Human Cytomegalovirus Transmission in Early Childhood and Impact of Maternal HIV in Zambia: A Molecular Study

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

KUNDA GEOFFREY MUSONDA

January 2017

Supervisor: Dr. Ursula A. Gompels
Pathogen Molecular Biology Department
Faculty of Infectious and Tropical Diseases
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Affiliations: Gompels Virology Research Group, PMBD/ITD, LSHTM, London, United Kingdom.
Virology Laboratory, Pathology & Microbiology Department, UTH, Lusaka, Zambia.

Kunda Geoffrey Musonda is a Commonwealth Scholar, funded by the UK government.

DECLARATION

I, KUNDA GEOFFREY MUSONDA, 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.

Signed: ……………………………………………………………...... Date:..

ABSTRACT

Human cytomegalovirus (HCMV) is the leading infectious cause of birth defects and neurodevelopmental delay, with worst outcomes in congenital infection. HCMV exacerbates neurological disease and progression to AIDS in HIV co-infected infants. In Zambia HCMV has adverse effects on growth and development even in maternally HIV-exposed uninfected infants, affecting a third. Transmission to infants can be congenital or postnatal primarily through breast milk, similar to HIV. However, the route affecting infant development is not established, but critical to understand for deployment of appropriate interventions and hospital diagnostics. In this thesis, congenital HCMV infection was investigated in newborns using PCR and nucleotide sequencing of saliva and umbilical cord DNA, and showed 1% prevalence, which was too low to account for the widespread adverse growth effects. Therefore, maternal HCMV secretion via breast milk and the effects of maternal HIV were investigated. HCMV viral loads were measured by qPCR in 461 breast milk samples from both HIV-infected and uninfected mothers at postpartum day 3 and weeks 2, 4, 9, 12, and 16. Results showed higher HCMV viral loads and longer secretion in HIV-positive compared to negative mothers, peaking at week 4 ($p=.026$) and remaining elevated at week 16 ($p<.001$). HCMV strains were investigated using the hypervariable glycoprotein gO, with all eight genotypes detected, but no relationship to viral load. Next, Illumina NGS was used to further analyse strains and mixed infections in order to assess the burden of infection. A script was developed to identify linked gO and gN 'Molecular tags' via analyses of FASTQ reads. In 21 samples, mixed infections were identified, showing 1-7 genotypes and a dominant genotype constituting 45-100%. ZMB240, a full-length HCMV genome - the first reference strain from a healthy donor representing a transmission population, and also first from Africa - was derived directly from breast milk DNA and found similar to elsewhere, including the reference Merlin. Overall, the thesis provided evidence that breast milk is the principal route for early HCMV infection in breastfed Zambian infants, and that maternal HIV increases HCMV load, burden of infection, and duration of secretion in breast milk, thereby escalating and prolonging risk of early infant HCMV infection.

TABLE OF CONTENTS

DECLARATION 2
ABSTRACT 3
TABLE OF CONTENTS 4
LIST OF FIGURES 7
LIST OF TABLES 10
LIST OF ABBREVIATIONS 11
ACKNOWLEDGEMENTS 13
DEDICATION 14

CHAPTER 1: INTRODUCTION .... **15 – 44**
1.1 Background 15
1.2 Virus Classification and Morphology 17
1.3 HCMV Genome Organisation and Content 20
1.4 HCMV Replication Cycle 24
1.5 HCMV Latency and Reactivation 27
1.6 Virus Gene Variation and Strain Identification 28
1.7 HCMV Epidemiology and Transmission 29
1.8 HCMV in Early Childhood 32
1.8.1 Congenital Infection 33
1.8.2 Postnatal Infection 34
1.8.3 HCMV and HIV/AIDS Infection in Childhood 35
1.9 HCMV Treatment and Control 36
1.9.1 Hyperimmune Globulin 36
1.9.2 Drugs 36
1.9.3 Vaccines 40
1.9.4 Non-pharmacological Control Methods 42
1.10 Thesis Aims 43

CHAPTER 2: METHODS .... **45 – 58**
2.1 Ethical Considerations 45
2.2 Study Cohorts and Sample Collection 45
2.2.1 Neonatal Cohort 46
2.2.2 Breastfeeding Cohort 47
2.3 DNA Extraction and PCR 48
2.3.1 DNA Extraction 48
2.3.2 Qualitative PCR 50
2.3.3 Quantitative Real-time PCR 53
2.4 Plasmid Cloning 53
2.5 Nucleotide Sequencing 55
2.5.1 Gel DNA Purification 55
2.5.2 Sanger Sequencing 56
2.5.3 Illumina Next Generation Sequencing 57
2.6 Bioinformatics 57

2.7 Statistical Analysis ... 58

CHAPTER 3: INVESTIGATION OF HCMV AT BIRTH IN ZAMBIA USING NON-INVASIVE SAMPLES ... 59 – 73

3.1 Introduction ... 59
3.2 Assay Optimisation ... 60
3.3 Pilot Studies Using Non-Invasive Samples ... 63
3.4 Evaluation of HCMV, HHV-6A, and HHV-6B in Neonates at UTH ... 66
3.5 Discussion ... 71

CHAPTER 4: ANALYSIS OF HCMV IN BREAST MILK OF HIV-POSITIVE AND NEGATIVE MOTHERS IN ZAMBIA ... 74 – 103

4.1 Introduction ... 74
4.2 Breast Milk DNA from BFPH Cohort ... 75
4.3 HCMV Detection in Breast Milk ... 76
4.4 HCMV Viral Loads in Breast Milk ... 77
4.5 HCMV gO and gN Genotypes in Breast Milk ... 89
4.6 Discussion ... 98

CHAPTER 5: GENOMIC ANALYSIS OF HCMV DIVERSITY IN BREAST MILK 104 – 195

5.1 Introduction ... 104
5.2 Identification of ‘Molecular Tags’ and Development of *Perl* Scripts .. 107
5.3 Analysis of HCMV NGS Sequences from Breast Milk ... 144
5.4 Assembly of Complete HCMV Genomes from Breast Milk ... 158
5.5 Discussion ... 193

CHAPTER 6: SUMMARY AND CONCLUSIONS ... 196 – 202

6.1 Summary ... 196
6.2 Future Work ... 199
6.3 Conclusions ... 201

REFERENCES ... 203 – 238

APPENDICES ... 239 – 338

1. Ethics Approval Letter – UNZABREC ... 239
2. Ethics Approval Letter – LSHTM REC ... 240
3. Research Approval – UTH ... 241
4. BFPH Sample Reuse Approval ... 242
5. Participant Information Sheet – Neonatal Study ... 243
6. Informed Consent Form – Neonatal Study ... 246
7. Data Capture Form – Neonatal Study ... 247
8. HCMV UL55 Oligonucleotide Primer & Probe Location ... 248
9. HCMV UL74 Oligonucleotide Primer Location ... 250
10. Complete HCMV Genomes Used in this Study ... 252
11. Partial and Complete HCMV gO CDSs Used in this Study ... 261
12. Partial and Complete HCMV gN CDSs Used in this Study ... 269

13. HCMV Strain ZMB240 Gene List 286
14. Manuscript 1 313
15. Manuscript 2 322
16. Manuscript 3 330

LIST OF FIGURES

Figure 1.1. Herpesvirus ultrastructure 18

Figure 1.2. Histological section of kidney showing HCMV-infected cells 19

Figure 1.3. The six classes of herpesvirus genomes 21

Figure 1.4. Structure of the four HCMV genome isomers 21

Figure 1.5. Genetic organization and content of wild-type HCMV 23

Figure 1.6. Summary of HCMV replication 26

Figure 1.7. HCMV latency and reactivation 28

Figure 1.8. Approved anti-HCMV drugs 38

Figure 3.1. PCR primer sensitivity 61

Figure 3.2. HCMV gB qPCR standard curve and amplification plot 62

Figure 3.3. PCR amplification of HCMV gB: saliva samples 64

Figure 3.4. PCR amplification of HHV-6A/B U38: saliva samples 64

Figure 3.5. PCR amplification of HCMV gB in sample 100S 69

Figure 3.6. Molecular phylogenetic analyses of sample 100S gO and gN 70

Figure 4.1. PCR amplification of GAPDH in breast milk 75

Figure 4.2. PCR amplification of HCMV gB in breast milk 76

Figure 4.3. HCMV viral load kinetics in breast milk from HIV-positive mothers 79-81

Figure 4.4. HCMV viral load kinetics in breast milk from HIV-negative mothers 82-84

Figure 4.5. Kinetics of HCMV shedding in breast milk of HIV-positive and negative mothers 86

Figure 4.6. Rate of HCMV viral load change 88

Figure 4.7. HCMV Viral Loads in Week 16 Breast milk: HIV-positive vs. HIV-negative mothers 89

Figure 4.8. PCR amplification of HCMV gO: breast milk 91

Figure 4.9. Phylogenetic analysis of gO genotypes in breast milk 91

Figure 4.10. Prevalence of HCMV gO genotypes in breast milk 92

Figure 4.11. Genotype independent increases in HCMV load in breast milk from HIV-positive and negative women 94

Figure 4.12. PCR amplification of HCMV gN: breast milk 95

Figure 4.13. Prevalence of HCMV gN genotype in W4 vs. W16 breast milk 96

Figure 4.14. HCMV gN genotype prevalence in breast milk of HIV-positive and negative mothers 97

Figure 4.15. Comparison of HCMV gO genotype prevalence in breast milk and other tissues 102

Figure 5.1. Molecular phylogenetic relationship of the eight HCMV gO genotypes 108

Figure 5.2. Amino acid sequence alignment of HCMV gO from 163 complete HCMV genomes 109-111

Figure 5.3. Nucleotide sequence alignment of HCMV gO from 163 complete HCMV genomes 112-114

Figure 5.4. Molecular phylogenetic analysis of HCMV gN genotypes 115

Figure 5.5. Amino acid sequence alignment of HCMV gN from 163 complete HCMV genomes 116-118

