Human cytomegalovirus chemokine receptor US28 induces migration of cells on a CX3CL1-presenting surface. *J Gen Virol.* **94**: 1111-1120.

Homa FL & Brown JC (1997) Capsid assembly and DNA packaging in herpes simplex virus. *Rev Med Virol.* **7**: 107-122.

Honess RW & Roizman B (1974) Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J Virol.* **14**: 8-19.

Honess RW, Gompels UA, Barrell BG, Craxton M, Cameron KR, Staden R, Chang YN & Hayward GS (1989) Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. *J Gen Virol.* **70**: 837-855.

Hook L, Hancock M, Landais I, Grabski R, Britt W & Nelson JA (2014) Cytomegalovirus microRNAs. *Curr Opin Virol.* **7**: 40-46.

Horvat RT, Wood C & Balachandran N (1989) Transactivation of human immunodeficiency virus promoter by human herpesvirus 6. *J Virol.* **63**: 970-973.

Huang Y, Hidalgo-Bravo A, Zhang E, Cotton VE, Mendez-Bermudez A, Wig G, Medina-Calzada Z, Neumann R, Jeffreys AJ, Winney B, Wilson JF, Clark DA, Dyer MJ & Royle NJ (2014) Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable facilitating release of the viral genome from the chromosome. *Nucleic Acids Res.* **42**: 315-327.

Hubacek P, Muzikova K, Hrdlickova A, Cinek O, Hyncicova K, Hrstkova H, Sedlacek P & Stary J (2009) Prevalence of HHV-6 integrated chromosomally among children treated for acute lymphoblastic or myeloid leukemia in the Czech Republic. *J Med Virol.* **81**: 258-263.

Hubacek P, Hrdlickova A, Spacek M, Zajac M, Muzikova K, Sedlacek P & Cetkovsky P (2013)

- Prevalence of chromosomally integrated HHV-6 in patients with malignant disease and healthy donors in the Czech Republic. *Folia Microbiol (Praha)*. **58**: 87-90.
- Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I & Le Mercier P (2011) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res*. **39**: D576-D582.
- Humar A, Kumar D, Caliendo AM, Moussa G, Ashi-Sulaiman A, Levy G & Mazzulli T (2002) Clinical impact of human herpesvirus 6 infection after liver transplantation. *Transplantation*. **73**: 599-604.
- Hurley EA, Agger S, McNeil JA, Lawrence JB, Calender A, Lenoir G & Thorley-Lawson DA (1991) When Epstein-Barr virus persistently infects B-cell lines, it frequently integrates. *J Virol*. **65**: 1245-1254.
- Huser D, Gogol-Doring A, Lutter T, Weger S, Winter K, Hammer EM, Cathomen T, Reinert K & Heilbronn R (2010) Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome. *PLoS Pathog*. **8**: e1000985.
- Hutt-Fletcher LM (2007) Epstein-Barr virus entry. *J Virol*. **81**: 7825-7832.
- Iampietro M, Morissette G, Gravel A & Flamand L (2014) Inhibition of interleukin-2 gene expression by human herpesvirus 6B U54 tegument protein. *J Virol*. **88**: 12452-12463.
- Ichimi R, Jin-no T & Ito M (1999) Induction of apoptosis in cord blood lymphocytes by HHV-6. *J Med Virol*. **58**: 63-68.
- Im DS & Muzyczka N (1990) The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell*. **61**: 447-457.
- Inazawa N, Hori T, Yamamoto M, Hatakeyama N, Yuko Y, Nojima M, Yasui H, Suzuki N, Shimizu N & Tsutsumi H (2015) HHV-6 encephalitis may complicate the early phase after allogeneic hematopoietic stem cell transplantation; detection by qualitative multiplex PCR and subsequent quantitative real-time PCR. *J Med Virol*. doi: 10.1002/jmv.24340.
- Inoue Y, Yasukawa M & Fujita S (1997) Induction of T-cell apoptosis by human herpesvirus 6. *J Virol*. **71**: 3751-3759.
- Inoue A, Hasegawa H, Kohno M, Ito MR, Terada M, Imai T, Yoshie O, Nose M & Fujita S (2005) Antagonist of fractalkine (CX3CL1) delays the initiation and ameliorates the progression of lupus nephritis in MRL/lpr mice. *Arthritis Rheum*. **52**: 1522-1533.
- Irannejad R, Tomshine JC, Tomshine JR, Chevalier M, Mahoney JP, Steyaert J, Rasmussen SG, Sunahara RK, El-Samad H, Huang B & von Zastrow M (2013) Conformational biosensors reveal GPCR signalling from endosomes. *Nature*. **495**: 534-538.
- Isaacson E, Glaser CA, Forghani B, Amad Z, Wallace M, Armstrong RW, Exner MM & Schmid S (2005) Evidence of human herpesvirus 6 infection in 4 immunocompetent patients with encephalitis. *Clin Infect Dis*. **40**: 890-893.

- Isaacson MK & Compton T (2009) Human cytomegalovirus glycoprotein B is required for virus entry and cell-to-cell spread but not for virion attachment, assembly, or egress. *J Virol.* **83**: 3891-3903.
- Isegawa Y, Ping Z, Nakano K, Sugimoto N & Yamanishi K (1998) Human herpesvirus 6 open reading frame U12 encodes a functional beta-chemokine receptor. *J Virol.* **72**: 6104-6112.
- Isegawa Y, Mukai T, Nakano K, Kagawa M, Chen J, Mori Y, Sunagawa T, Kawanishi K, Sashihara J, Hata A, Zou P, Kosuge H & Yamanishi K (1999) Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B. *J Virol.* **73**: 8053-8063.
- Isomura H, Stinski MF, Murata T, Yamashita Y, Kanda T, Toyokuni S & Tsurumi T (2011) The human cytomegalovirus gene products essential for late viral gene expression assemble into prereplication complexes before viral DNA replication. *J Virol.* **85**: 6629-9944.
- Issafras H, Angers S, Bulenger S, Blanpain C, Parmentier M, Labbe-Jullie C, Bouvier M & Marullo S (2002) Constitutive agonist-independent CCR5 oligomerisation and antibody-mediated clustering occurring at physiological levels of receptors. *J Biol Chem.* **277**: 34666-34673.
- Jacob RJ, Morse LS & Roizman B (1979) Anatomy of herpes simplex virus DNA. XII. Accumulation of head-to-tail concatemers in nuclei of infected cells and their role in the generation of the four isomeric arrangements of viral DNA. *J Virol.* **29**: 448-457.
- Jarnagin K, Grunberger D, Mulkins M, Wong B, Hemmerich S, Paavola C, Bloom A, Bhakta S, Diehl F, Freedman R, McCarley D, Polsky I, Ping-Tsou A, Kosaka A & Handel TM (1999) Identification of surface residues of the monocyte chemotactic protein 1 that affect signaling through the receptor CCR2. *Biochemistry.* **38**: 16167-16177.
- Jarrett RF, Clark DA, Josephs SF & Onions DE (1990) Detection of human herpesvirus-6 DNA in peripheral blood and saliva. *J Med Virol.* **32**: 73-76.
- Jasirwan C, Furusawa Y, Tang H, Maeki T & Mori Y (2014) Human herpesvirus-6A gQ1 and gQ2 are critical for human CD46 usage. *Microbiol Immunol.* **58**: 22-30.
- Jasirwan C, Tang H, Kawabata A & Mori Y (2015) The human herpesvirus 6 U21-U24 gene cluster is dispensable for virus growth. *Microbiol Immunol.* **59**: 48-53.
- Jaworska J, Gravel A, Fink K, Grandvaux N & Flamand L (2007) Inhibition of transcription of the beta interferon gene by the human herpesvirus 6 immediate-early 1 protein. *J Virol.* **81**: 5737-5748.
- Jaworska J, Gravel A & Flamand L (2010) Divergent susceptibilities of human herpesvirus 6 variants to type I interferons. *Proc Natl Acad Sci U S A.* **107**: 8369-8374.
- Jefford CE & Irminger-Finger I (2006) Mechanisms of chromosome instability in cancers. *Crit Rev Oncol Hematol.* **59**: 1-14.
- Jenner RG, Alba MM, Boshoff C & Kellam P (2001) Kaposi's sarcoma-associated herpesvirus latent and lytic gene expression as revealed by DNA arrays. *J Virol.* **75**: 891-902.

- Jin F, Zheng K, Zhuo C, Ma K, Chen M, Wang Q, Zhang P, Fan J, Ren Z & Wang Y (2014) Silencing herpes simplex virus type 1 capsid protein encoding genes by siRNA: a promising antiviral therapeutic approach. *PLoS One*. **9**: e96623.
- Jones DT, Taylor WR & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. **8**: 275-282.
- Jones MD & Teo IA (1993) Identification and analysis of the transport/capsid assembly protein (tp/cap) gene of human herpesvirus-6 (HHV6). *Virology*. **197**: 449-454.
- Jones MD (1995) DNA sequence analysis of the origin of DNA replication region of human herpesvirus 6A strain U1102. Unpublished.
- Jordan NJ, Kolios G, Abbot SE, Sinai MA, Thompson DA, Petraki K & Westwick J (1999) Expression of functional CXCR4 chemokine receptors on human colonic epithelial cells. *J Clin Invest*. **104**: 1061-1069.
- Jordan BA & Devi LA (1999) G-protein-coupled receptor heterodimerization modulates receptor function. *Nature*. **399**: 697-700.
- Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H & Xu J (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. **7**: 1511-1522.
- Kasolo FC, Mpabalwani E & Gompels UA (1997) Infection with AIDS-related herpesvirus in human immunodeficiency virus-negative infants and endemic childhood Kaposi's sarcoma in Africa. *J Gen Virol*. **78**: 847-855.
- Kaspersen MD, Larsen PB, Kofod-Olsen E, Fedder J, Bonde J & Hollsberg P (2012) Human herpesvirus-6A/B binds to spermatozoa acrosome and is the most prevalent herpesvirus in semen from sperm donors. *PLoS One*. **7**: e48810.
- Katancik JA, Sharma A & de Nardin E (2000) Interleukin 8, neutrophil-activating peptide-2 and GRO-alpha bind to and elicit cell activation via specific and different amino acid residues of CXCR2. *Cytokine*. **12**: 1480-1488.
- Katrich V, Cherezov V & Stevens RC (2013) Structure-function of the G protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol*. **53**: 531-556.
- Katsafanas GC, Schirmer EC, Wyatt LS & Frenkel N (1996) In vitro activation of human herpesvirus 6 and 7 from latency. *Proc Natl Acad Sci U S A*. **93**: 9788-9792.
- Kaufer BB, Jarosinski KW & Osterrieder N (2011) Herpesvirus telomeric repeats facilitate genomic integration into host telomeres and mobilization of viral DNA during reactivation. *J Exp Med*. **208**: 605-615.
- Kawabata A, Tang H, Huang H, Yamanishi K, Mori Y (2009) Human herpesvirus 6 envelope components enriched in lipid rafts: evidence for virion-associated lipid rafts. *Virology*. **6**: 127-133.

- Kawabata A, Oyaizu H, Maeki T, Tang H, Yamanishi K & Mori Y (2011) Analysis of a neutralising antibody for human herpesvirus 6B reveals a role for glycoprotein Q1 in viral entry. *J Virol.* **85**: 12962-12971.
- Kelly E, Bailey CP & Henderson G (2008) Agonist-selective mechanisms of GPCR desensitization. *Br J Pharmacol.* **153**: S379-S388.
- Kenneson A & Cannon MJ (2007) Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. *Rev Med Virol.* **17**: 253-276.
- Khurram SA, Whawell SA, Bingle L, Murdoch C, McCabe BM & Farthing PM (2010) Functional expression of the chemokine receptor XCR1 on oral epithelial cells. *J Pathol.* **221**: 153-163.
- Kikuta H, Nakane A, Lu H, Taguchi Y, Minagawa T & Matsumoto S (1990) Interferon induction by human herpesvirus 6 in human mononuclear cells. *J Infect Dis.* **162**: 35-38.
- King AMQ, Adams MJ, Carstens EB & Lefkowitz EJ (Eds.) (2012) Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic Press.
- Kishi M, Bradley G, Jessip J, Tanaka A & Nonoyama M (1991) Inverted repeat regions of Marek's disease virus DNA possess a structure similar to that of the a sequence of herpes simplex virus DNA and contain host cell telomere sequences. *J Virol.* **65**: 2791-2797.
- Klein E, Ben-Bassat H, Neumann H, Ralph P, Zeuthen J, Polliack A & Vanky F (1976) Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia. *Int J Cancer.* **18**: 421-431.
- Knox KK & Carrigan DR (1994) Disseminated active HHV-6 infections in patients with AIDS. *Lancet.* **343**: 577-578.
- Koboldt DC, Steinberg KM, Larson DE, Wilson RK & Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell.* **155**: 27-38.
- Kokkoli E, Kasinskas RW, Mardilovich A & Garg A (2005) Fractalkine targeting with a receptor-mimicking peptide-amphiphile. *Biomacromolecules.* **6**: 1272-1279.
- Kondo K, Kondo T, Okuno T, Takahashi M & Yamanishi K (1991) Latent human herpesvirus 6 infection of human monocytes/macrophages. *J Gen Virol.* **72**: 1401-1408.
- Kondo K, Shimada K, Sashihara J, Tanaka-Taya K & Yamanishi K (2002) Identification of human herpesvirus 6 latency-associated transcripts. *J Virol.* **76**: 4145-4151.
- Kondo K, Sashihara J, Shimada K, Takemoto M, Amo K, Miyagawa H & Yamanishi K (2003) Recognition of a novel stage of betaherpesvirus latency in human herpesvirus 6. *J Virol.* **77**: 2258-2264.
- Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, McLaughlin S,