Figure 5.6. Nucleotide sequence alignment of HCMV gN from 163 complete HCMV genomes 119-121

Figure 5.7. Polymorphic sites across 163 HCMV gO and gN sequences 122

Figure 5.8. Screenshots of search results returned by *perl* scripts 125

Figure 5.9. Full-length HCMV gO amino acid sequence alignment of BE/23/2010, AD169 (gO1a reference strain) and Toledo (gO1c reference strain) 140

Figure 5.10. Full-length HCMV gO nucleotide sequence alignment of BE/23/2010, AD169 (gO1a reference strain) and Toledo (gO1c reference strain) 141-142

Figure 5.11. HCMV gO genotype proportions in four breast milk samples from HIV-negative mothers 149

Figure 5.12. HCMV gO genotype proportions in 17 breast milk samples from HIV-positive mothers 150-151

Figure 5.13. HCMV gO genotypes in breast milk of three HIV-positive and three HIV-negative mothers at postpartum week 12 152

Figure 5.14. HCMV gO genotypes in breast milk from four mothers with milk samples collected from both left (L) and right (R) breasts at week 16 postpartum 153

Figure 5.15. HCMV gO genotypes and viral load: sample 243R 155

Figure 5.16. HCMV gO genotypes and viral load: sample 278R 156

Figure 5.17. HCMV gO genotypes and viral load: sample 288R 157

Figure 5.18. World map showing global distribution of complete HCMV genomes by country 160

Figure 5.19. HCMV strain ZMB240 genome layout 161

Figure 5.20. Nucleotide sequence alignment of the putative UL6 ORF of strain ZMB240 and 25 other HCMV strains 164

Figure 5.21. ZMB240 UL22A amino acid sequence alignment 167

Figure 5.22. ZMB240 UL29 amino acid sequence alignment 167

Figure 5.23. ZMB240 UL32 amino acid alignment 168

Figure 5.24. ZMB240 UL33 amino acid alignment 169

Figure 5.25. ZMB240 UL37 amino acid alignment 169-170

Figure 5.26. ZMB242 UL42 amino acid alignment 170

Figure 5.27. ZMB240 UL48 amino acid alignment 171

Figure 5.28. ZMB240 UL50 amino acid alignment 172
Figure 5.29. ZMB240 UL74 amino acid alignment 172
Figure 5.30. ZMB240 UL77 amino acid alignment 173
Figure 5.31. ZMB242 UL95 amino acid alignment 173-174
Figure 5.32. ZMB242 UL100 amino acid alignment 174-175
Figure 5.33. ZMB242 UL102 amino acid alignment 175
Figure 5.34. ZMB240 UL112 amino acid alignment 176
Figure 5.35. ZMB240 protein UL117 amino acid alignment 177
Figure 5.36. ZMB240 UL119 amino acid alignment 177-178
Figure 5.37. Polymorphisms in ZMB240 UL27 183
Figure 5.38. Polymorphisms in ZMB240 UL97 (Serine/Threonine Kinase) and UL54 (DNA polymerase) gene sequences 184

LIST OF TABLES

Table 1.1: Genotypes of selected Hypervariable HCMV genes 30

Table 2.1: Type of sample collected by ward 47

Table 2.2: Oligonucleotide primers and probes used in our PCR and qPCR assays 52

Table 2.3: Characteristics of plasmid clones...55

Table 3.1: HCMV gB qPCR Limit of Detection (LOD) determination 63

Table 3.2: Donor samples used in assay set-up and optimization 65

Table 3.3: Neonatal cohort: descriptive statistics 67

Table 3.4: Reason for presenting to UTH labour ward and infant's diagnosis in NICU .68

Table 3.5: Type of sample collected by ward 69

Table 4.1: Prevalence of HCMV gO genotypes in breast milk 92

Table 4.2. HCMV gO genotype in paired W4/W16 breast milk samples........................93

Table 4.3. HCMV gN genotype prevalence in W4 vs. W16 breast milk..........................96

Table 4.4. HCMV gN genotype prevalence in breast milk of HIV-positive and negative mothers 97

Table 4.5. HCMV gN genotype in paired W4/W16 breast milk samples........................98

Table 4.6. Effects of maternal HIV and breastfeeding duration on infant HCMV infection 100

Table 4.7. Proportion of breast milk samples with HCMV DNA load above 5.5 log_{10} copies/ml 101

Table 5.1: Molecular tags developed for detection of HCMV gO and gN genotypes ..123

Table 5.2: HCMV gO and gN genotype identification using molecular tags developed during this thesis 126-129

Table 5.3: HCMV gO genotype identification in partial HCMV sequences available in GenBank using molecular tags 130-134

Table 5.4: HCMV gN genotype identification using molecular tags 135-139

Table 5.5: Partial HCMV gN CDSs not assigned genotypes by our molecular tags 143

Table 5.6: Breast milk samples analysed by NGS 145

Table 5.7: Summary of FASTQ reads pre-processing for DNA derived from 21 breast milk samples 147

Table 5.8: Confirmation of HCMV gO / gN linkage in breast milk analysed by NGS 148

Table 5.9: Overview of HCMV strain ZMB240 genome content 162

Table 5.10: ZMB240 Core and Beta genes whose products are of variable length compared to HCMV reference strain Merlin (AY446894.2) 166

Table 5.11: HCMV microRNAs found in the ZMB240 genome 181

Table 5.12: HCMV strain ZMB240 gene list 185-192

LIST OF ABBREVIATIONS

aa	amino acid residues
AIDS	Acquired Immunodeficiency Syndrome
bp	base pairs
BFPH	Breast Feeding and Postpartum Maternal Health [study]
°C	degrees Celsius
CIGNIS	Chilenje Infant Growth Nutrition and Infection Study
CMV	Cytomegalovirus
DNA	Deoxyribonucleic Acid
EBV	Epstein-Barr virus
FAM	6-carboxyfluorescein
g	grams
GATK	Genome Analysis Toolkit
gp	glycoprotein
h	hour(s)
HAART	Highly active antiretroviral therapy
HCMV	Human Cytomegalovirus
HEU	HIV-exposed uninfected
HHV-5	Human betaherpesvirus 5
HHV-6	Human herpesvirus 6
HHV-6A	Human betaherpesvirus 6A
HHV-6B	Human betaherpesvirus 6B
HHV-7	Human betaherpesvirus 7
HHV-8	Human gammaherpesvirus 8
HIV	Human Immunodeficiency Virus
HSV-1	Herpes Simplex Virus type 1
HSV-2	Herpes Simplex Virus type 2
IRL	internal long repeat
IRS	internal short repeat
K_a	ratio of non-synonymous substitutions per possible non-synonymous site
kb	kilobase(s)
kbp	kilobase pairs
K_s	ratio of synonymous substitutions per possible synonymous site
KSHV	Kaposi sarcoma-associated herpesvirus
LSHTM	London School of Hygiene and Tropical Medicine
M	molar
mg	milligram(s)
min	minute(s)
ml	millilitres
mM	millimolar
mRNA	messenger ribonucleic acid
MS	multiple sclerosis

NGS	Next Generation Sequencing
NK	Natural killer cell
nm	nanometres
nt	nucleotides
ORF	open reading frame
PCR	Polymerase Chain Reaction
PMTCT	Prevention of mother-to-child transmission (of HIV)
qPCR	Quantitative Real-time Polymerase Chain Reaction
RNA	Ribonucleic acid
RT	reverse transcriptase
s	second(s)
SSA	Sub-Saharan Africa
TAMRA	5-Carboxytetramethylrhodamine
TBE	Tris/Borate/EDTA
TRL	long terminal repeat
TRS	short terminal repeat
UL	unique long
US	unique short
UTH	University Teaching Hospital
UTHVL	University Teaching Hospital Virology Laboratory
UV	ultraviolet
VZV	Varicella-Zoster Virus
WGS	Whole genome sequencing
µg	micrograms
µl	microlitres
µM	micromolar

ACKNOWLEDGEMENTS

I wish to express profound gratitude to the Commonwealth Scholarship Commission for awarding me a scholarship, which made it possible to undertake these PhD studies.

I sincerely thank my principal supervisor and teacher, Dr. Ursula A. Gompels, for guidance, support, encouragement, training, and exposure provided during my PhD journey. Thanks also go to Prof. Susanne Filtaeu, Dr. Lackson Kasonka, and BFPH investigators for graciously providing access to the breast milk samples analysed in my studies. I also pay special tribute to all participants and their families for willingly accepting to take part in my studies. Further gratitude goes to UTH Labour Ward, Neonatal Intensive Care Unit, and Virology Laboratory staff for their cooperation and assistance at various stages of my studies. I particularly wish to thank Dr. Mwaka Monze, Molly Chisenga, Joshua Siame, Edward Chentulo, Chisenga Musonda, and Paul Simusika for their immense support. I also acknowledge the invaluable support from Joshua Tweedy (Gompels research group, LSHTM) who helped me settle in the laboratory and got me started with NGS data analysis. Further, I thank Mary Nyonda (MSc. MBID, LSHTM 2013) and Margaret Njenga (MSc. MM, LSHTM 2015), who conducted some of the qPCR and genotyping experiments. Special thanks also go to the Andrew Davison group, at the University of Glasgow's MRC Center for Virus Research, for support with Next Generation Sequencing. I thank Jody Phelan (Taane Clarke group, LSHTM) for cheerfully navigating me through the UNIX system and NGS bioinformatics analyses.

I am eternally grateful and indebted to my dear wife Nyaxewo and precious daughters Chipo and Mapalo, for their steadfast faith, inspiration, and forbearance during my study and absence from them. Special thanks also go to my parents and the wider family, for unwavering encouragement and ceaseless prayers.

Above all, I give thanks and praise to the Lord God Almighty for His abundant grace and reassuring presence throughout this thrilling stage of my life's journey.

DEDICATION

I would like to dedicate this thesis to my

Wife Nyaxewo,

Daughters Chipo & Mapalo,

Parents Geoffrey Lingson Mututa & Elinah Chewe Ndoti Musonda, and

Brothers Chisenga Sr, Nkweto, Mutende, Chewe, Kalunga, Chisenga Jr, Kunda Jr, & Mwape

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Human Cytomegalovirus (HCMV), classified as Human betaherpesvirus 5 (HHV-5) (ICTV, 2016), is the leading infectious cause of birth defects including intrauterine growth restriction, stillbirth, low birthweight, preterm birth, microcephaly, neurodevelopmental delay, cerebral palsy, haematological disorders, pneumonitis, blindness, and sensorineural hearing loss (Cheeran et al., 2009, Iwasenko et al., 2011, Fowler et al., 1992, Noyola et al., 2001, Pass et al., 2006, McClure et al., 2010, Pereira et al., 2014, Boppana and Fowler, 2007, Boppana et al., 1999, Griffiths and Walter, 2005). Although HCMV infection is generally clinically silent in immunocompetent hosts, primary infection in young adults can manifest as a transient flu-like mononucleosis illness. In individuals with immature or compromised immunity – neonates, the HIV/AIDS infected, and those on long-term immunosuppressive treatments, such as transplant recipients – HCMV infection often results in clinically overt disease. The worst outcomes, including mortality and long-term morbidity, are with congenital infection, which however remains challenging to detect and mitigate early, since only up to about 10% of congenitally infected neonates are symptomatic at birth. Nevertheless, 15% of asymptomatic congenitally infected infants go on to develop long-term sequelae including progressive hearing loss and learning difficulties in later childhood (Kovacs et al., 1999, Rosenthal et al., 2009, Ross and Boppana, 2005). The gravity of these deleterious effects is compounded by HIV-1 immunosuppression, which enables new and reactivated HCMV infections to cause HIV/AIDS-associated disease, particularly in the absence of antiretroviral therapy. This is of particular relevance to Zambia, a country where HIV-1 is endemic. In HIV-infected infants, HCMV coinfection has been shown to accelerate progression to AIDS and to exacerbate neurological disease, leading to developmental delay and motor deficit (Kovacs et al., 1999, Doyle et al., 1996). Other studies, including from our laboratory, indicate highly prevalent infant HCMV infections and severe effects of HCMV and HIV-1 co-infection on child health, growth, development, and survival. The detrimental effects of HCMV have also been demonstrated in maternally HIV-exposed uninfected (HEU) infants, a growing cohort

with the largely successful implementation of prevention of mother-to-child HIV transmission (PMTCT) programs and widespread use of Highly Active Antiretroviral Therapy (HAART) in HIV-endemic regions (Bates et al., 2008, Gompels et al., 2012, Sanz-Ramos et al., 2013, Gantt et al., 2008, Slyker et al., 2009a, van der Sande et al., 2007).

In the Chilenje Infant Growth Nutrition and Infection Study (CIGNIS), a nutrition intervention study conducted in Zambia (ISRCTN37460449, http://www.controlled-trials.com/mrct), our laboratory showed that 83% of infants are HCMV seropositive by 18 months age in Zambia (Bates et al., 2008). This is similar in some developing countries but higher than in many global regions (van der Sande et al., 2007, Cannon et al., 2010, Dowd et al., 2009, Vyse et al., 2009, Bate et al., 2010). During the CIGNIS trial (Gompels et al., 2012), stunting was three times more prevalent among infants who were seropositive for HCMV compared to those seronegative. Those who were also maternally HIV-exposed had reduced head size and scored lower in psychomotor development tests. Further, HIV-exposed HCMV-viremic infants had more frequent hospital referrals compared to their HIV-unexposed peers. Immunological responses to routine childhood vaccines were assessed in the CIGNIS cohort, and sadly, HIV-exposed infants had weaker short-lived responses, which worsened if they were also infected with HCMV (Sanz-Ramos et al., 2013). These HCMV effects remained unaltered by dietary intervention. Since children were recruited to CIGNIS at 6 months age, the route of HCMV acquisition – whether congenital or postnatal – was not clear in this cohort. However this is important to determine so that appropriate, targeted interventions can be formulated to mitigate the detrimental HCMV effects highlighted above. With 40% (226/562) infants having HCMV DNAemia at recruitment to CIGNIS, it was unlikely that the observed effects could be attributed solely to congenital HCMV infection. A role for postnatal infection is thus implied, and investigated further in this thesis via analysis of breast milk. Despite the highlighted possible serious effects of HCMV, there are currently no routine diagnoses established for its detection in Zambia. This thesis therefore also trialled the use of non-invasive samples including saliva and breast milk, in molecular diagnoses of HCMV, for possible future application in the local setting.

1.2 VIRUS CLASSIFICATION AND MORPHOLOGY

With the recent recognition of human herpesvirus 6A (HHV-6A) and human herpesvirus 6B (HHV-6B) as distinct virus species (Ablashi et al., 2013, Adams and Carstens, 2012), nine distinct herpesviruses, including HCMV, are currently known to infect man. These viruses are grouped under the human Herpesviridae family, and on the basis of replication and latency, are further classified into three subfamilies – the Alpha, Beta, and Gammaherpesvirinae (ICTV, 2016). The Alphaherpesvirinae are mostly neurotropic, with a relatively short replication cycle, quick spread, and efficient cell lysis in culture. They exhibit a relatively broad host species range and establish latency primarily in neuronal ganglia. Members include herpes simplex virus types 1 and 2 (HSV-1 and HSV-2) and varicella-zoster virus (VZV) (Roizman and Pellett, 2007). The Betaherpesvirinae, the topic of this thesis, have a rather restricted host range and long replication cycles. Infection of cell cultures is slow, and infected cells become enlarged. Viruses establish latency in haematopoietic cells in the bone marrow (Kondo et al., 1994, Kondo et al., 1991, Reeves and Sinclair, 2008, Sinclair, 2008, Sinclair and Sissons, 2006, Roizman and Pellett, 2007). Members include HCMV, HHV-6A, HHV-6B, and human herpesvirus 7 (HHV-7). Finally, subfamily Gammaherpesvirinae encompasses the lymphotropic Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV), which infect and establish latency in lymphocytes. Both EBV and KSHV are oncogenic viruses (ICTV, 2016, Roizman and Pellett, 2007).

The main characteristic shared by members of the Herpesviridae family is the ability to remain in a latent or persistent state for the lifetime of the host, with reactivations occurring periodically. Herpes viruses are distinctively large, double-stranded DNA viruses belonging to Baltimore group I (Baltimore, 1971). Virions are 150–300nm spherical to pleomorphic particles that possess a 240kbp monopartite double stranded linear DNA genome. The genome is packaged together with a lytic origin of DNA replication (oriLyt)-associated RNA in an icosahedral capsid, which is encased by a thick protein-rich amorphous tegument, in turn surrounded by a lipid bi-layer membrane, studded with glycoproteins (Figure 1.1A).

Apart from encoding structural proteins, the HCMV genome encodes numerous non-structural proteins and non-coding RNAs that regulate the complex viral replication cycle

and modulate interactions with the host's immune system. The capsid is 110-130nm in diameter and constructed from mainly five herpesvirus core proteins: major capsid protein (MCP, the UL86 gene product), minor capsid protein (TRI1, the UL46 gene product), minor capsid protein binding protein (TRI2, the UL86 gene product); the smallest capsid protein (SCP, the UL48A gene product), and portal protein (PORT, the UL104 gene product). These proteins are organised into 162 capsomeres (150 hexamers and 12 pentamers), with 320 triplexes between the capsomeres. A portal complex (composed of the PORT protein) occupies one of the 12 pentamer positions, and facilitates viral genome entry into / exit from the capsid. Surrounding the capsid is the approximately 50nm thick tegument, which includes in excess of 27 relatively abundant virus-encoded proteins, plus a selection of cellular and viral RNAs. The tegument is clothed by a 10nm-thick lipid bilayer envelope derived from the host cell endoplasmic reticulum-Golgi intermediate compartment (ERGIC). The virion envelope contains no less than 20 virus-encoded glycoproteins, which serve various functions including adsorption to host cells. (Gibson, 2008, Mocarski Jr and Pass, 2008, Mocarski Jr et al., 2007). With the envelope peeled back, negatively stained virions take on a characteristic 'fried egg' appearance on electron micrographs (figure 1.1B).

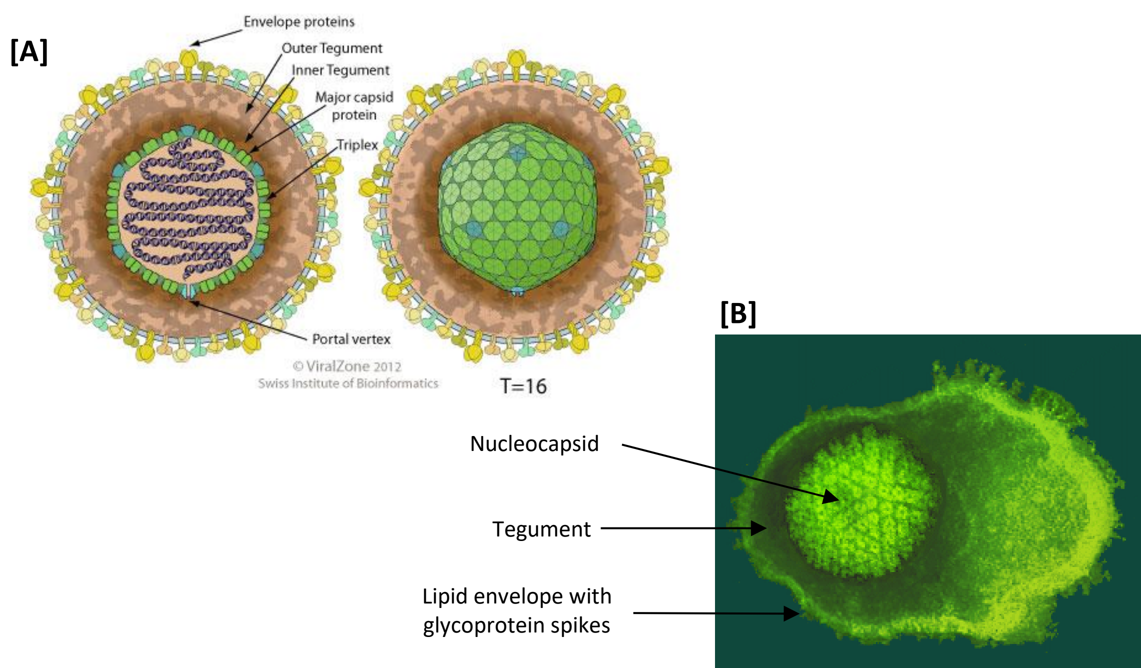


Figure 1.1. Herpesvirus ultrastructure. [A] Key components of Cytomegalovirus virions. Image available online at www.viralzone.expasy.org/all_by_species/180.html [B] Electron micrograph of negatively stained virion showing the typical 'fried egg' appearance. Image Copyright Dr Linda M Stannard, 1995, University of Cape Town (accessed at http://www.virology.uct.ac.za/vir/teaching/linda-stannard/herpesvirus).