- Muzyczka N, Rocchi M & Berns KI (1990) Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci U S A.* **87**: 2211-2215.
- Kremlev SG, Roberts RL & Palmer C (2004) Differential expression of chemokines and chemokine receptors during microglial activation and inhibition. *J Neuroimmunol.* **149**: 1-9.
- Kripalani-Joshi S & Law HY (1994) Identification of integrated Epstein-Barr virus in nasopharyngeal carcinoma using pulse field gel electrophoresis. *Int J Cancer.* **56**: 187-192.
- Krosky PM, Underwood MR, Turk SR, Feng KW, Jain RK, Ptak RG, Westerman AC, Biron KK, Townsend LB & Drach JC (1998) Resistance of human cytomegalovirus to benzimidazole ribonucleosides maps to two open reading frames: UL89 and UL56. *J Virol.* **72**: 4721-4728.
- Krug LT, Inoue N & Pellet PE (2001) Differences in DNA binding specificity among Roseolovirus origin binding proteins. *Virology.* **288**: 145-153.
- Krummenacher C, Carfi A, Eisenberg RJ & Cohen GH (2013) Entry of herpesviruses into cells: the enigma variations. *Adv Exp Med Biol.* **790**: 178-195.
- Kuhl U, Pauschinger M, Seeberg B, Lassner D, Noutsias M, Poller W & Schultheiss HP (2005) Viral persistence in the myocardium is associated with progressive cardiac dysfunction. *Circulation.* **112**: 1965-1970.
- Kuhl U, Lassner D, Wallaschek N, Gross UM, Krueger GR, Seeberg B, Kaufer BB, Escher F, Poller W & Schultheiss HP (2015) Chromosomally integrated human herpesvirus 6 in heart failure: prevalence and treatment. *Eur J Heart Fail.* **17**: 9-19.
- Kuhn DE, Beall CJ & Kolattukudy PE (1995) The cytomegalovirus US28 protein binds multiple CC chemokines with high affinity. *Biochem Biophys Res Commun.* **211**: 325-330.
- Kurth R & Bannert N (2010) Beneficial and detrimental effects of human endogenous retroviruses. *Int J Cancer.* **126**: 306-314.
- Kuschert GS, Coulin F, Power CA, Proudfoot Aem Hubbard RE, Hoogewerf AJ & Wells TN (1999) Glycosaminoglycans interact selectively with chemokines and modulate receptor binding and cellular responses. *Biochemistry.* **38**: 12959-12968.
- Lamartina S, Ciliberto G & Toniatti C (2000) Selective cleavage of AAVS1 substrates by the adeno-associated virus type 2 rep68 protein is dependent on topological and sequence constraints. *J Virol.* **74**: 8831-8842.
- Lander ES et al. (2001) Initial sequencing and analysis of the human genome. *Nature.* **409**: 860-921.
- Langemeijer EV, Slinger E, de Munnik S, Schreiber A, Maussang D, Vischer H, Verkaar F, Leurs R, Siderius M and Smit MJ (2012) Constitutive β -catenin signaling by the viral chemokine receptor US28. *PLoS One.* **7**: e48925.
- Lansdorp PM, Verwoerd NP, van de Rijke FM, Dragowska V, Little MT, Dirks RW, Raap AK & Tanke HJ (1996) Heterogeneity in telomere length of human chromosomes. *Hum Mol*

Genet. **5**: 685-691.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ & Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics.* **23**: 2947-2948.

LaRosa GJ, Thomas KM, Kaufmann ME, Mark R, White M, Taylor L, Gray G, Witt D & Navarro J (1992) Amino terminus of the interleukin-8 receptor is a major determinant of receptor subtype specificity. *J Biol Chem.* **267**: 25402-25406.

Latek D, Modzelewska A, Trzaskowski B, Palczewski K & Filipek S (2012) G protein-coupled receptors – recent advances. *Acta Biochim Pol.* **59**: 515-529.

Lau EK, Paavola CD, Johnson Z, Gaudry JP, Geretti E, Borlat F, Kungl AJ, Proudfoot AE, Handel TM (2004) Identification of the glycosaminoglycan binding site of the CC chemokine, MCP-1: implications for structure and function in vivo. *J Biol Chem.* **279**: 22294-22305.

Laurence JS, Blanpain C, Burgner JW, Parmentier M & LiWang PJ (2000) CC chemokine MIP-1 beta can function as a monomer and depends on Phe13 for receptor binding. *Biochemistry.* **39**: 3401-3409.

Leach CT, Newton ER, McParlin S, Jenson HB (1994) Human herpesvirus 6 infection of the female genital tract. *J Infect Dis.* **169**: 1281-1283.

Lederberg EM & Lederberg J (1953) Genetic Studies of Lysogenicity in Escherichia Coli. *Genetics.* **38**: 51-64.

Lee H & Schatz MC (2012) Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. *Bioinformatics.* **28**: 2097-2105.

Lee SO, Brown RA & Razonable RR (2012) Chromosomally integrated human herpesvirus-6 in transplant recipients. *Transpl Infect Dis.* **14**: 346-354.

Leibovitch E, Wohler JE, Cummings Macri SM, Montanic K, Harberts E, Gaitan MI, Maggi P, Ellis M, Westmoreland S, Silva A, Reich DS & Jacobson S (2013) Novel marmoset (*Callithrix jacchus*) model of human herpesvirus 6A and 6B infections: immunologic, virologic and radiologic characterization. *PLoS Pathog.* **9**: e1003138.

Leibovitch EC, Brunetto GS, Caruso B, Fenton K, Ohayon J, Reich DS & Jacobson S (2014) Coinfection of human herpesvirus 6A and HHV-6B as demonstrated by novel digital droplet PCR assay. *PLoS One.* **9**: e92328.

Leong HN, Tuke PW, Tedder RS, Khanom AB, Eglin RP, Atkinson CE, Ward KN, Griffiths PD & Clark DA (2007) The prevalence of chromosomally integrated human herpesvirus 6 genomes in the blood of UK blood donors. *J Med Virol.* **79**: 45-51.

Leveque N, Boulagnon C, Brasselet C, Lesaffre F, Boutolleau D, Metz D, Fornes P & Andreoletti L (2011) A fatal case of human herpesvirus 6 chronic myocarditis in an immunocompetent adult. *J Clin Virol.* **52**: 142-145.

- Levoye A, Balabanian K, Baleux F, Bachelier F & Lagane B (2009) CXCR7 heterodimerizes with CXCR4 and regulates CXCL12-mediated G protein signaling. *Blood*. **113**: 6085-6093.
- Li WH, Gu Z, Wang H & Nekrutenko A (2001) Evolutionary analyses of the human genome. *Nature*. **409**: 847-849.
- Li M, Mizuuchi M, Burke TR & Craigie R (2006) Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J*. **25**: 1295-1304.
- Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**: 1754-1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*. **25**: 2078-2079.
- Lin Z, Wang X, Strong MJ, Concha M, Baddoo M, Xu G, Baribault C, Fewell C, Hulme W, Hedges D, Taylor CM & Flemington EK (2013) Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. *J Virol*. **87**: 1172-1182.
- Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM & Trask BJ (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature*. **437**: 94-100.
- Linden RM, Winocour E & Berns KI (1996) The recombination signals for adeno-associated virus site-specific integration. *Proc Natl Acad Sci U S A*. **93**: 7966-7972.
- Lindquister GJ & Pellet PE (1991) Properties of the human herpesvirus 6 strain Z29 genome: G+C contents, length, and presence of variable-length directly repeated terminal sequence elements. *Virology*. **182**: 102-110.
- Liou JW, Chang FT, Chung Y, Chen WY, Fischer WB & Hsu HJ (2014) In silico analysis reveals sequential interactions and protein conformational changes during the binding of chemokine CXCL-8 to its receptor CXCR1. *PLoS One*. **9**: e94178.
- Liston A, Kohler RE, Townley S, Haylock-Jacobs S, Comerford I, Caon AC, Webster J, Harrison JM, Swann J, Clark-Lewis I, Korner H & McColl SR (2009) Inhibition of CCR6 function reduces the severity of experimental autoimmune encephalomyelitis via effects on the priming phase of the immune response. *J Immunol*. **182**: 3121-3130.
- Liszewski MK, Post TW & Atkinson JP (1991) Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annu Rev Immunol*. **9**: 431-455.
- Littler E, Lawrence G, Liu MY, Barrell BG & Arrand JR (1990) Identification, cloning, and expression of the major capsid protein gene of human herpesvirus 6. *J Virol*. **64**: 714-722.
- Liu DX, Gompels UA, Foa-Tomasi L & Campadelli-Fiume G (1993) Human herpesvirus-6 glycoprotein H and L homologs are components of the gp100 complex and the gH

- external domain is the target for neutralizing monoclonal antibodies. *Virology*. **197**: 12-22.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L & Law M (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. **2012**: 251364.
- Ljungman P, Wang FZ, Clark DA, Emery VC, Remberger M, Ringden O & Linde A (2000) High levels of human herpesvirus 6 DNA in peripheral blood leucocytes are correlated to platelet engraftment and disease in allogenic stem cell transplant patients. *Br J Haematol*. **111**: 774-781.
- Ljungman P & Singh N (2006) Human herpesvirus-6 infection in solid organ and stem cell transplant recipients. *J Clin Virol*. **37**: S87-S91.
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J & Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. **30**: 434-439.
- Longnecker R, Kieff E & Cohen JI (2013) Epstein-Barr Virus. In Knipe D & Howley PM (Eds.) *Fields Virology* (6th edition), Lippincott, Williams and Wilkins, Philadelphia.
- Lopez C, Pellet P, Stewart J, Goldsmith C, Sanderlin K, Black J, Warfield D & Feorino P (1988) Characteristics of human herpesvirus-6. *J Infect Dis*. **157**: 1271-1273.
- Lopez R, Silventoinen V, Robinson S, Kibria A and Gish W (2003) WU-BLAST2 server at the European Bioinformatics Institute. *Nucleic Acids Res*. **31**: 3795-3798.
- Lou J, Wu Y, Cai M, Wu X & Shang S (2011) Subtype-specific, probe-based, real-time PCR for detection and typing of human herpesvirus-6 encephalitis from pediatric patients under the age of 2 years. *Diagn Microbiol Infect Dis*. **70**: 223-229.
- Love M, Sandberg JL, Ziarek JJ, Gerarden KP, Rode RR, Jensen DR, McCaslin DR, Peterson FC & Veldkamp CT (2012) Solution structure of CCL21 and identification of a putative CCR7 binding site. *Biochemistry*. **51**: 733-735.
- Lowman HB, Slagle PH, DeForge LE, Wirth CM, Gillece-Castro BL, Bourell JH & Fairbrother WJ (1996) Exchanging interleukin-8 and melanoma growth-stimulating activity receptor binding specificities. *J Biol Chem*. **271**: 14344-14352.
- Lubkowski J, Bujacz G, Boque L, Domaille PJ, Handel TM & Wlodawer A (1997) The structure of MCP-1 in two crystal forms provides a rare example of variable quaternary interactions. *Nat Struct Biol*. **4**: 64-69.
- Ludwig A, Ehlert JE, Flad HD & Brandt E (2000) Identification of distinct surface-expressed and intracellular CXC-chemokine receptor 2 glycoforms in neutrophils: N-glycosylation is essential for maintenance of receptor surface expression. *J Immunol*. **165**: 1044-1052.
- Luppi M, Marasca R, Barozzi P, Ferrari S, Ceccherini-Nelli L, Batoni G, Merelli E & Torelli G (1993) Three cases of human herpesvirus-6 latent infection: integration of viral genome in peripheral blood mononuclear cell DNA. *J Med Virol*. **40**: 44-52.