A brief historical note …

The earliest histopathological evidence of the characteristic intranuclear inclusions (figure 1.2) typical of cytomegalovirus infection was by German pathologist Hugo Ribbert, who reported his observations in 1881 of enlarged cells with eccentrically placed nuclei that contained a "central nuclear body" surrounded by a clear halo in kidney and parotid gland sections taken from demised syphilitic infants. Later these intranuclear inclusions were also observed in lung and liver cells.

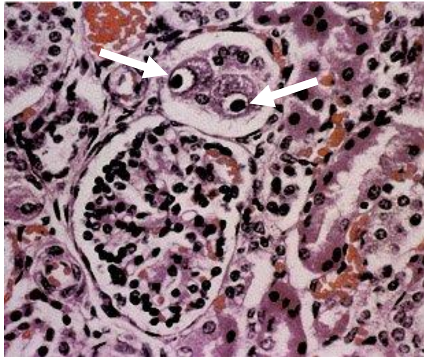


Figure 1.2. Histological section of kidney showing HCMV-infected cells with the typical 'Owl's eye' appearance due to intranuclear inclusions (arrows). Image reproduced with permission from *Herriot R. (1994) NEJM Vol 331:649 No.10* (Herriot and Gray, 1994), Copyright Massachusetts Medical Society.

At that time, and for much of the early 20th century when more reports of similar histopathological observations continued to emerge from other scientists, the inclusion bodies were attributed to syphilis or protozoa. However, in 1921 Goodpasture and Talbert suggested that a viral agent could cause these "cytomegalic inclusions", and with the independent finding of the inclusions in herpes zoster and herpes genitalis lesions, doubt was cast on a protozoal cause and the suggestion of a viral link strengthened. This was further consolidated after the development of electron microscopy in the 1930s. Because the pathological changes in the liver and kidney were noted to occur in the absence of inflammatory signs, Goodpasture and Talbert postulated that the causal agent could be blood-borne, and in 1950 Smith and Vellios demonstrated that the infection could even be spread *in utero*. The establishment of tissue culture systems in the 1950s finally led to the isolation of HCMV from salivary glands independently by three research teams led by Smith, Rowe, and Weller. Weller subsequently proposed the term "cytomegalovirus" in 1960 (Ho, 2008, Plachter et al., 1996, Riley, 1997). Since then the virus has been isolated from every human population group and continues to be topical in the medical research community. HCMV is increasingly recognized as a critically important, complex, and highly successful pathogen.

1.3 GENOME ORGANISATION AND CONTENT

Based on a scheme proposed by Roizman and Pellet (Roizman and Pellett, 2001), herpesvirus genome architecture can be divided into six classes, A to F (figure 1.3), with typical representatives HHV-6B, herpesvirus saimiri, EBV, VZV, HCMV, and tupaia herpesvirus, respectively. The linear, double-stranded HCMV DNA genome comprises two covalently linked regions, long unique (U_L) and short unique (U_S), each flanked by inverted repeats [Long Terminal Repeat (TR_L) and Long Internal Repeat (IR_L), and Short Terminal Repeat (TR_S) and Short Internal Repeat (IR_S) respectively]. This gives the overall genome configuration TR_L–U_L–IR_L–IR_S–U_S–TR_S (Davison et al., 2003b, Dolan et al., 2004, Mocarski Jr et al., 2007), typical of class E herpesviral genomes. Additionally, the HCMV genome possesses a short region, termed the *a* sequence, as a direct repeat at the termini and also in inverse orientation at the IR_L–IR_S junction (figure 1.3) (Davison et al., 2003b). The direct terminal repeats contain cis-acting signals for cleavage and packaging of progeny genomes during replication. Because U_L and U_S can invert relative to each other by recombination between inverted repeats, four equimolar genome arrangements are possible in virion DNA (figure 1.4).

HCMV is the largest and most complex of the nine human herpesviruses. The DNA sequence of strain AD169 was the first complete HCMV genome to be published (Chee et al., 1990). Revisions have since been made to the initial AD169 genome sequence, and in recent years full genome sequences for several other HCMV strains – from both laboratory-adapted viruses and viruses derived from pathological specimens – have been published. Currently, the strain Merlin (GenBank Accession number AY446894.2), with a genome size of 235,646bp, is widely regarded as the consensus HCMV reference genome. Wild type HCMV strain Merlin contains at least 170 protein-coding genes (Dolan et al., 2004, Gatherer et al., 2011). With technological advancement, the coding potential of HCMV is beginning to be understood better. Recent high-definition analyses of the HCMV transcriptome using RNA-Seq, RNA mapping, and ribosome profiling have revealed that gene expression in HCMV is even more complex than previously thought, in part owing to highly regulated non-canonical translation patterns, overlapping ORFs, and use of non-conventional initiation codons (Gatherer et al., 2011, Stern-Ginossar et al., 2012). One study has identified over 600 additional protein coding ORFs, the majority

of which are very short and situated upstream of longer ORFs (Stern-Ginossar et al., 2012).

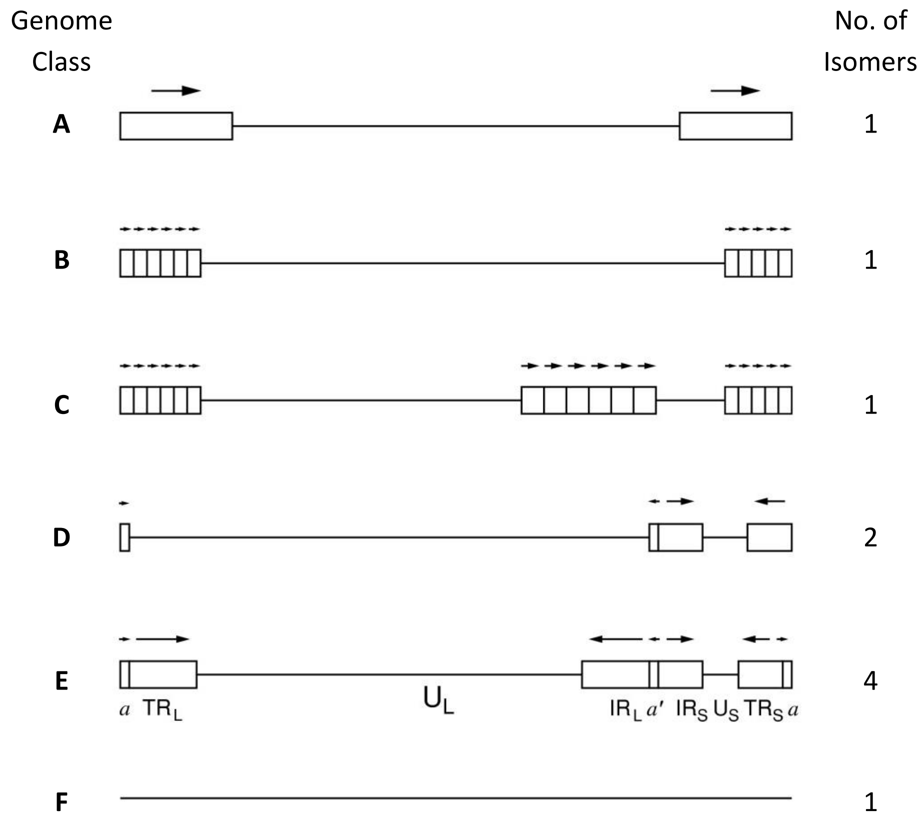


Figure 1.3. The six classes of herpesvirus genomes, according to Roizman and Pellet (Roizman and Pellett, 2001). Horizontal lines represent unique / quasiunique regions, while rectangles depict reiterated domains. Arrows indicate the orientations of repeats. The nomenclature of unique and repeat regions, including the terminal redundancy (a) and its internal inverted copy (a'), is indicated for the class E genome. Image obtained from https://www.ncbi.nlm.nih.gov/books/NBK47439/ (Davison, 2007).

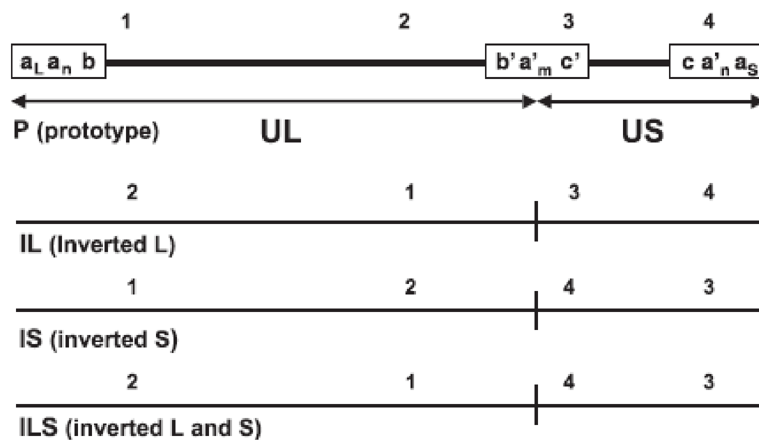


Figure 1.4. Structure of the four HCMV genome isomers produced by inversion of UL and US regions (Landolfo et al., 2003).

In addition to protein coding genes, HCMV also produces polyadenylated non-coding RNAs, including four long non-coding RNAs – RNA2.7, RNA1.2, RNA4.9, and RNA5.0 – that are produced in abundance and do not overlap with protein coding regions. Additional non-coding RNAs are transcribed anti-sense of protein coding regions (Gatherer et al., 2011, Zhang et al., 2007). Non-polyadenylated HCMV RNAs include microRNAs, which are involved in numerous processes related to regulating host cell metabolism, immune evasion, and maintenance of latency (Dhuruvasan et al., 2011, Grey et al., 2007, Stern-Ginossar et al., 2007, Stern-Ginossar et al., 2009, Lau et al., 2016a, Lau et al., 2016b, Poole et al., 2011, Shen et al., 2014).

Overall, there are 40 core genes conserved across the human herpesvirus family. No additional genes are shared between HCMV and the alphaherpesviruses, but at least 7 more genes are shared with gammaherpesviruses, and 27 are betaherpesvirus specific. (Mocarski Jr et al., 2007, Dolan et al., 2004, Murphy et al., 2003, Roizman and Pellett, 2007, Davison et al., 2003b, Mocarski Jr, 2007). The number of genes shared across herpesvirus subfamilies is subject to revision as HCMV coding potential and gene function become clearer. Based on relatedness, HCMV non-core genes are grouped into at least 13 gene families (figure 1.5), many of which contain duplicated genes arranged as 'gene accordions'. This level of gene organisation is one of the several strategies that enable HCMV to rapidly adapt and counteract host defences, and has also been demonstrated in other DNA viruses such as poxviruses (Elde et al., 2012). Functionally, HCMV genes can be broadly categorised into two classes: core genes, which encode virus structural components and key proteins involved in replication, and accessory genes encoding modulators of virus-host interaction. Unlike core genes, accessory genes are more divergent and therefore facilitate virus diversity at various levels including cellular tropism, immune evasion, and latency adaptations.

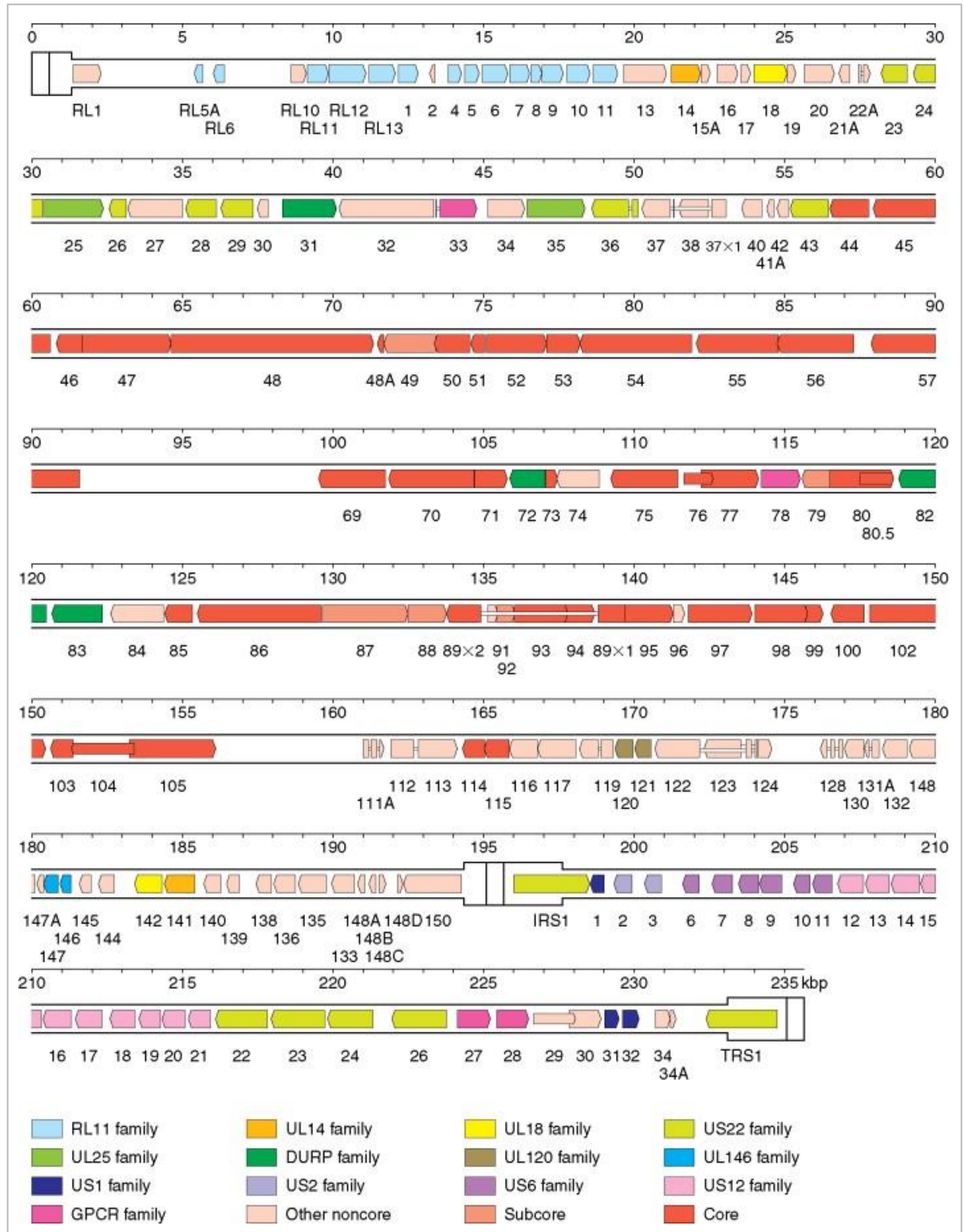


Figure 1.5. Genetic organization and content of wild-type HCMV, based on strain Merlin, sequenced from the original clinical source material (urine). Image adapted from Dolan and co-workers (Dolan et al., 2004).

1.4 REPLICATION CYCLE

HCMV can infect various cell types including fibroblasts, endothelial cells, epithelial cells and cells of the monocyte-myeloid lineage (Arrode and Davrinche, 2003, DuRose et al., 2012, O'Connor and Shenk, 2012, Bayer et al., 2013, Mocarski Jr et al., 2007). Compared to other herpesviruses, HCMV replicates relatively slowly, with production of progeny viruses peaking 72-96 hours post infection (Mocarski Jr et al., 2007). Virus-driven modulation of host cell gene expression, immunity, and cell survival is crucial for successful progression of the viral replication cycle, but the virus also has to concurrently regulate its own gene expression in a stringently temporal manner.

The first stage in the complex HCMV replication cycle involves entry into a susceptible host cell, and includes several distinct steps from attachment to specific cell surface receptors to release of the viral genome into the nucleus (figure 1.6). Three distinct viral envelope glycoprotein complexes, gCI, gCII, and gCIII, are involved in recognition and cellular uptake of HCMV. gCI is composed of glycoprotein gB (UL55 product), while gCII is composed of glycoproteins gM (UL100 product) and gN (UL73 product). Two forms of gCIII are known to exist: a trimeric complex composed of gH (UL75 product), gL (UL115 product), and gO (UL74 product), and a pentameric complex formed by gH, gL, pUL128, pUL130, and pUL131A (products of the UL128 locus). gCI and gCII mediate the initial binding and adsorption of HCMV to host-cell heparan sulfate glycosaminoglycans (Compton et al., 1993, Kari and Gehrz, 1992, Britt and Boppana, 2004). At the start of this thesis, additional binding of gCI to integrins (Feire et al., 2004), epidermal growth factor receptor (EGFR) (Wang et al., 2003, Chan et al., 2009), and platelet-derived growth factor receptor α (PDGFRα) (Soroceanu et al., 2008) had been shown to occur, although some other studies reported conflicting findings (Isaacson et al., 2007, Vanarsdall et al., 2012). Virus adsorption is followed by a fusion step, involving gCI and gCIII (Vanarsdall et al., 2008, Wille et al., 2013). In fibroblasts, this involves the trimeric gH/gL/gO form of the gCIII complex, followed by direct membrane fusion. In epithelial, endothelial and myeloid cells however, virus entry necessarily requires the pentameric gH/gL/pUL128/pUL130/pUL131 gCIII complex, with subsequent pH-mediated endocytosis (Ryckman et al., 2006, Ryckman et al., 2008b, Wang and Shenk, 2005b, Murrell et al., 2013, Hahn et al., 2004). Together with RL13, UL128 is among a select set of genes that rapidly mutate upon culture of virus in fibroblasts; this mutation appears

vital for optimal viral replication in fibroblasts (Stanton et al., 2010). The gCIII complex is therefore a major determinant for virus cell tropism (Scrivano et al., 2011, Murrell et al., 2013). After delivery of the nucleocapsid into the cytoplasm, it is trafficked via microtubules to the nucleus, into which the viral DNA is released (Dohner and Sodeik, 2005). The large tegument protein (UL48 product) and large tegument protein binding protein (UL47 product) are key in nucleocapsid translocation, uncoating, and viral genome release (Dunn et al., 2003).

In the nucleus, HCMV genes are transcribed in a temporal fashion, starting with immediate-early (IE), then early (E), delayed-early (DE), and finally late (L) transcripts (Mocarski Jr and Pass, 2008, Mocarski Jr et al., 2007). Genes UL36/UL37, UL122/UL123, TRS1/IRS1, and US3 encode numerous transcripts and products involved in IE gene regulation and evasion of cellular immune mechanisms to guarantee progression of replication (Mocarski Jr et al., 2007). Immune evasion mechanisms include: IE1 (UL123 product) suppression of STAT signalling, with consequent blockade of IFN activation; cell death suppression by viral inhibitor of caspase 8 activation (vICA) and viral mitochondrial inhibitor of apoptosis (vMIA), encoded by UL36 and UL37 respectively; inhibition of IFN-inducible protein kinase R (PKR) by IRS1 and TRS1 products; and, US3-induced disruption of MHC class I antigen presentation, making infected cells less attractive to cytotoxic T-cells (Mocarski Jr et al., 2007).