- Luppi M, Barozzi P, Garber R, Maiorana A, Bonacorsi G, Artusi T, Trovato R, Marasca R & Torelli G (1998a) Expression of human herpesvirus-6 antigen in benign and malignant lymphoproliferative diseases. *Am J Pathol.* **153**: 815-823.
- Luppi M, Barozzi P, Morris CM, Merelli E & Torelli G (1998b) Integration of human herpesvirus 6 genome in human chromosomes. *Lancet.* **352**: 1707-1708.
- Luppi M, Barozzi P, Morris C, Maiorana A, Garber R, Bonacorsi G, Donelli A, Marasca R, Tabilio A & Torelli G (1999) Human herpesvirus 6 latently infects early bone marrow progenitor in vivo. *J Virol.* **73**: 754-759.
- Lusso P, Markham PD, Tschachler E, di Marzo Veronesi F, Salahuddin SZ, Ablashi DV, Pahwa S, Krohn K & Gallo RC (1988) In vitro cellular tropism of human B-lymphotropic virus (human herpesvirus-6). *J Exp Med.* **167**: 1659-1670.
- Lusso P, Ensoli B, Markham PD, Ablashi DV, Salahuddin SZ, Tschachler E, Wong-Staal F & Gallo RC (1989) Productive dual infection of human CD4+ T lymphocytes by HIV-1 and HHV-6. *Nature.* **337**: 370-373.
- Lusso P, De Maria A, Malnati M, Lori F, DeRocco SE, Baseler M & Gallo RC (1991) Induction of CD4 and susceptibility to HIV-1 infection in human CD8+ T lymphocytes by human herpesvirus 6. *Nature.* **349**: 533-535.
- Lusso P, Malnati MS, Garzino-Demo A, Crowley RW, Long EO & Gallo RC (1993) Infection of natural killer cells by human herpesvirus 6. *Nature.* **362**: 458-462.
- Lusso P, Garzino-Demo A, Crowley RW & Malnati MS (1995) Infection of gamma/delta T lymphocytes by human herpesvirus 6: transcriptional induction of CD4 and susceptibility to HIV infection. *J Exp Med.* **181**: 1303-1310.
- Lusso P, Crowley RW, Malnati MS, Di Serio C, Ponzoni M, Biancotto A, Markham PD & Gallo RC (2007) Human herpesvirus 6A accelerates AIDS progression in macaques. *Proc Natl Acad Sci U S A.* **104**: 5067-5072.
- Luttichau HR, Clark-Lewis I, Jensen PO, Moser C, Gerstoft J & Schwartz TW (2003) A highly selective CCR2 chemokine agonist encoded by human herpesvirus 6. *J Biol Chem.* **278**: 10928-10933.
- Ma Y, Wang N, Li M, Gao S, Wang L, Ji Y, Qi Y, He R, Sun Z & Ruan Q (2011) An antisense transcript in the human cytomegalovirus UL87 gene region. *Virology.* **418**: 515.
- Mackay CR (2001) Chemokines: immunology's high impact factors. *Nat Immunol.* **2**: 95-101.
- MacLean HJ & Douen AG (2002) Severe amnesia associated with human herpesvirus 6 encephalitis after bone marrow transplantation. *Transplantation.* **73**: 1086-1089.
- Maicher A, Kastner L, Dees M & Luke B (2012) Deregulated telomere transcription causes replication-dependent telomere shortening and promotes cellular senescence. *Nucleic Acids Res.* **40**: 6649-6659.

- Majka J & Speck C (2007) Analysis of protein-DNA interactions using surface plasmon resonance. *Adv Biochem Eng Biotechnol.* **104**: 13-36.
- Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J & Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods.* **7**: 111-118.
- Manchester KL (1996) Use of UV methods for measurement of protein and nucleic acid concentrations. *Biotechniques.* **20**: 968-970.
- Manichanh C, Grenot P, Gautheret-Dejean A, Debre P, Huraux JM & Agut H (2000) Susceptibility of human herpesvirus 6 to antiviral compounds by flow cytometry analysis. *Cytometry.* **40**: 135-140.
- Mantovani A, Bonecchi R & Locati M (2006) Tuning inflammation and immunity by chemokine sequestration: decoys and more. *Nat Rev Immunol.* **6**: 907-918.
- Mardis ER (2013) Next-generation sequencing platforms. *Annu Rev Anal Chem.* **6**: 287-303.
- Margulies BJ & Gibson W (2007) The chemokine receptor homologue encoded by US27 of human cytomegalovirus is heavily glycosylated and is present in infected human foreskin fibroblasts and enveloped virus particles. *Virus Res.* **123**: 57-71.
- Marsh M, Tu O, Dolnik V, Roach D, Solomon N, Bechtol K, Smietana P, Wang L, Li X, Cartwright P, Marks A, Barker D, Harris D & Bashkin J (1997) High-throughput DNA sequencing on a capillary array electrophoresis system. *J Capillary Electrophor.* **4**: 83-89.
- Martens UM, Zijlmans JM, Poon SS, Dragowska W, Yui J, Chavez EA, Ward RK & Lansdorp PM (1998) Short telomeres on human chromosome 17p. *Nat Genet.* **18**: 76-80.
- Martin ME, Thomason BJ, Honess RW, Craxton MA, Gompels UA, Liu MY, Littler E, Arrand JR, Teo I & Jones MD (1991a) The genome of human herpesvirus 6: maps of unit-length and concatemeric genomes for nine restriction endonucleases. *J Gen Virol.* **72**: 157-168.
- Martin ME, Nicholas J, Thomson BJ, Newman C & Honess RW (1991b) Identification of a transactivating function mapping to the putative immediate-early locus of human herpesvirus 6. *J Virol.* **65**: 5381-5390.
- Matloubian M, David A, Engel S, Ryan JE & Cyster JG (2000) A transmembrane CXC chemokine is a ligand for HIV-coreceptor Bonzo. *Nat Immunol.* **1**: 298-304.
- Matranga CB, Andersen KG, Winnicki S, Busby M, Gladden AD, Tewhey R, Stremlau M, Berlin A, Gire SK, England E, Moses LM, Mikkelsen TS, Odia I, Ehiane PE, Folarin O, Goba A, Kahn SH, Grant DS, Honko A, Hensley L, Happi C, Garry RF, Malboeuf CM, Birren BW, Gnirke A, Levin JZ & Sabeti PC (2014) Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.* **15**: 519-531.
- Maussang D, Verzijl D, van Walsum M, Leurs R, Holl J, Pleskoff O, Michel D, van Dongen GA & Smit MJ (2006) Human cytomegalovirus-encoded chemokine receptors US28 promotes tumorigenesis. *Proc Natl Acad Sci U S A.* **103**: 13068-13073.

- Maussang D, Langemeijer E, Fitzsimons CP, Stigter-van Walsum M, Dijkman R, Borg MK, Slinger E, Schreiber A, Michel D, Tensen CP, van Dongen GA, Leurs R and Smit MJ (2009) The human cytomegalovirus-encoded chemokine receptor US28 promotes angiogenesis and tumour formation via cyclooxygenase-2. *Cancer Res.* **69**: 2861-2869.
- Mayer MR & Stone MJ (2001) Identification of receptor binding and activation determinants in the N-terminal and N-loop of the CC chemokine eotaxin. *J Biol Chem.* **276**: 13911-13916.
- McCarty DM, Ryan JH, Zolotukhin S, Zhou X & Muzyczka N (1994) Interaction of the adeno-associated virus Rep protein with a sequence within the A palindrome of the viral terminal repeat. *J Virol.* **68**: 4998-5006.
- McCarthy M, Auger D, He J & Wood C (1998) Cytomegalovirus and human herpesvirus-6 trans-activate the HIV-1 long terminal repeat via multiple response regions in human fetal astrocytes. *J Neurovirol.* **4**: 495-511.
- McGeoch DJ, Rixon FJ & Davison AJ (2006) Topics in herpesvirus genomics and evolution. *Virus Res.* **117**: 90-104.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M & DePristo MA (2010) The Genome Analysis Toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297-1303.
- McQuibban GA, Butler GS, Gong JH, Bendall L, Power C, Clark-Lewis I & Overall CM (2001) Matrix metalloproteinase activity inactivates the CXC chemokine stromal cell-derived factor-1. *J Biol Chem.* **276**: 43503-43508.
- McQuibban GA, Gong JH, Wong JP, Wallace JL, Clark-Lewis I & Overall CM (2002) Matrix metalloproteinase processing of monocyte chemoattractant proteins generates CC chemokine receptor antagonists with anti-inflammatory properties in vivo. *Blood.* **100**: 1160-1167.
- McSharry BP, Avdic S & Slobedman B (2013) Human cytomegalovirus encoded homologs of cytokines, chemokines and their receptors: roles in immunomodulation. *Viruses.* **4**: 2448-2470.
- Mefford HC & Trask BJ (2002) The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet.* **3**: 91-102.
- Megaw AG, Rapaport D, Avidor B, Frenkel N & Davison AJ (1998) The DNA sequence of the RK strain of human herpesvirus 7. *Virology.* **244**: 119-132.
- Mehrle S, Rohde V & Schlehofer JR (2004) Evidence of chromosomal integration of AAV DNA in human testis tissue. *Virus Genes.* **28**: 61-69.
- Melnychuk RM, Streblow DN, Smith PP, Hirsch AJ, Pancheva D & Nelson JA (2004) Human cytomegalovirus-encoded G protein-coupled receptor US28 mediates smooth muscle cell

migration through Galpha12. *J Virol.* **78**: 8382-8391.

Menotti L, Mirandola P, Locati M and Campadelli-Fiume G (1999) Trafficking to the plasma membrane of the seven-transmembrane protein encoded by human herpesvirus 6 U51 gene involves a cell-specific function present in T lymphocytes. *J Virol.* **73**: 325-333.

Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H & Brookes AJ (2011) Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics.* **10**: 374-386.

Mettenleiter TC (2002) Herpesvirus assembly and egress. *J Virol.* **76**: 1537-1547.

Mettenleiter TC, Klupp BG & Granzow H (2009) Herpesvirus assembly: an update. *Virus Res.* **143**: 222-234.

Mewborn SK, Lese Martin C & Ledbetter DH (2005) The dynamic nature and evolutionary history of subtelomeric and pericentromeric regions. *Cytogenet Genome Res.* **108**: 22-25.

Middleton J, Patterson AM, Gardner L, Schmutz C & Ashton BA (2002) Leukocyte extravasation: chemokine transport and presentation by the endothelium. *Blood.* **100**: 3853-3860.

Milbradt J, Auerochs S, Korn K & Marschall M (2009) Sensitivity of human herpesvirus 6 and other human herpesviruses to the broad spectrum antiinfective drug artesunate. *J Clin Virol.* **46**: 24-28.

Miller WE, Zagorski WA, Brenneman JD, Avery D, Miller JL & O'Connor CM (2012) US28 is a potent activator of phospholipase C during HCMV infection of clinically relevant target cells. *PloS One.* **7**: e50524.

Milne RS, Mattick C, Nicholson L, Devaraj P, Alcami A & Gompels UA (2000) RANTES binding and down-regulation by a novel human herpesvirus-6 beta chemokine receptor. *J Immunol.* **164**: 2396-2404.

Minisini R, Tulone C, Luske A, Michel D, Mertens T, Gierschik P & Moepps B (2003) Constitutive inositol phosphate formation in cytomegalovirus-infected human fibroblasts is due to expression of the chemokine receptor homologue pUS28. *J Virol.* **77**: 4489-4501.

Minjolle S, Michelet C, Jusselin I, Joannes M, Cartier F & Colimon R (1999) Amplification of the six major human herpesviruses from cerebrospinal fluid by a single PCR. *J Clin Microbiol.* **37**: 950-953.

Mirzadegan T, Benko G, Filipek S & Palczewski K (2003) Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry.* **42**: 2759-2767.

Miyake F, Yoshikawa T, Sun H, Kakimi A, Ohashi M, Akimoto S, Nishiyama Y & Asano Y (2006) Latent infection of human herpesvirus 7 in CD4(+) T lymphocytes. *J Med Virol.* **78**: 112-116.

Mizoue LS, Bazan JF, Johnson EC & Handel TM (1999) Solution structure and dynamics of

the CX3C chemokine domain of fractalkine and its interaction with an N-terminal fragment of CX3CR1. *Biochemistry*. **38**: 1402-1414.

Monteclaro FS & Charo IF (1996) The amino-terminal extracellular domain of the MCP-1 receptor, but not the RANTES/MIP-1alpha receptor, confers chemokine selectivity. Evidence for a two-step mechanism for MCP-1 receptor activation. *J Biol Chem*. **271**: 19084-19092.

Monteclaro FS & Charo IF (1997) The amino-terminal domain of CCR2 is both necessary and sufficient for high affinity binding of monocyte chemoattractant protein 1. Receptor activation by a pseudo-tethered ligand. *J Biol Chem*. **272**: 23186-23190.

Montejo M, Ramon Fernandez J, Testillano M, Valdivieso A, Aguirrebengoa K, Varas C, Olaizola A & De Urbina JO (2002) Encephalitis caused by human herpesvirus-6 in a liver transplant recipient. *Eur Neurol*. **48**: 234-235.

Montoya JG, Neely MN, Gupta S, Lunn MR, Loomis KS, Pritchett JC, Polsky B & Medveczky PG (2012) Antiviral therapy of two patients with chromosomally-integrated human herpesvirus-6A presenting with cognitive dysfunction. *J Clin Virol*. **55**: 40-45.

Moody CA & Laimins LA (2010) Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer*. **10**: 550-560.

Mori Y, Dhepakson P, Shimamoto T, Ueda K, Gomi Y, Tani H, Matsuura Y & Yamanishi K (2000) Expression of human herpesvirus 6B rep within infected cells and binding of its gene product to the TATA-binding protein in vitro and in vivo. *J Virol*. **74**: 6069-6104.

Mori Y, Seya T, Huang HL, Akkapaiboon P, Dhepakson P & Yamanishi K (2002) Human herpesvirus 6 variant A but not variant B induces fusion from without in a variety of human cells through a human herpesvirus 6 entry receptor, CD46. *J Virol*. **76**: 6750-6761.

Mori Y, Akkapaiboon P, Yang X & Yamanishi K (2003a) The human herpesvirus 6 U100 gene product is the third component of the gH-gL glycoprotein complex on the viral envelope. *J Virol*. **77**: 2452-2458.