Figure 1.6 summarises the subsequent steps and viral products involved in the HCMV lytic replication cycle, leading to the production of progeny virions.

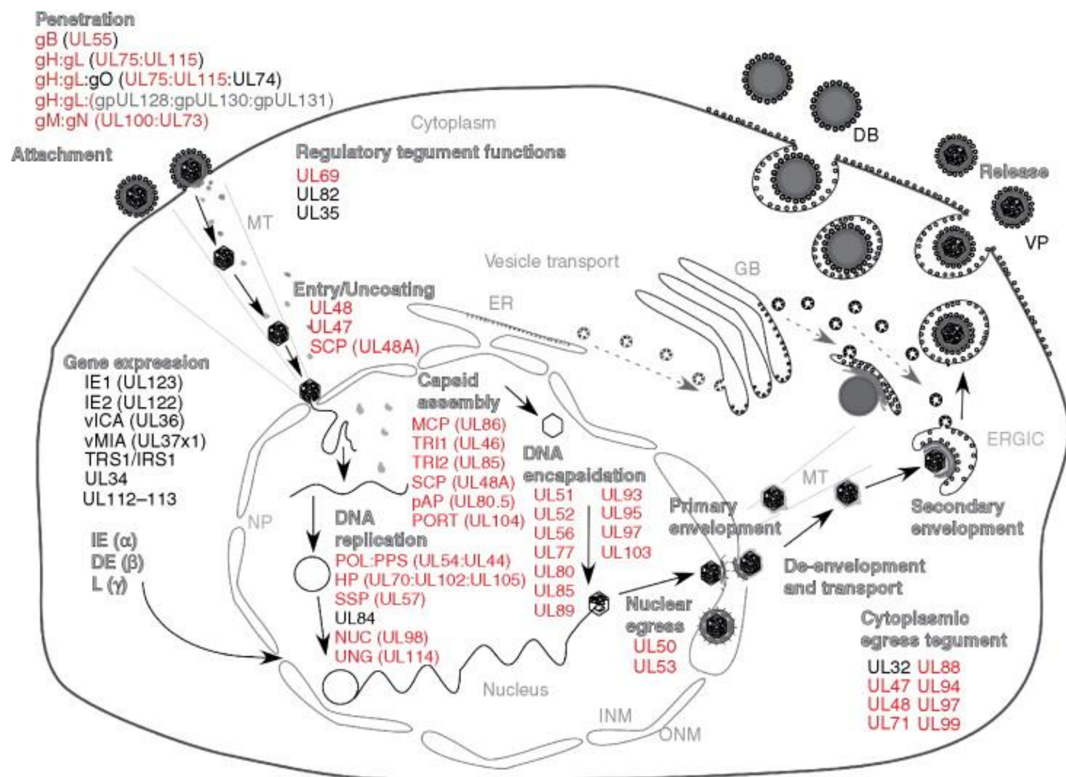


Figure 1.6. Summary of HCMV replication (Mocarski Jr and Pass, 2008, Mocarski Jr et al., 2007)**.** Infectious virions enter the cell via interaction of viral envelope glycoprotein complexes with cellular receptors. Following membrane fusion (fibroblasts) or endocytosis (epithelial & endothelial cells), tegument and capsid proteins are delivered to the cytosol, and after delivery into the nucleus, the genome is circularized. Host cell responses are regulated by tegument proteins, which also initiate the temporal expression of viral immediate early (IE) followed by delayed early (DE) genes, that initiate viral genome replication, and late (L) gene expression. Late gene expression gives rise to capsid assembly, followed by nuclear egress to the cytosol where capsids associate with tegument proteins and are trafficked to endoplasmic reticulum (ER) / Golgi assembly complexes. Capsids then acquire tegument and viral envelope components. Finally, enveloped infectious virus particles (VP) and non-infectious dense bodies (DB) are released. MT, microtubules; ERGIC, endoplasmic reticulum and Golgi intermediate compartment; NP, nuclear pores; INM, inner nuclear membrane; ONM, outer nuclear membrane.

1.5 HCMV LATENCY AND REACTIVATION

The ability to establish latency is a defining characteristic of herpesviruses. HCMV can setup and maintain latent infection predominantly in CD14+ monocytes and CD34+ progenitor cells. During the latency state, only a restricted set of transcripts is expressed and no virion production occurs (Reeves and Sinclair, 2010, Poole et al., 2013, Goodrum et al., 2007, Cheung et al., 2006, Hargett and Shenk, 2010, Goodrum et al., 2002, Reeves et al., 2005a, Reeves et al., 2005b, Rossetto et al., 2013, Hahn et al., 1998, Mendelson et al., 1996). Remodelling of chromatin around the viral major immediate early promoter (MIEP) to a heavily repressive state is a main driver of suppression of lytic transcription and maintenance of latent infection (Reeves et al., 2005a, Reeves et al., 2005b, Reeves, 2011, Murphy et al., 2002, Reeves et al., 2006). This state is enhanced by a deficit of viral activators (such as pp71), or in the presence of latency-associated repressors (such as HCMV RNA 4.9), or if cellular repressors of MIEP dominate (Saffert et al., 2010, Rossetto et al., 2013, Wright et al., 2005, Noriega et al., 2014). Latency is maintained by various cellular and virally encoded factors including the cellular repressor complex hDaxx:ATRX (Reeves et al., 2010, Woodhall et al., 2006, Saffert and Kalejta, 2007, Saffert and Kalejta, 2006, Lukashchuk et al., 2008); the latency unique nuclear antigen (LUNA), derived from the UL81-UL82 antisense transcript (Keyes et al., 2012, Bego et al., 2011),; histone modifying enzymes such as histone methyltransferases (HMT) and histone deacetylases (HDACs) (Murphy et al., 2002); and various microRNAs. Virus reactivation from the latent state enables resumption of the lytic replication program, and is driven by various physiological and external stimuli that include cellular differentiation, conditions of stress, and drugs. Differentiation of progenitor cells into macrophages or dendritic cells alters the nuclear environment leading to changes in the conformation of chromatin and activation of the MIEP and the lytic replication program (Reeves et al., 2005a, Reeves et al., 2005b, Poole et al., 2014, Reeves, 2011) (figure 1.7). This can result in virus dissemination and clinically overt disease, particularly in the absence of robust host immunity. Histone deacetylase inhibitors such as valproic acid, trichostatin A, and MC1568 have been shown experimentally to alleviate the repressive chromatin structure, allowing IE gene expression, at least transiently (Murphy et al., 2002). LUNA has also been shown to be in fact necessary for both the establishment of and successful rescue from the latency state (Keyes et al., 2012). The ability of HCMV to reactivate is a critical characteristic that gives rise to development of pathology at

individual level, and enables sustenance of virus circulation at population level. Much more remains to be understood regarding the precise mechanisms that regulate establishment and maintenance of HCMV latency in myeloid-lineage cells, as well as the interplay of factors governing reactivation from the latency state. Understanding these mechanisms is crucial for the identification of novel therapeutic targets and development of strategies to eliminate virus from latent reservoirs.

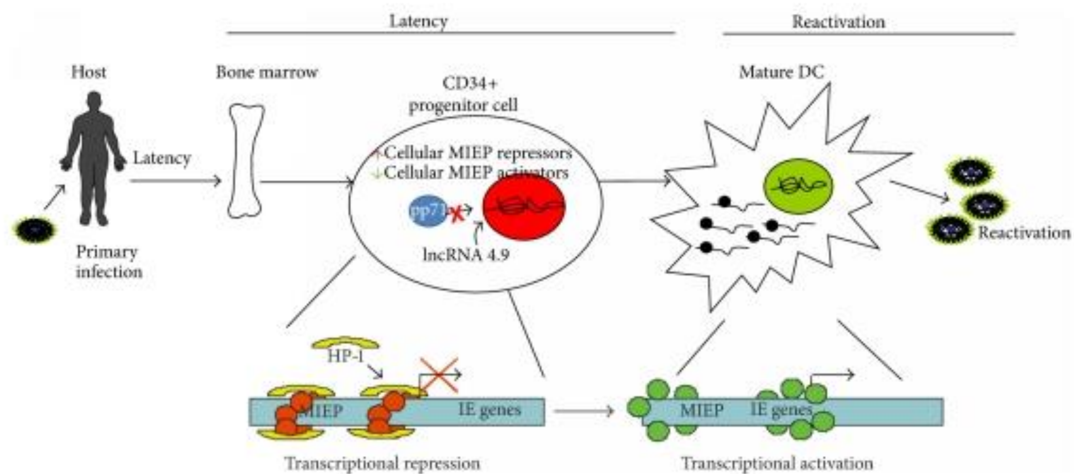


Figure 1.7. HCMV latency and reactivation. After primary HCMV infection, latency is established in bone marrow CD34+myeloid progenitor cells. In these cells, the viral genome is associated with repressive chromatin markers. Cellular differentiation into dendritic cells (DCs) is accompanied by chromatin remodelling, which enables the MIEP to associate with transcription activators, with subsequent induction of immediate early gene expression and reactivation of the virus (Poole et al., 2014).

1.6 VIRUS GENE VARIATION AND STRAIN IDENTIFICATION

In HCMV, gene variability appears to be restricted to particular genome locations, mostly near the ends of U_L and U_S regions (Dolan et al., 2004). Nucleotide variation in genes involved in host cell entry, tissue tropism, or replication has long been thought to contribute to apparent differences in virulence among HCMV strains, and is a risk factor for subsequent infection. However no clear link has yet been shown between a particular HCMV genotype and clinical disease (Kari et al., 1986, Pignatelli and Dal Monte, 2009, Pignatelli et al., 2004, Pignatelli et al., 2010, Puchhammer-Stockl and Gorzer, 2011). HCMV encodes in excess of 20 membrane glycoproteins, including gB, gH, gL, gM, gN, and gO. Variation in these individual glycoproteins and the complexes

they form has been of interest because of their key role as determinants of virulence and tissue tropism, as exemplified by the gH/gL complex (gCIII), which forms two different heterologous complexes (Ryckman et al., 2008a, Ryckman et al., 2008b, Rasmussen et al., 2002, Mocarski Jr, 2007, Mocarski Jr et al., 2007). Other hypervariable HCMV surface glycoproteins that have become the subject of intense study include gO and gN. The genes encoding these two glycoproteins are located at the centre of the genome, and eight genotypes have been identified for each, with evidence for linkage (See Table1.1). In both gO and gN variability is highest toward the N-terminus (Bates et al., 2008, Dolan et al., 2004, Mattes et al., 2005, Mattick et al., 2004, Pignatelli and Dal Monte, 2009, Stanton et al., 2005). Because of its hypervariability and critical role in tissue tropism via the trimeric gCIII complex, gO is of particular interest in genotyping studies in this thesis. Other hypervariable genes in HCMV include UL139, which encodes a putative membrane glycoprotein, and for which eight genotypes are defined; and UL146, encoding a viral chemokine (vCXCL1), which has at least 14 distinct genotypes (summarised with gO genotypes in Table 1.1) (Bradley et al., 2008, Kenneson and Cannon, 2007, Hassan-Walker et al., 2004, Lurain et al., 2006, Dolan et al., 2004). Both UL139 and UL146 are components of the so called 'virulence region' – a 13-15kb region that is deleted or modified in laboratory-adapted HCMV strains and encodes at least 20 genes, many of which are involved in immunomodulation (Cha et al., 1996, Murphy et al., 2003, Dolan et al., 2004). High variability also exists at the left end of U_L, particularly in the RL11 family of genes (Davison et al., 2003a, Sekulin et al., 2007, Dolan et al., 2004). One RL11 member, RL13, mutates rapidly in fibroblast cultures and is therefore linked to a role in tissue tropism (Dargan et al., 2010, Stanton et al., 2010). Understanding the extent of these hypervariable genes is essential in diagnosing strain variants, their transmission, and disease associations. It is also necessary to investigate fully whether there may be restriction of strains in the various body compartments.

1.7 EPIDEMIOLOGY AND TRANSMISSION

HCMV is a ubiquitous virus found in every human population group. Globally, seroprevalence ranges from 60 to 100%. Infection shows no seasonality, but seroprevalence varies geographically and with socioeconomic status, gender, and age. The virus is acquired in early life in most populations except in the more economically

developed northern European and North American regions. Seroprevalence thus increases with age, is higher in developing countries, and inversely related to socioeconomic status (Lopo et al., 2011, Aarnisalo et al., 2003, de Ory et al., 2004, Kothari et al., 2002, Hecker et al., 2004, Adjei et al., 2006, Cannon et al., 2010, Dowd et al., 2009, van der Sande et al., 2007, Vyse et al., 2009, Crough and Khanna, 2009, Staras et al., 2008a, Staras et al., 2008b, Bate et al., 2010).

Genotypes		Reference strain	GenBank Accession No.
gO	**gN linkage**		
gO1a	gN1	AD169	FJ527563.1
gO1b	gN3a	TR	KF021605.1
gO1c	gN4c	Toledo	GU937742.1
gO2a	gN3b	HAN38	GQ396662.1
gO2b	gN2	AF1	GU179291.1
gO3	gN4a	HAN16	JX512204.1
gO4	gN4b	Towne	FJ616285.1
gO5	gN4d	Merlin	AY446894.2
UL139			
G1		Merlin	AY446894.2
G2		TR	KF021605.1
G3		HAN38	GQ396662.1
G4		Toledo	GU937742.1
G5		Towne	FJ616285.1
G6		3301	GQ466044.1
G7		AD169	FJ527563.1
G8		AF1	GU179291.1
UL146			
G1		Toledo	GU937742.1
G2		Merlin	AY446894.2
G3		KSG	AY446889.1
G4		AF1	GU179291.1
G5		Davis	JX512198.1
G6		ML1	AY446880.1
G7		Towne	FJ616285.1
G8		TB40/E	KF297339.1
G9		AD169	FJ527563.1
G10		JP	GQ221975.1
G11		JHC	HQ380895.1
G12		6397	JX512197.1
G13		HAN8	JX512202.1
G14		HAN1	JX512199.1

Table 1.1. Genotypes of selected hypervariable HCMV genes. There are eight genotypes each for gO and gN, and they are linked as shown in this table (Bates et al., 2008, Dolan et al., 2004, Mattes et al., 2005, Mattick et al., 2004, Pignatelli and Dal Monte, 2009, Stanton et al., 2005). UL139 has eight genotypes and 14 genotypes are known for UL146 (Bradley et al., 2008, Kenneson and Cannon, 2007, Hassan-Walker et al., 2004, Lurain et al., 2006, Dolan et al., 2004).

In some countries, high HCMV seroprevalence has been found in women attending sexual health clinics and in young homosexual men (Crough and Khanna, 2009, Dowd et al., 2009, Boppana and Fowler, 2007, Chandler et al., 1985). In the USA, higher seroprevalence was found in African Americans and Hispanics compared to Caucasians (Boppana and Fowler, 2007, Crough and Khanna, 2009, Dowd et al., 2009).

HCMV is transmitted through close personal contact via infected bodily fluids including oropharyngeal secretions / saliva, urine, cervical and vaginal secretions, semen, breast milk, tears, blood products, and allografts (Boppana and Fowler, 2007, Cannon et al., 2011). Vertical transmission of HCMV may occur transplacentally / *in utero*, during parturition, or via breastfeeding (see section 1.8). After early infancy, young children acquire HCMV through horizontal transmission from other children or indirectly through environmental contamination. This is evidenced by the finding of higher HCMV seropositivity among children who attend day care centres compared to those who do not. Further, families living in crowded conditions also show early acquisition of the virus. Young children shed virus in saliva and urine, thereby facilitating virus exposure to other children, their parents, and to the day-care workers (Adler et al., 2004, Noyola et al., 2000, Boppana and Fowler, 2007). Although both children and adults can shed HCMV for several months following primary infection, children are particularly important as reservoirs of infection in communities. Some studies have demonstrated that children continue to shed virus in urine for as long as 9 years (Adler et al., 2004, Noyola et al., 2000). Women of childbearing age and immunocompromised individuals are particularly at high risk for infection (Boppana and Fowler, 2007). In adolescence and early adulthood, HCMV rates continue to increase and this is attributed in part to increased sexual contact. The finding of HCMV in cervical and vaginal secretions and semen support the categorisation of HCMV as a sexually transmitted infection (Crough and Khanna, 2009, Chandler et al., 1985, Sohn et al., 1991). Risk factors for HCMV seropositivity in studies of sexual health clinic attendees include young age of sexual debut, recent / new sexual partners, multiple sexual partners, lack of barrier contraception, history of STIs, chlamydial infection, female gender, and African American race (Boppana and Fowler, 2007, Chandler et al., 1985, Staras et al., 2008b). Other occasions of HCMV infection include iatrogenic settings such as blood transfusion or allograft transplantation. HCMV infection is associated with transfusion of cellular

components of blood but not fresh-frozen plasma or cryoprecipitate (Bowden and Sayers, 1990). Transfusion of unscreened cellular components results in 10%–70% HCMV transmission largely from reactivation of latent virus from myeloid lineage cells (Preiksaitis et al., 1988, Wilhelm et al., 1986). The risk of transfusion associated HCMV infection is directly related the number of components, particularly the quantity of leukocytes, transfused (Boppana and Fowler, 2007). Leucocyte depletion prior to use of blood products has proved to be a useful strategy to limit transfusion-transmitted HCMV. In the transplant setting, HCMV infection may occur due to transmission from the transplanted organ, due to reactivation of latent infection, or following primary infection in a seronegative transplant recipient. Development of HCMV disease depends on the type of transplant, the HCMV serological match between donor and recipient, the immunosuppressive drug regimen used, occurrence of acute tissue rejection, donor and/or recipient age, HLA mismatch, genetic polymorphisms, and other concurrent infections.

Chronic pathological conditions linked to HCMV

Evidence continues to emerge for a role of HCMV in a range of chronic inflammatory conditions, autoimmune disorders, and neoplastic lesions. This is postulated to arise partly from the long-term immune activation and chronic inflammatory state induced by the frequent, albeit subclinical, HCMV reactivations. Early acquisition of HCMV is thus seen as a risk factor for premature aging and reduced lifespan. HCMV has been implicated in gliomas, some types of bowel and breast cancer, cardiovascular diseases (including atherosclerosis and hypertension), immune senescence, and aging (Söderberg-Nauclér, 2008, Barami, 2010, Soroceanu and Cobbs, 2011, Hawkins and Croul, 2011, Lepiller et al., 2011, Bhattacharjee et al., 2012, Fonseca et al., 2012, Grahame-Clarke, 2005, Cheng et al., 2009, Bishop et al., 2015, Li et al., 2015, Pawelec, 2012, Pawelec et al., 2009, Strindhall et al., 2007, Derhovanessian et al., 2009, Stassen et al., 2006, Wikby et al., 2002, Roberts et al., 2010, Courivaud et al., 2013).

1.8 HCMV IN EARLY CHILDHOOD

HCMV commonly infects in childhood and then persists for life via establishing latent infections. The virus can be transmitted pre-, peri- or post-natally. Prenatal transmission

occurs transplacentally, more frequently during maternal primary infection, but can also occur during viral reactivation or reinfection with a different strain (Boppana and Fowler, 2007, Boppana et al., 1999, Boppana et al., 2001). Perinatal infection occurs through cervical and other genital fluids during parturition, or via milk during breastfeeding. Postnatally, HCMV infection occurs via infected bodily fluids, commonly saliva, urine, and breast milk. Children are a critical reservoir for HCMV and play a central role in transmission and maintenance of the virus within communities, as described in section 1.7 (Adler et al., 2004, Noyola et al., 2000).