Mori Y, Yang X, Akkapaiboon P, Okuno T & Yamanishi K (2003b) Human herpesvirus 6 variant A glycoprotein H-glycoprotein L-glycoprotein Q complex associates with human CD46. *J Virol*. **77**: 4992-4999.

Mori Y, Akkapaiboon P, Yonemoto S, Koike M, Takemoto M, Sadaoka T, Sasamoto Y, Konishi S, Uchiyama Y & Yamanishi K (2004) Discovery of a second form of tripartite complex containing gH-gL of human herpesvirus 6 and observations on CD46. *J Virol*. **78**: 4609-4616.

Mori Y, Koike M, Moriishi E, Kawabata A, Tang H, Oyaizu H, Uchiyama Y & Yamanishi K (2008) Human herpesvirus-6 induces MVB formation, and virus egress occurs by an exosomal release pathway. *Traffic*. **9**: 1728-1742.

Mori Y (2009a) Recent topics related to human herpesvirus 6 cell tropism. *Cell Microbiol*. **11**: 1001-1006.

- Mori T, Tanaka-Taya K, Satoh H, Aisa Y, Yamazaki R, Kato J, Ikeda Y & Okamoto S (2009b) Transmission of chromosomally integrated human herpesvirus 6 (HHV-6) variant A from a parent to children leading to misdiagnosis of active HHV-6 infection. *Transpl Infect Dis.* **11**: 503-506.
- Morris C, Luppi M, McDonald M, Barozzi P & Torelli G (1999) Fine mapping of an apparently latent human herpesvirus type 6 integration site in chromosome band 17p13.3. *J Med Virol.* **58**: 69-75.
- Morissette G & Flamand L (2010) Herpesviruses and chromosomal integration. *J Virol.* **84**: 12100-12109.
- Mortier A, Van Damme J & Proost P (2012) Overview of the mechanisms regulating chemokine activity and availability. *Immunol Lett.* **145**: 2-9.
- Moser B, Dewald B, Barella L, Schumacher C, Baggiolini M & Clark-Lewis I (1993) Interleukin-8 antagonists generated by N-terminal modification. *J Biol Chem.* **268**: 7125-7128.
- Murakami Y, Minami M, Daimon Y & Okanoue T (2004) Hepatitis B virus DNA in liver, serum, and peripheral blood mononuclear cells after the clearance of serum hepatitis B virus surface antigen. *J Med Virol.* **72**: 203-214.
- Murata T, Goshima F, Koshizuka T, Takakuwa H & Nishiyama Y (2001) A single amino acid substitution in the ICP27 protein of herpes simplex virus type 1 is responsible for its resistance to leptomycin B. *J Virol.* **75**: 1039-1043.
- Murphy PM, Baggiolini M, Charo IF, Hebert CA, Horuk R, Matsushima K, Miller LH, Oppenheim JJ & Power CA (2000) International union of pharmacology. XXII. Nomenclature for chemokine receptors. *Pharmacol Rev.* **52**: 145-176.
- Musnier A, Blanchot B, Reiter E & Crepieux P (2010) GPCR signalling to the translation machinery. *Cell Signal.* **22**: 707-716.
- Nacheva EP, Ward KN, Brazma D, Virgili A, Howard J, Leong HN & Clark DA (2008) Human herpesvirus 6 integrates within telomeric regions as evidenced by five different chromosomal sites. *J Med Virol.* **80**: 1952-1958.
- Nakagawa N, Mukai T, Sakamoto J, Hata A, Okuno T, Takeda K & Yamanishi K (1997) Antigenic analysis of human herpesvirus 7 (HHV-7) and HHV-6 using immune sera and monoclonal antibodies against HHV-7. *J Gen Virol.* **78**: 1131-1137.
- Neptune ER & Bourne HR (1997) Receptor induce chemotaxis by releasing the $\beta\gamma$ subunit of Gi not by activating Gq or Gs. *Proc Natl Acad Sci U S A.* **94**: 14489-14494.
- Neote K, DiGregorio D, Mak JY, Horuk R & Schall TJ (1993) Molecular cloning, functional expression, and signaling characteristic of a C-C chemokine receptor. *Cell.* **72**: 415-425.
- New DC & Wong YH (2003) CC chemokine receptor-coupled signalling pathways. *Acta*

Bioch Bioph Sin. **35**: 779-788.

Nibbs RJ & Graham (2013) Immune regulation by atypical chemokine receptors. *Nat Rev Immunol.* **13**: 815-829.

Nicola AV, McEvoy AM & Straus SE (2003) Roles for endocytosis and low pH in herpes simplex virus entry into HeLa and Chinese hamster ovary cells. *J Virol.* **77**: 5324-5332.

Niederman JC, Liu CR, Kaplan MH & Brown NA (1988) Clinical and serological features of human herpesvirus-6 infection in three adults. *Lancet.* **2**: 817-819.

Nomiyama H, Osada N, & Yoshie O (2010) The evolution of mammalian chemokine genes. *Cytokine Growth Factor Rev.* **21**: 253-262.

Nomiyama H, Osada N, & Yoshie O (2013) Systematic classification of vertebrate chemokines based on conserved synteny and evolutionary history. *Genes Cells.* **18**: 1-16.

Nowak RM (2015) Assembly of repetitive regions using next-generation sequencing data. *Biocybern Biomed Eng.* doi: 10.1016/j.bbe.2014.12.001.

Nukui M, Mori Y & Murphy EA (2015) A human herpesvirus 6A-encoded microRNA: role in viral lytic replication. *J Virol.* **89**: 2615-2627.

Oakes B, Hoagland-Henefield M, Komaroff AL, Erickson JL & Huber BT (2013) Human endogenous retrovirus-K18 superantigen expression and human herpesvirus-6 and human herpesvirus-7 viral loads in chronic fatigue patients. *Clin Infect Dis.* **56**: 1394-1400.

Offermanns S (2003) G-proteins as transducers in transmembrane signalling. *Prog Biophys Mol Biol.* **83**: 101-130.

Ogata M, Satou T, Kadota J, Saito N, Yoshida T, Okumura H, Ueki T, Nagafuji K, Kako S, Uoshima N, Tsudo M, Itamura H & Fukuda T (2013) Human herpesvirus 6 (HHV-6) reactivation and HHV-6 encephalitis after allogeneic hematopoietic cell transplantation: a multicenter, prospective study. *Clin Infect Dis.* **57**: 671-681.

Ogawa-Goto K, Tanaka K, Gibson W, Moriishi E, Miura Y, Kurata T, Irie S & Sata T (2003) Microtubule network facilitates nuclear targeting of human cytomegalovirus capsid. *J Virol.* **77**: 8541-8547.

Ohashi M, Sugata K, Ihira M, Asano Y, Egawa H, Takada Y, Uemoto S & Yoshikawa T (2008) Human herpesvirus 6 infection in adult living related liver transplant recipients. *Liver Transpl.* **14**: 100-109.

Ohye T, Inagaki H, Ihira M, Higashimoto Y, Kato K, Oikawa J, Yagasaki H, Niizuma T, Takahashi Y, Kojima S, Yoshikawa T & Kurahashi H (2014) Dual roles for the telomeric repeats in chromosomally integrated human herpesvirus-6. *Sci Rep.* **4**: 4559.

Ohye T, Kawamura Y, Inagaki H, Yoshikawa A, Ihira M, Yoshikawa T & Kurahashi H (2016) A simple cytogenetic method to detect chromosomally integrated human herpesvirus 6. *J Virol Methods.* **228**: 74-78.

Ojala PM, Sodeik B, Ebersold MW, Kutay U & Helenius A (2000) Herpes simplex virus type 1 entry into host cell: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro. *Mol Cell Biol.* **20**: 4922-4931.

Okuno T, Takahashi K, Balachandra K, Shiraki K, Yamanishi M & Baba K (1989) Seroepidemiology of human herpesvirus 6 infection in normal children and adults. *J Clin Microbiol.* **27**: 651-653.

Okuno T, Oishi H, Hayashi K, Nonogaki M, Tanaka K & Yamanishi K (1995) Human herpesvirus 6 and 7 in cervixes of pregnant women. *J Clin Microbiol.* **33**: 1968-1970.

Olshevsky U & Becker Y (1970) Herpes simplex virus structural proteins. *Virology.* **40**: 948-960.

Opsahl ML & Kennedy PG (2005) Early and late HHV-6 gene transcripts in multiple sclerosis lesion and normal appearing white matter. *Brain.* **128**: 516-527.

Osterrieder N, Kamil JP, Schumacher D, Tischer BK & Trapp S (2006) Marek's disease virus: from miasma to model. *Nat Rev Microbiol.* **4**: 283-294.

O'Sullivan JN, Bronner MP, Brentall TA, Finley JC, Shen WT, Emerson S, Emond MJ, Gollahon KA, Moskovitz AH, Crispin DA, Potter JD & Rabinovitch PS (2002) Chromosomal instability in ulcerative colitis is related to telomere shortening. *Nat Genet.* **32**: 280-284.

O'Sullivan RJ & Karlseder J (2010) Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol.* **11**: 171-181.

Ott TR, Lio FM, Olshefski D, Liu XJ, Struthers RS & Ling N (2004) Determinants of high-affinity binding and receptor activation in the N-terminus of CCL-19 (MIP-3 beta). *Biochemistry.* **43**: 3670-3678.

Ott TR, Lio FM, Olshefski D, Liu XJ, Ling N & Struthers RS (2006) The N-terminal domain of CCL21 reconstitutes high affinity binding, G protein activation, and chemotactic activity, to the C-terminal domain of CCL19. *Biochem Biophys Res Commun.* **348**: 1089-1093.

Otto TD, Dillon GP, Degraeve WS & Berriman M (2011) RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res.* **39**: e57.

Paavola CD, Hemmerich S, Grunberger D, Polsky I, Bloom A, Freedman R, Mulkins M, Bhakta S, McCarley D, Wiesent L, Wong B, Jarnagin K & Handel TM (1998) Monomeric monocyte chemoattractant protein-1 (MCP-1) binds and activates the MCP-1 receptor CCR2B. *J Biol Chem.* **273**: 33157-33165.

Pakianathan DR, Kuta EG, Artis DR, Skelton NJ & Hebert CA (1997) Distinct but overlapping epitopes for the interaction of a CC-chemokine with CCR1, CCR3 and CCR5. *Biochemistry.* **36**: 9642-9648.

Panussis DA, Cook MW, Rifkin LL, Snider JE, Strong JT, McGrane RM, Wilson RK & Mardis ER (1998) A pneumatic device for rapid loading of DNA sequencing gels. *Genome Res.* **8**:

543-548.

Pantry SN, Medveczky MM, Arbuckle JH, Luka J, Montoya JG, Hu J, Renne R, Peterson D, Pritchett JC, Ablashi DV & Medveczky PG (2013) Persistent human herpesvirus-6 infection in patients with an inherited form of the virus. *J Med Virol.* **85**: 1940-1946.

Papanikolaou E, Kouvatsis V, Dimitriadis G, Inoue N & Arsenakis M (2002) Identification and characterization of the gene products of open reading frame U86/U87 of human herpesvirus 6. *Virus Res.* **89**: 89-101.

Parenty G, Appelbe S & Milligan G (2008) CXCR2 chemokine receptor antagonism enhances DOP opioid receptor function via allosteric regulation of the CXCR2-DOP receptor heterodimer. *Biochem J.* **412**: 245-256.

Park SH, Casagrande F, Cho L, Albrecht L & Opella SJ (2011) Interactions of interleukin-8 with the human chemokine receptor CXCR1 in phospholipid bilayers by NMR spectroscopy. *J Mol Biol.* **414**: 194-203.

Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM & Opella SJ (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature.* **491**: 779-783.

Park D, Lengyel J & Rice SA (2013) Role of immediate early protein ICP27 in the differential sensitivity of herpes simplex viruses 1 and 2 to leptomycin B. *J Virol.* **87**: 8940-8951.

Paterson DL, Singh N, Gayowski T, Carrigan DR & Marino IR (1999) Encephalopathy associated with human herpesvirus 6 in a liver transplant recipient. *Liver Transpl Surg.* **5**: 454-455.

Paterson DA, Dyer AP, Milne RS, Sevilla-Reyes E & Gompels UA (2002) A role for human cytomegalovirus glycoprotein O (gO) in cell fusion and a new hypervariable locus. *Virology.* **293**: 281-294.

Paulsen SJ, Rosenkilde MM, Eugen-Olsen J & Kledal TN (2005) Epstein-Barr virus-encoded BILF1 is a constitutively active G protein-coupled receptor. *J Virol.* **79**: 536-546.

Pawelec G, Derhovanessian E, Larbi A, Strindhall J & Wikby A (2009) Cytomegalovirus and human immunosenescence. *Rev Med Virol.* **19**: 47-56.

Pease JE, Wang J, Ponath PD & Murphy PM (1998) The N-terminal extracellular segments of the chemokine receptors CCR1 and CCR3 are determinants for MIP-1alpha and eotaxin binding, respectively, but a second domain is essential for efficient receptor activation. *J Biol Chem.* **273**: 19972-19976.

Pease J & Williams TJ (2006) The attraction of chemokines as a target for specific anti-inflammatory therapy. *Br J Pharmacol.* **147**: S212-221.

Pellet PE, Ablashi DV, Ambros PF, Agut H, Caserta MT, Descamps V, Flamand L, Gautheret-Dejean A, CB, Kamble RT, Kuehl U, Lassner D, Lautenschlager I, Loomis KS, Luppi M, Lusso P, Medveczky PG, Montoya JG, Mori Y, Ogata M, Pritchett JC, Rogez S, Seto E, Ward KN,

- Yoshikawa T & Razonable RR (2012) Chromosomally integrated human herpesvirus 6: questions and answers. *Rev Med Virol.* **22**: 144-155.
- Pellet P & Roizman B (2013) The family *Herpesviridae*: a brief introduction. In Knipe D & Howley PM (Eds.) *Fields Virology* (6th edition), Lippincott, Williams and Wilkins, Philadelphia.
- Pello OM, Martinez-Munoz L, Parrillas V, Serrano A, Rodriguez-Frade JM, Toro MJ, Lucas P, Monterrubio M, Martinez AC & Mellado M (2008) Ligand stabilization of CXCR4/delta-opioid receptor heterodimers reveals a mechanism for immune response regulation. *Eur J Immunol.* **38**: 537-549.
- Percherancier Y, Berchiche YA, Slight I, Volkmer-Engert R, Tamamura H, Fujii N, Bouvier M & Heveker N (2005) Bioluminescence resonance energy transfer reveals ligand-induced conformational changes in CXCR4 homo- and heterodimers. *J Biol Chem.* **280**: 9895-9903.
- Pereira L (1994) Functions of glycoprotein B homologues of the family herpesviridae. *Infect Agents Dis.* **3**: 9-28.
- Pereira DJ, McCarty DM & Muzyczka N (1997) The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J Virol.* **71**: 1079-1088.
- Perzova R, Graziano E, Sanghi S, Welch C, Benz P, Abbott L, Lalone D, Glaser J, Loughran T, Sheremata W & Poiesz BJ (2011) Increased seroreactivity to HERV-K10 peptides in patients with HTLV myelopathy. *Virol J.* **10**: 360.
- Pfeiffer B, Berneman ZN, Neipel F, Chang CK, Tirwatnapong S & Chandran B (1993) Identification and mapping of the gene encoding the glycoprotein complex gp82-gp105 of human herpesvirus 6 and mapping of the neutralizing epitope recognized by monoclonal antibodies. *J Virol.* **67**: 4611-4620.
- Pfeiffer B, Thomson B & Chandran B (1995) Identification and characterization of a cDNA derived from multiple splicing that encodes envelope glycoprotein gp105 of human herpesvirus 6. *J Virol.* **69**: 3490-3500.
- Pfeiffer V & Lingner J (2012) TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS Genet.* **8**: e1002747.
- Pleskoff O, Treboute C & Alizon M (1998) The cytomegalovirus-encoded chemokine receptor US28 can enhance cell-cell fusion mediated by different viral proteins. *J Virol.* **72**: 6389-6397.
- Pond SL, Scheffler K, Gravenor MB, Poon AF & Frost SD (2010) Evolutionary fingerprinting of genes. *Mol Biol Evol.* **27**: 520-536.
- Potenza L, Barrozzi P, Masetti M, Pecorari M, Bresciani P, Gautheret-Dejean A, Riva G, Vallerini D, Tagliazucchi S, Codeluppi M, Di Benedetto F, Gerunda GE, Narni F, Torelli G & Luppi M (2009) Prevalence of human herpesvirus-6 chromosomal integration (CIHHV-6) in Italian solid organ and allogenic stem cell transplant patients. *Am J Transplant.* **9**: 1690-

1697.

Prado GN, Suetomi K, Shumate D, Maxwell C, Ravindran A, Rajarathnam K & Navarro J (2007) Chemokine signaling specificity: essential role for the N-terminal domain of chemokine receptors. *Biochemistry*. **46**: 8961-8968.

Preobrazhensky AA, Dragan S, Kawano T, Gavrilin MA, Gulina IV, Chakravarty L & Kolattukudy PE (2000) Monocyte chemotactic protein-1 receptor CCR2B is a glycoprotein that has tyrosine sulfation in a conserved extracellular N-terminal region. *J Immunol*. **165**: 5295-5303.

Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA & Baumeister K (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science*. **238**: 336-341.

Proost P, Struyf S & Van Damme J (2006) Natural post-translational modifications of chemokines. *Biochem Soc Trans*. **34**: 997-1001.

Proudfoot AE, Power CA, Hoogewerf AJ, Montjovent MO, Borlat F, Offord RE & Wells TN (1996) Extension of recombinant human RANTES by the retention of the initiating methionine produces a potent antagonist. *J Biol Chem*. **271**: 2599-2603.

Proudfoot AE, Handel TM, Johnson Z, Lau EK, LiWang P, Clark-Lewis I, Borlat F, Wells TN & Kosco-Vilbois MH (2003) Glycosaminoglycan binding and oligomerization are essential for the in vivo activity of certain chemokines. *Proc Natl Acad Sci U S A*. **100**: 1885-1890.

Proudfoot AE (2006) The biological relevance of chemokine-proteoglycan interactions. *Biochem Soc Trans*. **34**: 422-426.

Prusty BK, Siegl C, Hauck P, Hain J, Korhonen SJ, Hiltunen-Back E, Puolakkainen M & Rudel T (2013a) Chlamydia trachomatis infection induces replication of latent HHV-6. *PLoS One*. **8**: e61400.

Prusty BK, Krohne G & Rudel T (2013b) Reactivation of chromosomally integrated human herpesvirus-6 by telomeric circle formation. *PLoS Genet*. **9**: e1004033.

Qin L, Kufareva I, Holden LG, Wang C, Zheng Y, Zhao C, Fenalti G, Wu H, Han GW, Cherezov V, Abagyan R, Stevens RC & Handel TM (2015) Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science*. **347**: 1117-1122.

Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC & Hall N (2012) Application of next-generation sequencing technologies in virology. *J Gen Virol*. **93**: 1853-1868.

Radtke K, Dohner K & Sodeik B (2006) Viral interactions with the cytoskeleton: a hitchhiker's guide to the cell. *Cell Microbiol*. **8**: 387-400.

Rajagopal S, Bassoni DL, Campbell JJ, Gerard NP, Gerard C & Wehrman TS (2013) Biased agonism as a mechanism for differential signaling by chemokine receptors. *J Biol Chem*. **288**: 35039-35048.

- Rajagopalan L & Rajarathnam K (2004) Ligand selectivity and affinity of chemokine receptor CXCR1. Role of N-terminal domain. *J Biol Chem.* **279**: 30000-30008.
- Rajagopalan L & Rajarathnam K (2006) Structural basis of chemokine function – a model for binding affinity and ligand selectivity. *Biosci Rep.* **26**: 325-339.
- Rajarathnam K, Sykes BD, Kay CM, Dewald B, Geiser T, Baggiolini M & Clark-Lewis I (1994) Neutrophil activation by monomeric interleukin-8. *Science.* **264**: 90-92.
- Rana S & Baranski TJ (2010) Third extracellular loop (EC3)-N terminus interaction is important for seven transmembrane domain receptor functions: implications for an activation switch. *J Biol Chem.* **285**: 31472-31483.
- Randolph-Habecker JR, Rahill B, Torok-Storb B, Vieira J, Kolattukudy PE, Rovin BH and Sedmak DD (2002) The expression of the cytomegalovirus chemokine receptor homolog US28 sequesters biologically active CC chemokines and alters IL-8 production. *Cytokine.* **19**: 37-46.
- Rapp JC, Krug LT, Inoue N, Dambaugh TR & Pellet PE (2000) U94, the human herpesvirus 6 homolog of the parvovirus nonstructural gene, is highly conserved among isolates and is expressed at low mRNA levels as a spliced transcript. *Virology.* **268**: 504-516.
- Reardon JE & Spector T (1989) Herpes simplex virus type 1 DNA polymerase. Mechanism of inhibition by acyclovir triphosphate. *J Biol Chem.* **264**: 7405-7411.
- Renzette N, Bhattacharjee B, Jensen JD, Gibson L & Kowalik TF (2011) Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog.* **7**: e1001344.
- Reske A, Pollara G, Krummenacher C, Chain BM & Katz DR (2007) Understanding HSV-1 entry glycoproteins. *Rev Med Virol.* **17**: 205-215.
- Reynaud JM, Jegou JF, Welsch J & Horvat B (2014) Human herpesvirus 6A infection in CD46 transgenic mice: viral persistence in the brain and increased production of proinflammatory chemokines via TLR9. *J Virol.* **88**: 5421-5436.
- Rice SA & Knipe DM (1990) Genetic evidence for two distinct transactivation functions of the herpes simplex virus alpha protein ICP27. *J Virol.* **64**: 1704-1715.
- Richardson RM, Ali H, Tomhave ED, Haribabu B & Snyderman R (1995) Cross-desensitization of chemoattractant receptors occurs at multiple levels. Evidence for a role for inhibition of phospholipase C activity. *J Biol Chem.* **270**: 27829-27833.
- Riethman H, Ambrosini A, Castaneda C, Finklestein J, Hu XL, Mudunuri U, Paul S & Wei J (2004) Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res.* **14**: 18-28.
- Rieux C, Gautheret-Dejean A, Chaille-Lehmann D, Kirch C, Agut H & Vernant JP (1998) Human herpesvirus-6 meningoencephalitis in a recipient of an unrelated allogeneic bone marrow transplantation. *Transplantation.* **65**: 1408-1411.

- Ritter SL & Hall RA (2009) Fine-tuning of GPCR activity by receptor-interacting proteins. *Nat Rev Mol Cell Biol.* **10**: 819-830.
- Roizmann B, Desrosiers RC, Fleckenstein B, Lopez C, Minson AC & Studdert MJ (1992) The family Herpesviridae: an update. The Herpesvirus Study Group of the International Committee on Taxonomy of Viruses. *Arch Virol.* **123**: 425-449.
- Roizman B, Knipe DM & Whitley RJ (2013) Herpes Simplex Viruses. In Knipe D & Howley PM (Eds.) *Fields Virology* (6th edition), Lippincott, Williams and Wilkins, Philadelphia.
- Rosenfeld CS, Rybka WB, Weinbaum D, Carrigan DR, Knox KK, Andrews DF & Shadduck RK (1995) Late graft failure due to dual bone marrow infection with variants A and B of human herpesvirus-6. *Exp Hematol.* **23**: 626-629.
- Ross EM & Wilkie TM (2000) GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins. *Annu Rev Biochem.* **69**: 795-827.
- Rossi D & Zlotnik A (2000) The biology of chemokines and their receptors. *Annu Rev Immunol.* **18**: 217-242.
- Rotola A, Ravaioli T, Gonelli A, Dewhurst S, Cassai E & Di Luca D (1998) U94 of human herpesvirus 6 is expressed in latently infected peripheral blood mononuclear cells and blocks viral gene expression in transformed lymphocytes in culture. *Proc Natl Acad Sci U S A.* **95**: 13911-13916.
- Rotola A, Ravaioli T, Gonelli A, Dewhurst S, Cassai E, Pfeiffer V & Lingner J (2012) TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS Genet.* **8**: e1002747.
- Roush KS, Domiati-Saad RK, Margraf LR, Krisher K, Scheuermann RH, Rogers BB & Dawson DB (2001) Prevalence and cellular reservoir of latent human herpesvirus 6 in tonsillar lymphoid tissue. *Am J Clin Pathol.* **116**: 648-654.
- Ruan J, Jiang L, Chong Z, Gong Q, Li H, Li C, Tao Y, Zheng C, Zhai W, Turissini D, Cannon CH, Lu X and Wu CI (2013) Pseudo-Sanger sequencing: massively parallel production of long and near error-free reads using NGS technology. *BMC Genomics.* **14**: 711.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA & Barrell B (2000) Artemis: sequence visualization and annotation. *Bioinformatics.* **16**: 944-945.
- Ryckman BJ, Chase MC & Johnson DC (2008) HCMV gH/gL/UL128-131 interferes with virus entry into epithelial cells: evidence for cell type-specific receptors. *Proc Natl Acad Sci U S A.* **105**: 14118-14123.
- Rye HS, Dabora JM, Quesada MA, Mathies RA & Glazer AN (1993) Fluorometric assay using dimeric dyes for double- and single-stranded DNA and RNA with picogram sensitivity. *Anal Biochem.* **208**: 144-150.

- Sabroe I, Jorritsma A, Stubbs VE, Xanthou G, Jopling LA, Ponath PD, Williams TJ, Murphy PM & Pease JE (2005) The carboxyl terminus of the chemokine receptor CCR3 contains distinct domains which regulate chemotactic signaling and receptor down-regulation in a ligand-dependent manner. *Eur J Immunol.* **35**: 1301-1310.
- Sahin E, Colla S, Liesa M, Moslehi J, Müller FL, Guo M, Cooper M, Kotton D, Fabian AJ, Walkey C, Maser RS, Tonon G, Foerster F, Xiong R, Wang YA, Shukla SA, Jaskelioff M, Martin ES, Heffernan TP, Protopopov A, Ivanova E, Mahoney JE, Kost-Alimova M, Perry SR, Bronson R, Liao R, Mulligan R, Shirihai OS, Chin L & DePinho RA (2011) Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature.* **470**: 359-365.
- Salahuddin SZ, Ablashi DV, Markham PD, Josephs SF, Sturzenegger S, Kaplan M, Halligan G, Biberfeld P, Wong-Staal F, Kramarsky B & Gallo RC (1986) Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. *Science.* **234**: 596-601.
- Salanga CL & Handel TM (2011) Chemokine oligomerization and interactions with receptors and glycosaminoglycans: the role of structural dynamics in function. *Exp Cell Res.* **317**: 590-601.
- Samulski RJ, Berns KI, Tan M & Muzyczka N (1982) Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. *Proc Natl Acad Sci U S A.* **79**: 2077-2081.
- Samulski RJ, Zhu X, Xiao X, Brook JD, Housman DE, Epstein N & Hunter LA (1991) Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J.* **10**: 3941-3950.
- Sandri-Goldin RM (2011) The many roles of the highly interactive HSV protein ICP27, a key regulator of infection. *Future Microbiol.* **6**: 1261-1277.
- Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* **74**: 5463-5467.
- Santoro F, Kennedy PE, Locatelli G, Malnati MS, Berger EA & Lusso P (1999) CD46 is a cellular receptor for human herpesvirus 6. *Cell.* **99**: 817-827.
- Santoro F, Greenstone HL, Insinga A, Liszewski MK, Atkinson JP, Lusso P & Berger EA (2003) Interaction of glycoprotein H of human herpesvirus 6 with the cellular receptor CD46. *J Biol Chem.* **278**: 25964-25969.
- Sarmiento J, Shumate C, Suetomi K, Ravindran A, Villegas L, Rajarathnam K & Navarro J (2011) Diverging mechanisms of activation of chemokine receptors revealed by novel chemokine agonists. *PLoS One.* **6**: e27967.
- Sato M, Li H, Ikizler MR, Werkhaven JA, Williams JV, Chappell JD, Tang YW & Wright PF (2009) Detection of viruses in human adenoid tissues by use of multiplex PCR. *J Clin Microbiol.* **47**: 771-773.
- Schall TJ & Proudfoot AE (2011) Overcoming hurdles in developing successful drugs targeting chemokine receptors. *Nat Rev Immunol.* **11**: 355-363.

- Sheaffer AK, Newcomb WW, Gao M, Yu D, Weller SK, Brown JC & Tenney DJ (2001) Herpes simplex virus DNA cleavage and packaging proteins associate with the procapsid prior to its maturation. *J Virol.* **75**: 687-698.
- Scheibye-Alsing K, Hoffmann S, Frankel A, Jensen P, Stadler PF, Mang Y, Tommerup N, Gilchrist MJ, Nygard AB, Cirera S, Jorgensen CB, Fredholm M & Gorodkin J (2009) Sequence assembly. *Comput Biol Chem.* **33**: 121-136.
- Schermuly J, Greco A, Hartle S, Osterrieder N, Kaufer BB & Kaspers B (2015) In vitro model for lytic replication, latency, and transformation of an oncogenic alphaherpesvirus. *Proc Natl Acad Sci U S A.* **112**: 7279-7284.
- Schiffer M & Edmundson AB (1967) Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys J.* **7**: 121-135.
- Schmidt M, Afione S & Kotin RM (2000) Adeno-associated virus type 2 Rep78 induces apoptosis through caspase activation independently of p53. *J Virol.* **74**: 9441-9450.
- Schnepf BC, Jensen RL, Chen CL, Johnson PR & Clark KR (2005) Characterization of adeno-associated virus genomes isolated from human tissues. *J Virol.* **79**: 14793-14803.
- Scholz B, Rechter S, Drach JC, Townsend LB & Bogner E (2003) Identification of the ATP-binding site in the terminase subunit pUL56 of human cytomegalovirus. *Nucleic Acids Res.* **31**: 1426-1433.
- Schioth HB & Fredriksson R (2005) The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen Comp Endocrinol.* **142**: 94-101.
- Schirmer EC, Wyatt LS, Yamanishi K, Rodriguez WJ & Frenkel N (1991) Differentiation between two distinct classes of viruses now classified as human herpesvirus 6. *Proc Natl Acad Sci U S A.* **88**: 5922-5926.
- Schleimann MH, Hoberg S, Solhøj Hansen A, Bundgaard B, Witt CT, Kofod-Olsen E & Hollsbørg P (2014) The DR6 protein from human herpesvirus-6B induces p53-independent cell cycle arrest in G2/M. *Virology.* **452**: 254-263.
- Schnur E, Kessler N, Zherdev Y, Noah E, Scherf T, Ding FX, Rabinovich S, Arshava B & Kurbatska V, Leonciks A, Tsimanis A, Rosen O, Naider F & Anglister J (2013) NMR mapping of RANTES surfaces interacting with CCR5 using linked extracellular domains. *FEBS J.* **280**: 2068-2084.
- Schraufstatter IU, Barritt DS, Ma M, Oades ZG & Cochrane CG (1993) Multiple sites on IL-8 responsible for binding to alpha and beta IL-8 receptors. *J Immunol.* **151**: 6418-6428.
- Secchiero P, Carrigan DR, Asano Y, Benedetti L, Crowley RW, Komaroff AL, Gallo RC & Lusso P (1995) Detection of human herpesvirus 6 in plasma of children with primary infection and immunosuppressed patients by polymerase chain reaction. *J Infect Dis.* **171**: 273-280.

- Sedlak RH & Jerome KR (2013) Viral diagnostics in the era of digital polymerase chain reaction. *Diagn Microbiol Infect Dis.* **75**: 1-4.
- Sedlak RH, Cook L, Huang ML, Magaret A, Zerr DM, Boeckh M & Jerome KR (2014) Identification of chromosomally integrated human herpesvirus 6 by droplet digital PCR. *Clin Chem.* doi: 10.1373/clinchem.2013.217240.
- Shang J, Zhu F, Vongsangnak W, Tang Y, Zhang W & Shen B (2014) Evaluation and comparison of multiple aligners for next-generation sequencing data analysis. *Biomed Res Int.* doi: 10.1155.
- Shay JW & Wright WE (2005) Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis.* **26**: 867-874.
- Shi G, Partida-Sanchez S, Misra RS, Tighe M, Borchers MT, Lee JJ, Simon MI & Lund FE (2007) Identification of an alternative Galphaq-dependent chemokine receptor signal transduction pathway in dendritic cells and granulocytes. *J Exp Med.* **204**: 2705-2718.
- Shinkai A, Komuta-Kunitomo M, Sato-Nakamura N & Anazawa H (2002) N-terminal domain of eotaxin-3 is important for activation of CC chemokine receptor 3. *Protein Eng.* **15**: 923-929.
- Shukla D & Spear PG (2001) Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. *J Clin Invest.* **108**: 503-510.
- Sieczkowski L, Chandran B & Wood C (1995) The human immunodeficiency virus tat gene enhances replication of human herpesvirus-6. *Virology.* **211**: 544-553.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **7**: 539.
- Sijmons S, Van Ranst M & Maes P (2014) Genomic and functional characteristics of human cytomegalovirus revealed by next-generation sequencing. *Viruses.* **6**: 1049-1072.
- Simmons G, Clapham PR, Picard L, Offord RE, Rosenkilde MM, Schwartz TW, Buser R, Wells TN & Proudfoot AE (1997) Potent inhibition of HIV-1 infectivity in macrophages and lymphocytes by a novel CCR5 antagonist. *Science.* **276**: 276-279.
- Simon MI, Strathmann MP & Gautam N (1991) Diversity of G proteins in signal transduction. *Science.* **252**: 802-808.
- Simpson S Jr, Taylor B, Dwyer DE, Taylor J, Blizzard L, Ponsonby AL, Pittas F, Dwyer T & van der Mei I (2012) Anti-HHV-6 IgG titer significantly predicts subsequent relapse risk in multiple sclerosis. *Mult Scler.* **18**: 799-806.
- Sinclair J (2008) Human cytomegalovirus: latency and reactivation in the myeloid lineage. *J Clin Virol.* **41**: 180-185.
- Singer II, Scot S, Kawka DW, Chin J, Daugherty BL, DeMartino JA, Gould SL, Lineberger JE,

- Malkowitz L, Miller MD, Mitnaul L, Siciliano SJ, Staruch MJ, Williams HR, Zweerink HJ & Springer MS (2001) CCR5, CXCR4, and CD4 are clustered and closely apposed on microvilli of human macrophages and T Cells. *J Virol.* **75**: 3779-3790.
- Sjahril R, Isegawa Y, Tanaka T, Nakano K, Yoshikawa T, Asano Y, Ohshima A, Yamanishi K & Sugimoto N (2009) Relationship between U83 gene variation in human herpesvirus 6 and secretion of U83 gene product. *Arch Virol.* **154**: 273-283.
- Skelton NJ, Quan C, Reilly D & Lowman H (1999) Structure of a CXC chemokine-receptor fragment in complex with interleukin-8. *Structure.* **7**: 157-168.
- Slinger E, Maussang D, Schreiber A, Siderius M, Rahbar A, Fraile-Ramos A, Lira SA, Soderberg-Naucler C & Smit MJ (2010) HCMV-encoded chemokine receptor US28 mediates proliferative signaling through the IL-6-STAT3 axis. *Sci Signal.* **3**: ra58.
- Smit MJ, Vischer HF, Bakker RA, Jongejan A, Timmerman H, Pardo L & Leurs R (2007) Pharmacogenomic and structural analysis of constitutive G protein-coupled receptor activity. *Annu Rev Pharmacol Toxicol.* **47**: 53-87.
- Smith A, Santoro F, Di Lullo G, Dagna L, Verani A & Lusso P (2003) Selective suppression of IL-12 production by human herpesvirus 6. *Blood.* **102**: 2877-2884.
- Snoeck R, Andrei G, Bodaghi B, Lagneaux L, Daelemans D, De Clerq E, Neyts J, Schols D, Naesens L, Michelson S, Bron D, Otto MJ, Boussesu A, Nemecek C & Roy C (2002) 2-Chloro-3-pyridin-3-yl-6,6,7,8-tetrahydroindolizine-1-carboxamide (CMV423), a new lead compound for the treatment of human cytomegalovirus infections. *Antiviral Res.* **55**: 413-424.
- Sodeik B (2000) Mechanisms of viral transport in the cytoplasm. *Trends Microbiol.* **8**: 465-472.
- Sohy D, Parmentier M & Springael JY (2007) Allosteric transinhibition by specific antagonists in CCR2/CXCR4 heterodimers. *J Biol Chem.* **282**: 30062-30069.
- Soldan SS, Berti R, Salem N, Secchiero P, Flamand L, Calabresi PA, Brennan MB, Maloni HW, McFarland HF, Lin HC, Patnaik M & Jacobson S (1997) Association of human herpesvirus 6 (HHV-6) with multiple sclerosis: increased IgM response to HHV-6 early antigen and detection of serum HHV-6 DNA. *Nat Med.* **3**: 1394-1397.
- Soldan SS, Leist TP, Juhng KN, McFarland HF & Jacobson S (2000) Increased lymphoproliferative response to human herpesvirus type 6A variant in multiple sclerosis. *Ann Neurol.* **47**: 306-313.
- Soliman TM, Sandri-Goldin RM & Silverstein SJ (1997) Shuttling of the herpes simplex virus type 1 regulatory protein ICP27 between the nucleus and cytoplasm mediates the expression of late proteins. *J Virol.* **71**: 9188-9197.
- Sorkin A & von Zastrow M (2002) Signal transduction and endocytosis: close encounters of many kinds. *Nat Rev Mol Cell Biol.* **3**: 600-614.
- Spear PG & Longnecker R (2003) Herpesvirus entry: an update. *J Virol.* **77**: 10179-10185.

Spiess K, Jeppesen MG, Malmgaard-Clausen M, Krzywkowski K, Dulal K, Cheng T, Hjorto GM, Larsen O, Burg JS, Jarvis MA, Garcia KC, Zhu H, Kledal TN & Rosenkilde MM (2015) Rationally designed chemokine-based toxin targeting the viral G protein-coupled receptor US28 potentially inhibits cytomegalovirus infection in vivo. *Proc Natl Acad Sci U S A.* pii: 201509392.

Sprang SR (2011) Cell signalling: Binding the receptor at both ends. *Nature.* **469**: 172-173.

Srivastava A, Lusby EW & Berns KI (1983) Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol.* **45**: 555-564.

Streblow DN, Soderberg-Naucler C, Vieira J, Smith P, Wakabayashi E, Ruchti F, Mattison K, Altschuler Y & Nelson JA (1999) The human cytomegalovirus chemokine receptor US28 mediates vascular smooth muscle cell migration. *Cell.* **99**: 511-520.

Strenger V, Aberle SW, Wendelin G, Pfurtscheller K, Nacheva EP, Zobel G & Nagel B (2010) Chromosomal integration of the HHV-6 genome as a possible cause of HHV-6 detection in cardiac tissues. *J Clin Pathol.* **63**: 1129-1130.

Strenger V, Caselli E, Lautenschlager I, Schwinger W, Aberle SW, Loginov R, Gentili V, Nacheva E, DiLuca D & Urban C (2014) Detection of HHV-6-specific mRNA and antigens in PBMCs of individuals with chromosomally integrated HHV-6 (ciHHV-6). *Clin Microbiol Infect.* **20**: 1027-1032.

Springael JY, Le Minh PN, Urizar E, Costagliola S, Vassart G & Parmentier M (2006) Allosteric modulation of binding properties between units of chemokine receptor homo- and hetero- oligomers. *Mol Pharmacol.* **69**: 1652-1661.

Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, Mann M, Ingolia NT & Weissman JS (2012) Decoding human cytomegalovirus. *Science.* **338**: 1088-1093.

Struyf S, De Meester I, Scharpe S, Lenaerts JP, Menten P, Wang JM, Proost P & Van Damme J (1998) Natural truncation of RANTES abolishes signaling through the CC chemokine receptors CCR1 and CCR3, impairs its chemotactic potency and generates a CC chemokine inhibitor. *Eur J Immunol.* **28**: 1262-1271.

Struyf S, Proost P, Schols D, De Clercg E, Opdenakker G, Lenaerts JP, Detheux M, Parmentier M, De Meester I, Scharpe S & Van Damme J (1999) CD26/dipeptidase IV down-regulates the eosinophil chemotactic potency, but not the anti-HIV activity of human eotaxin by affecting its interaction with CC chemokine receptor 3. *J Immunol.* **162**: 4903-4909.

Stutika C, Hüser D, Weger S, Rutz N, Heßler M & Heilbronn R (2015) Definition of herpes simplex virus helper functions for the replication of adeno-associated virus type 5. *J Gen Virol.* **96**: 840-850.

Subramanian S & Lambert DM (2012) Selective constraints determine the time dependency of molecular rates for human nuclear genomes. *Genome Biol Evol.* **4**: 1127-

1132.

Suga S, Yoshikawa T, Kajita Y, Ozaki T & Asano Y (1998) Prospective study of persistence and excretion of human herpesvirus-6 in patients with exanthem subitum and their parents. *Pediatrics*. **102**: 900-904.

Sullivan BM & Coscoy L (2008) Downregulation of the T-cell receptor complex and impairment of T-cell activation by human herpesvirus 6 u24 protein. *J Virol*. **82**: 602-608.

Swingler S, Easton A & Morris A (1992) Cytokine augmentation of HIV-1 LTR-driven gene expression in neural cells. *AIDS Res Hum Retroviruses*. **8**: 487-493.

Szabo I, Wetzel MA & Rogers TJ (2001) Cell-density-regulated chemotactic responsiveness of keratinocytes in vitro. *J Invest Dermatol*. **117**: 1083-1090.

Szpakowska M, Fievez V, Arumugan K, van Nuland N, Schmit JC & Chevigne A (2012) Functional diversity and therapeutic potential of the N-terminal domain of human chemokine receptors. *Biochem Pharmacol*. **84**: 1366-1380.

Szpara ML, Gatherer D, Ochoa A, Greenbaum B, Dolan A, Bowden RJ, Enquist LW, Legendre M & Davison AJ (2014) Evolution and diversity in human herpes simplex virus genomes. *J Virol*. **88**: 1209-1227.

Tadagaki K, Yamanishi K & Mori Y (2007) Reciprocal roles of cellular chemokine receptors and human herpesvirus 7-encoded chemokine receptors, U12 and U51. *J Gen Virol*. **88**: 1423-1428.

Tafreshi NK, Sadeghizadeh M, Amini-Bavil-Olyaei S, Ahadi AM, Jahanzad I & Roostaei MH (2005) Development of a multiplex nested consensus PCR for detection and identification of major human herpesviruses in CNS infections. *J Clin Virol*. **32**: 318-324.

Tai AK, Luka J, Ablashi D & Huber BT (2009) HHV-6A infection induces expression of HERV-K18-encoded superantigen. *J Clin Virol*. **46**: 47-48.

Takahashi K, Sonoda S, Higashi K, Takahashi H, Takahashi M & Yamanishi K (1989) Predominant CD4 T-lymphocyte tropism of human herpesvirus 6-related virus. *J Virol*. **63**: 3161-3163.

Takahashi K, Segal E, Kondo T, Mukai T, Moriyama M, Takahashi M & Yamanishi K (1992) Interferon and natural killer cell activity in patients with exanthem subitum. *Pediatr Infect Dis J*. **11**: 369-373.

Takeda K, Okuno T, Isegawa Y & Yamanishi K (1996) Identification of a variant A-specific neutralizing epitope on glycoprotein B (gB) of human herpesvirus-6 (HHV-6). *Virology*. **222**: 176-183.

Tamura K & Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. **10**: 512-526.

- Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* **30**: 2725-2729.
- Tan JH, Canals M, Ludeman JP, Wedderburn J, Boston C, Butler SJ, Carrick AM, Parody TR, Taleski D, Christopoulos A, Payne RJ & Stone MJ (2012) Design and receptor interactions of obligate dimeric mutant of chemokine monocyte chemoattractant protein-1 (MCP-1). *J Biol Chem.* **287**: 14692-14702.
- Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q & Wu B (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor complex. *Science.* **341**: 1387-1390.
- Tanaka K, Kondo T, Torigoe S, Okada S, Mukai T & Yamanishi K (1994) Human herpesvirus 7: another causal agent for roseola (exanthem subitum). *J Pediatr.* **125**: 1-5.
- Tanaka Y, Suenaga T, Matsumoto M, Seya T & Arase H (2013) Herpesvirus 6 glycoproteins B (gB), gH, gL, and gQ are necessary and sufficient for cell-to-cell fusion. *J Virol.* **87**: 10900-10903.
- Tanaka-Taya K, Kondo T, Mukai T, Miyoshi H, Yamamoto Y, Okado S & Yamanishi K (1996) Seroepidemiological study of human herpesvirus-6 and -7 in children of different ages and detection of these two viruses in throat swabs by polymerase chain reaction. *J Med Virol.* **48**: 88-94.
- Tanaka-Taya K, Kondo T, Nakagawa N, Inagi R, Miyoshi H, Sunagawa T, Okada S & Yamanishi K (2000) Reactivation of human herpesvirus 6 by infection of human herpesvirus 7. *J Med Virol.* **60**: 284-289.
- Tanaka-Taya K, Sashihara J, Kurahashi H, Amo K, Miyagawa H, Kondo K, Okado S & Yamanishi K (2004) Human herpesvirus 6 (HHV-6) is transmitted from parent to child in an integrated form and characterisation of cases with chromosomally integrated HHV-6 DNA. *J Med Virol.* **73**: 465-473.
- Tang H, Kawabata A, Yoshida M, Oyaizu H, Maeki T, Yamanishi K & Mori Y (2010) Human herpesvirus 6 encoded glycoprotein Q1 gene is essential for virus growth. *Virology.* **407**: 360-367.
- Tang H, Hayashi M, Maeki T, Yamanishi K & Mori Y (2011) Human herpesvirus 6 glycoprotein complex formation is required for folding and trafficking of the gH/gL/gQ1/gQ2 complex and its cellular receptor binding. *J Virol.* **85**: 11121-11130.
- Tang H, Serada S, Kawabata A, Ota M, Hayashi E, Naka T, Yamanishi K & Mori Y (2013) CD134 is a cellular receptor specific for human herpesvirus-6B entry. *Proc Natl Acad Sci U S A.* **110**: 9096-9099.
- Tedder RS, Briggs M, Cameron CH, Honess R, Robertson D & Whittle H (1987) A novel lymphotropic herpesvirus. *Lancet.* **2**: 390-392.
- Telford EA, Watson MS, Aird HC, Perry J & Davison AJ (1995) The DNA sequence of equine herpesvirus 2. *J Mol Biol.* **249**: 520-528.

- Tembo J, Kabwe M, Chilukutu L, Chilufya M, Mwaanza N, Chabala C, Zumla A & Bates M (2015) Prevalence and risk factors for betaherpesvirus DNAemia in children >3 weeks and <2 years of age admitted to a large referral hospital in sub-saharan Africa. *Clin Infect Dis.* **60**: 423-431.
- Tengelsen LA, Pederson NE, Shaver PR, Wathen MW & Homa FL (1993) Herpes simplex virus type 1 DNA cleavage and encapsidation require the product of the UL28 gene: isolation and characterization of two UL28 deletion mutants. *J Virol.* **67**: 3470-3480.
- Terrillon S & Bouvier M (2004) Roles of G-protein-coupled receptor dimerisation. *EMBO Rep.* **5**: 30-34.
- Thomson BJ, Efsthathiou S & Honess RW (1991) Acquisition of the human adeno-associated virus type-2 rep gene by human herpesvirus type-6. *Nature.* **351**: 78-80.
- Thomson BJ, Dewhurst S & Gray D (1994a) Structure and heterogeneity of the sequences of human herpesvirus 6 strain variants U1102 and Z29 and identification of human telomeric repeat sequences at the genomic termini. *J Virol.* **68**: 3007-3014.
- Thomson BJ, Weindler FW, Gray D, Schwaab V & Heilbronn R (1994b) Human herpesvirus 6 (HHV-6) is a helper virus for adeno-associated virus type 2 (AAV-2) and the AAV-2 rep gene homologue in HHV-6 can mediate AAV-2 DNA replication and regulate gene expression. *Virology.* **204**: 304-311.
- Tong CY, Bakran A, Peiris JS, Muir P & Herrington CS (2002) The association of viral infection and chronic allograft nephropathy with graft dysfunction after renal transplantation. *Transplantation.* **74**: 576-578.
- Torelli G, Barozzi P, Marasca R, Cocconcelli P, Merelli E, Ceccherini-Nelli L, Ferrari S & Luppi M (1995) Targeted integration of human herpesvirus 6 in the p arm of chromosome 17 of human peripheral blood mononuclear cells in vivo. *J Med Virol.* **46**: 178-188.
- Torrisi MR, Gentile M, Cardinali G, Cirone M, Zompetta C, Lotti LV, Frati L & Faggioni A (1999) Intracellular transport and maturation pathway of human herpesvirus 6. *Virology.* **257**: 460-471.
- Toufaily C, Landry S, Leib-Mosch C, Rassart E & Barbeau B (2011) Activation of LTRs from different human endogenous retrovirus (HERV) families by the HTLV-1 tax protein and T-cell activators. *Viruses.* **3**: 2146-2159.
- Tratschin JD, Tal J & Carter BJ (1986) Negative and positive regulation in trans of gene expression from adeno-associated virus vectors in mammalian cells by a viral rep gene product. *Mol Cell Biol.* **6**: 2884-2894.
- Treangen TJ & Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet.* **13**: 36-46.
- Trempe F, Gravel A, Dubuc I, Wallaschek N, Collin V, Gilbert-Girard S, Morissette G, Kaufert BB & Flamand L (2015) Characterization of human herpesvirus 6A/B U94 as ATPase,

helicase, exonuclease and DNA-binding proteins. *Nucleic Acids Res.* pii: gkv503.

Trott O & Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* **31**: 455-461.

Tsao EH, Kellam P, Sin CS, Rasaiyaah J, Griffiths PD & Clark DA (2009) Microarray-based determination of the lytic cascade of human herpesvirus 6B. *J Gen Virol.* **90**: 2581-2591.

Tsuchiya S, Yamabe M, Yamaguchi Y, Kobayashi Y, Konno T & Tada K (1980) Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int J Cancer.* **26**: 171-176.

Tuddenham L, Jung JS, Chane-Woon-Ming B, Dolken L & Pfeffer S (2012) Small RNA deep sequencing identifies microRNAs and other small noncoding RNAs from human herpesvirus 6B. *J Virol.* **86**: 1638-1649.

Tullman ER, Afonso CL, Lu Z, Zsak L, Rock DL & Kutish GF (2000) The genome of a very virulent Marek's disease virus. *J Virol.* **74**: 7980-7988.

Turcanova VL, Bundgaard B & Hollberg P (2009) Human herpesvirus-6B induces expression of the human endogenous retrovirus K18-encoded superantigen. *J Clin Virol.* **46**: 15-19.

Turner S, DiLuca D & Gompels U (2002) Characterisation of a human herpesvirus 6 variant A 'amplicon' and replication modulation by U94-Rep 'latency gene'. *J Virol Methods.* **105**: 331-341.

Tweedy J, Spyrou MA, Hubacek P, Kuhl U, Lassner D & Gompels UA (2015a) Analyses of germline chromosomally intergrated human herpesvirus 6A and B indicate emergent infection and new inflammatory mediators. *J Gen Virol.* **96**: 370-389.

Tweedy J, Spyrou MA, Donaldson CD, Depledge D, Breuer J & Gompels UA (2015b) Complete genome sequence of the human herpesvirus 6A strain AJ from Africa resembles strain GS from North America. *Genome Announc.* **3**: e01498-14.

Tweedy J, Spyrou MA, Pearson M, Lassner D, Kuhl U & Gompels UA (2016) Complete genome sequence of germline chromosomally integrated human herpesvirus 6A and analyses of integration site define a new human endogenous virus with potential to reactivate as an emerging infection. *Viruses.* **8**: E19.

Tyler SD, Peters GA, Grose C, Severini A, Gray MJ, Upton C & Tipples GA (2007) Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. *Virology.* **359**: 447-458.

Tyler S, Severini A, Black D, Walker M & Eberle R (2011) Structure and sequence of the saimiriine herpesvirus 1 genome. *Virology.* **410**: 181-191.

Udi J, Schuler J, Wider D, Ihorst G, Catusse J, Waldschmidt J, Schnerch D, Follo M, Wasch R & Engelhardt M (2013) Potent in vitro and in vivo activity of sorafenib in multiple myeloma:

- induction of cell death, CD138-downregulation and inhibition of migration through actin depolymerisation. *Br J Haematol.* **161**: 104-116.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**: W71-W74.
- Uprichard SL & Knipe DM (1996) Herpes simplex ICP27 mutant viruses exhibit reduced expression of specific DNA replication genes. *J Virol.* **70**: 1969-1980.
- Vaidehi N, Pease JE, & Horuk R (2009) Modeling small molecule-compound binding to G-protein-coupled receptors. *Methods Enzymol.* **460**: 263-288.
- van der Kuyl AC (2012) HIV infection and HERV expression: a review. *Retrovirology.* **9**: 6.
- van Dijk EL, Auger H, Jaszczyszyn Y & Thermes C (2014) Ten years of next-generation sequencing technology. *Trends Genet.* **30**: 418-426.
- van Koppen CJ & Jakobs KH (2004) Arrestin-independent internalization of G protein-coupled receptors. *Mol Pharmacol.* **66**: 365-367.
- Veldkamp CT, Seibert C, Peterson FC, Sakmar TP & Volkman BF (2006) Recognition of a sulfotyrosine by the chemokine stromal cell-derived factor-1alpha (SDF-1alpha/CXCL12). *J Mol Biol.* **359**: 1400-1409.
- Veldkamp CT, Seibert C, Peterson FC, De la Cruz NB, Haugner JC, Basnet H, Sakmar TP & Volkman BF (2008) Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Sci Signal.* **1**: doi 10.1126/scisignal.1160755.
- Vieira J, Schall TJ, Corey L & Geballe AP (1998) Functional analysis of the human cytomegalovirus US28 gene by insertion mutagenesis with the green fluorescent protein gene. *J Virol.* **72**: 8158-8165.
- Violin JD & Lefkowitz RJ (2007) Beta-arrestin-biased ligands at seven-transmembrane receptors. *Trends Pharmacol Sci.* **28**: 416-422.
- Vischer HF, Vink C & Smit MJ (2006) A viral conspiracy: hijacking the chemokine system through virally encoded pirated chemokine receptors. *Curr Top Microbiol Immunol.* **303**: 121-154.
- Vischer HF, Siderius M, Leurs R & Smit MJ (2014) Herpesvirus-encoded GPCRs: neglected players in inflammatory and proliferative diseases? *Nat Rev Drug Discov.* **13**: 123-139.
- Vomaske J, Nelson JA & Streblow DN (2009a) Human cytomegalovirus US28: a functionally selective chemokine binding receptor. *Infect Disord Drug Targets.* **9**: 548-556.
- Vomaske J, Melnychuk RM, Smith PP, Powell J, Hall L, DeFilippis V, Fruh K, Smit M, Schlaepfer DD, Nelson JA & Streblow DN (2009b) Differential ligand binding to a human cytomegalovirus chemokine receptor determines cell type-specific motility. *PLoS Pathogen.* **5**: e1000304.

- Vroon A, Heijnen CJ & Kavelaars A (2006) GRKs and arrestins: regulators of migration and inflammation. *J Leukoc Biol.* **80**: 1214-1221.
- Wagstaff AJ & Bryson HM (1994) Foscarnet. A reappraisal of its antiviral activity, pharmacokinetic properties and therapeutic use in immunocompromised patients with viral infections. *Drugs.* **48**: 199-226.
- Wang FZ, Linde A, Dahl H & Ljungman P (1999) Human herpesvirus 6 infection inhibits specific lymphocyte proliferation responses and is related to lymphocytopenia after allogeneic stem cell transplantation. *Bone Marrow Transplant.* **24**: 1201-1206.
- Ward P, Urcelay E, Kotin R, Safer B & Berns KI (1994) Adeno-associated virus DNA replication in vitro: activation by a maltose binding protein/Rep 68 fusion protein. *J Virol.* **68**: 6029-637.
- Ward P, Elias P & Linden RM (2003) Rescue of the adeno-associated virus genome from a plasmid vector: evidence for rescue by replication. *J Virol.* **77**: 11480-11490.
- Ward KN (2005) Human herpesviruses-6 and -7 infections. *Curr Opin Infect Dis.* **18**: 247-252.
- Ward KN, Thiruchelvam AD & Couto-Parada X (2005) Unexpected occasional persistence of high levels of HHV-6 DNA in sera: detection of variants A and B. *J Med Virol.* **76**: 563-570.
- Ward KN, Leong HN, Nacheva EP, Howard J, Atkinson CE, Davies NW, Griffiths PD & Clark DA (2006) Human herpesvirus 6 chromosomal integration in immunocompetent patients results in high levels of viral DNA in blood, sera, and hair follicles. *J Clin Microbiol.* **44**: 1571-1574.
- Watanabe H, Daibata M, Tohyama M, Batchelor J, Hashimoto K & Iijima M (2008) Chromosomal integration of human herpesvirus 6 DNA in anticonvulsant hypersensitivity syndrome. *Br J Dermatol.* **158**: 640-642.
- Weiss RA (2006) The discovery of endogenous retroviruses. *Retrovirology.* **3**: 67.
- Weitzman MD, Kyostio SR, Kotin RM & Owens RA (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc Natl Acad Sci U S A.* **91**: 5808-5812.
- Wettschureck N & Offermanns S (2005) Mammalian G Proteins and their cell type specific functions. *Physiol Rev.* **85**: 1159-1204.
- White CA, Stow ND, Patel AH, Hughes M & Preston VG (2003) Herpes simplex virus type 1 portal protein UL6 interacts with the putative terminase subunits UL15 and UL28. *J Virol.* **77**: 6351-6358.
- Whitley RJ (2004) Congenital cytomegalovirus infection: epidemiology and treatment. *Adv Exp Med Biol.* **549**: 155-160.
- Wilborn F, Schmidt CA, Brinkmann V, Jendroska K, Oettle H & Siegert W (1994) A potential role for human herpesvirus type 6 in nervous system disease. *J Neuroimmunol.* **49**: 213-

214.

Williams G, Borkakoti N, Bottomley GA, Cowan I, Fallowfield AG, Jones PS, Kirtland SJ & Price L (1996) Mutagenesis studies of interleukin-8. Identification of a second epitope involved in receptor binding. *J Biol Chem.* **271**: 9579-9586.

Wilson S, Wilkinson G & Milligan G (2005) The CXCR1 and CXCR2 receptors form constitutive homo- and heterodimers selectively and with equal apparent affinities. *J Biol Chem.* **280**: 28663-28674.

Winkler M, Rice SA & Stamminger T (1994) UL69 of human cytomegalovirus, an open reading frame with homology to ICP27 of herpes simplex virus, encodes a transactivator of gene expression. *J Virol.* **68**: 3943-3954.

Winkler M & Stamminger T (1996) A specific subform of the human cytomegalovirus transactivator protein pUL69 is contained within the tegument of virus particles. *J Virol.* **70**: 8984-8987.

Wistrand M, Kall L & Sonnhammer EL (2006) A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Sci.* **15**: 509-521.

Witt DP & Lander AD (1994) Differential binding of chemokines to glycosaminoglycan subpopulations. *Curr Biol.* **4**: 394-400.

Wolfe BL & Trejo J (2007) Clathrin-dependent mechanisms of G protein-coupled receptor endocytosis. *Traffic.* **8**: 462-470.

Wonderling RS, Kyostio SR & Owens RA (1995) A maltose-binding protein/adeno-associated virus Rep68 fusion protein has DNA-RNA helicase and ATPase activities. *J Virol.* **69**: 3542-3548.

Wu L, Ruffing N, Shi X, Newman W, Soler D, Mackay CR & Qin S (1996) Discrete steps in binding and signaling of interleukin-8 with its receptor. *J Biol Chem.* **271**: 31202-31209.

Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katrich V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V & Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science.* **330**: 1066-1071.

Wu H, Liu L, Xiao J, Chi M, Qu Y & Feng H (2015) Glycosylation of KSHV encoded vGPCR functions in its signaling and tumorigenicity. *Viruses.* **7**: 1627-1641.

Wyatt LS, Balachandran N & Frenkel N (1990) Variations in the replication and antigenic properties of human herpesvirus 6 strains. *J Infect Dis.* **162**: 852-857.

Xi R, Hadjipanayis AG, Luquette LJ, Kim TM, Lee E, Zhang J, Johnson MD, Muzny DM, Wheeler DA, Gibbs RA, Kucherlapati R & Park PJ (2011) Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *Proc Natl Acad Sci U S A.* **108**: e1128-1136.

- Yamanishi K, Okuno T, Shiraki K, Takahashi M, Kondo T, Asano Y & Kurata T (1988) Identification of human herpesvirus-6 as a causal agent for exanthem subitum. *Lancet*. **1**: 1065-1067.
- Yasukawa M, Inoue Y, Ohminami H, Terada K & Fujita S (1998) Apoptosis of CD4+ T lymphocytes in human herpesvirus-6 infection. *J Gen Virol*. **79**: 143-147.
- Yasukawa M, Ohminami H, Sada E, Yakushijin Y, Kaneko M, Yanagisawa K, Kohno H, Bando S & Fujita S (1999) Latent infection and reactivation of human herpesvirus 6 in two novel myeloid cell lines. *Blood*. **93**: 991-999.
- Ye J, Kohli LL & Stone MJ (2000) Characterisation of binding between the chemokine eotaxin and peptides derived from the chemokine receptor CCR3. *J Biol Chem*. **275**: 27250-27257.
- Yoshida H, Matsunaga K, Ueda T, Yasumi M, Ishikawa J, Tomiyama Y & Matsuzawa Y (2002) Human herpesvirus 6 meningoencephalitis successfully treated with ganciclovir in a patient who underwent allogenic bone marrow transplantation from an HLA-identical sibling. *Int J Hematol*. **75**: 421-425.
- Yoshikawa T, Asano Y, Akimoto S, Ozaki T, Iwasaki T, Kurata T, Goshima F & Nishiyama Y (2002) Latent infection of human herpesvirus 6 in astrocytoma cell line and alteration of cytokine synthesis. *J Med Virol*. **66**: 497-505.
- Yu D, Silva MC & Shenk T (2003) Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc Natl Acad Sci U S A*. **100**: 12396-12401.
- Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. **18**: 821-829.
- Zerr DM, Gooley TA, Yeung L, Huang ML, Carpenter P, Wade JC, Corey L & Anasetti C (2001) Human herpesvirus 6 reactivation and encephalitis in allogenic bone marrow transplant recipients. *Clin Infect Dis*. **33**: 763-771.
- Zerr DM, Meier AS, Selke SS, Frenkel LM, Huang ML, Wald A, Rhoads MP, Nguy L, Bornemann R, Morrow RA & Corey L (2005a) A population-based study of primary human herpesvirus 6 infection. *N Engl J Med*. **352**: 768-776.
- Zerr DM, Corey L, Kim HW, Huang ML, Nguy L & Boeckh M (2005b) Clinical outcomes of human herpesvirus 6 reactivation after hematopoietic stem cell transplantation. *Clin Infect Dis*. **40**: 932-940.
- Zerr DM (2012) Human herpesvirus 6 (HHV-6) disease in the setting of transplantation. *Curr Opin Infect Dis*. **25**: 438-444.
- Zhang W, Chen J, Yang Y, Tang Y, Shang J & Shen B (2011) A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. *PLoS One*. **6**: e17915.

Zhao J, Fan H, Mu G, Shen X & Cheng X (1997) Detection of human herpesvirus 6 (HHV-6) DNA in salivary glands by polymerase chain reaction. *Chin Med Sci J.* **12**: 126-128.

Zhen Z, Bradel-Tretheway B, Sumagin S, Bidlack JM & Dewhurst S (2005) The human herpesvirus 6 G protein-coupled receptor homolog U51 positively regulates virus replication and enhances cell-cell fusion in vitro. *J Virol.* **79**: 11914-11924.

Zielke B, Thomas M, Giede-Jeppe A, Muller R & Stamminger T (2011) Characterization of the betaherpesviral pUL69 protein family reveals binding of the cellular mRNA export factor UAP56 as a prerequisite for stimulation of nuclear mRNA export and for efficient viral replication. *J Virol.* **85**: 1804-1819.

Zimmermann J & Hammerschmidt W (1995) Structure and role of the terminal repeats of Epstein-Barr virus in processing and packaging of virion DNA. *J Virol.* **69**: 3147-3155.

Zlotnik A & Yoshie O (2000) Chemokines: a new classification system and their role in immunity. *Immunity.* **12**: 121-127.

Zlotnik A, Yoshie O, & Nomiya H (2006) The chemokine and chemokine receptor superfamilies and their molecular evolution. *Genome Biol.* **7**: 243-254.

Zlotnik A & Yoshie O (2012) The chemokine superfamily revisited. *Immunity.* **36**: 705-716.

Zoffmann S, Chollet A & Galzi JL (2002) Identification of the extracellular loop 2 as the point of interaction between the N terminus of the chemokine MIP-1alpha and its CCR1 receptor. *Mol Pharmacol.* **62**: 729-736.

Zou P, Isegawa Y, Nakano K, Haque M, Horiguchi Y & Yamanishi K (1999) Human herpesvirus 6 open reading frame U83 encodes a functional chemokine. *J Virol.* **73**: 5926-5933.