1.8.1 CONGENITAL HCMV

HCMV is the most frequent cause of congenital infection, during which it can cause serious damage to the baby (Boppana and Fowler, 2007, Boppana et al., 1999). Prevalence of congenital HCMV can range from 0.2 to 5.4% (Manicklal et al., 2013, van der Sande et al., 2007, Dar et al., 2008, Manicklal et al., 2014), although one study found a prevalence of 29% in a limited sample of HIV-infected infants (Slyker et al., 2009a). Studies have shown that prevalence is highest, at around 1–5% of births, in populations with high maternal seroprevalence, commonly in the developing world, compared to developed countries with low maternal seroprevalence, where incidence is 0.2–1.8% of births (Fowler and Boppana, 2006, Fowler et al., 1999, Fowler et al., 2003, Kaye et al., 2008, Yamamoto et al., 2013, Yamamoto et al., 2010). HCMV remains the leading non-genetic cause of childhood hearing loss. The incidence of HCMV-related hearing loss among infants born in populations of high maternal seroprevalence is similar to or even higher than that reported in populations from developed countries (Boppana, 1999; 2001; Dar, 2008; Kenneson, 2007; Manicklal, 2013; Mussi-Pinhata, 2009; Sohn, 1992; van der Sande, 2007). Congenital HCMV infection has been shown to occur in up to 40% of cases following primary maternal infection during pregnancy (Liesnard et al., 2000, Revello and Gerna, 2002), while in non-primary maternal infection, it occurs in 2.2% of cases (Revello and Gerna, 2002, Fowler et al., 2003, Kenneson and Cannon, 2007), but can approach 20%, as reported in one Israeli study (Rahav et al., 2007). Congenital HCMV from non-primary maternal infection occurs if the mother has a reactivation of virus or gets re-infected with a new strain. HCMV is therefore unique, in that unlike all other known congenital pathogens, it can even be transmitted by women with pre-existing immunity (Mussi-Pinhata et al., 2009, Boppana and Fowler, 2007, Boppana et al., 1999,

Boppana et al., 2001). Although congenital infection leading to long-term sequelae can result from maternal primary infection during any stage of pregnancy, primary infection in the first trimester is associated with a higher rate of foetal damage (Foulon et al., 2008, Pass et al., 2006). However, transmission rates are reportedly lower in early compared to later stages of pregnancy (Revello and Gerna, 2004, Gindes et al., 2008, Enders et al., 2011, Daiminger et al., 2005, Bodeus et al., 2010). During delivery, neonatal HCMV infection can occur by virus shedding in the birth canal followed by ingestion of infected secretions by the neonate (Boppana and Fowler, 2007). Diagnosis of congenital HCMV requires virus isolation (Gold standard) or PCR using saliva, urine or other samples collected within two weeks of birth (Ross et al., 2011, de Vries et al., 2012). Beyond this two-week window, congenital infection cannot be reliably distinguished from postnatally acquired virus. Breastfeeding and intimate nursing can profoundly confound the diagnosis of congenital HCMV.

1.8.2 POSTNATAL HCMV INFECTION

Although breastfeeding has been recognised as a route for postnatal HCMV transmission for several decades (Hayes et al., 1972, Stagno et al., 1980, Schleiss, 2006b, Schleiss, 2006a), it is traditionally believed that HCMV transmitted through this route carries no major health risks except in very low birthweight preterm infants, where acute deleterious effects have been widely documented (Hamprecht et al., 2005, Hamprecht et al., 2008, Hamprecht et al., 2004b, Hamprecht et al., 2001, Maschmann et al., 2006, Jim et al., 2004). Nevertheless, long-term follow-up of preterm infants infected with HCMV via milk demonstrates that breastmilk-acquired HCMV infection has detrimental effects on intellectual development (Goelz et al., 2013). Possible severe life-threatening effects of breastmilk-transmitted HCMV in term babies have been highlighted by a recent German case of severe colitis (Novakova et al., 2014). Early studies of HCMV in breast milk in Africa suggest differences with data in other regions. In Europe and the far East, studies indicate HCMV is generally absent in colostrum and ‘early milk’, being only detectable in the milk from 2-4 weeks postpartum, with peak viral load around the fourth postpartum week, before waning and finally disappearing by week 12. It has been shown that HCMV in breast milk results from local virus reactivation within the mammary gland and that cell-free virus, frequently shed into the whey, is of greater importance for vertical transmission than the infrequently occurring virus within milk

cells (Asanuma et al., 1996, Hamprecht et al., 1998, Hamprecht et al., 2001, Jim et al., 2009). A Zimbabwean study showed that HCMV and EBV interact locally with HIV-1 within breast tissue, and associate with higher HIV-1 viral loads, thereby impacting on HIV-1 transmission via breast milk (Gantt et al., 2008). As described earlier in this chapter, widespread negative effects of HCMV on child health and development have also been demonstrated in Zambia where breastfeeding is widely practiced, including among otherwise healthy children, suggesting again more widespread pathological effects of postnatal infection (Gompels et al., 2012, Sanz-Ramos et al., 2013).

1.8.3 HCMV AND HIV/AIDS INFECTION IN CHILDHOOD

In the setting of HIV-1 infection, HCMV is the most common viral opportunistic infection, with severe sight- and life-threatening effects, especially in the absence of antiviral therapy. Children co-infected with HIV-1 and HCMV suffer greater morbidity and mortality compared to those infected with either virus alone (Kovacs et al., 1999, Slyker et al., 2009b). Concomitant HCMV in HIV-1 infection accelerates progression to AIDS (Kovacs et al., 1999, Slyker et al., 2009b, Marinda et al., 2007, Ross and Boppana, 2005). This phenomenon was also noted among haemophiliacs co-infected with HIV and HCMV (Webster et al., 1989). On the other hand, HIV-1 infection increases the likelihood of HCMV infection and/or reactivation (Slyker et al., 2009a, Slyker et al., 2009b). In Zambia, a lung necropsy study ranked HCMV among the top three causes of death among HIV-1 infected children dying of respiratory illness (Chintu et al., 2002). In the CIGNIS cohort, even maternal HIV-1 exposure without child infection increased the negative effects of HCMV on child psychomotor development, reduced infant growth in a third of the cohort, and markedly reduced immunological responses to the polio vaccine (Gompels et al., 2012, Sanz-Ramos et al., 2013).

Although infant HCMV infection was assessed in Zambia through the Chilenje Infant Growth and Nutrition Study (CIGNIS) (Sanz-Ramos et al., 2013, Gompels et al., 2012), neither the actual route of transmission (congenital or postnatal) was established, nor how maternal HIV affects this, and therefore is a key premise for studies in this thesis.

1.9 HCMV CONTROL

1.9.1 HYPERIMMUNE GLOBULIN

Passive immunization using anti-HCMV hyperimmune globulin (HIG) – an immunoglobulin preparation pooled from healthy blood donors with high titers – has been of interest as a potentially safe intervention, particularly in women at risk of transmitting HCMV during pregnancy, and as prophylaxis in solid organ transplant recipients. Active research is ongoing, but some early reports indicate beneficial outcomes. For instance, in a group of 'at risk' women who received HIG, there were significant reductions in placental pathology, markedly less incidences of congenital HCMV in neonates, and reversal of foetal cerebral abnormalities compared to peers who did not receive HIG (Nigro et al., 2005, Nigro et al., 2008, La Torre et al., 2006, Buxmann et al., 2012). Improved infant neurodevelopmental outcomes have also been associated with maternal use of HIG during pregnancy (Visentin et al., 2012).

1.9.2 DRUGS

There are four anti-HCMV drugs currently in use for prevention and treatment of HCMV infection/disease: Ganciclovir (GCV), Valganciclovir (vGCV), Cidofovir (CDV) and Foscarnet (FOS) (figure 1.8). These drugs have tremendously improved HCMV management and outcomes, particularly in adults in the transplantation setting. However, the usefulness of these antivirals is restricted by their low bioavailability (except for vGCV, which has good oral bioavailability), toxicities, and the emergence of resistance mutations. There are currently no suitable anti-HCMV drugs for neonates, although GCV and vGCV have been undergoing clinical trials in the context of symptomatic congenital infection, and have shown good efficacy in improving neurodevelopmental outcomes, particularly with longer (6 months versus 6 weeks) drug treatment (Oliver et al., 2009, Acosta et al., 2007, Kimberlin et al., 2003, Kimberlin et al., 2015, Nassetta et al., 2009).

Ganciclovir (GCV) and Valganciclovir (vGCV)

GCV (9-[1, 3-dyhydroxy-2-propoxymethyl] guanine) was the first antiviral approved for HCMV treatment in 1989 (Chou, 2008). GCV is a guanosine analogue and homologue of acyclovir (ACV). It is activated to its triphosphate form by sequential phosphorylation involving both viral and cellular kinases. UL97 (serine/threonine phosphokinase) and

UL54 (DNA polymerase) are two key viral molecules involved in GCV metabolism. In its active form as GCV triphosphate, the drug competitively inhibits viral DNA polymerisation, leading to chain termination. The requirement for hospitalization to administer the drug intravenously, and the associated need for intravenous catheterisation for six weeks of therapy limit GCV use. Furthermore, utility of GCV is hampered by the serious hematologic side effects that can occur, particularly neutropenia and thrombocytopenia, and by the development of drug resistance mutations (Biron, 2006, Chou, 2008). In the 1990s, vGCV, an orally administered L-valyl ester prodrug of GCV, was developed. Following oral administration vGCV is rapidly metabolized in the intestinal wall and liver and achieves up to 10 times the bioavailability of GCV. vGCV has a similar toxicity profile as GCV. Because of its convenient mode of administration, vGCV is now more widely used in the transplant setting (Biron, 2006, Len et al., 2008).

Foscarnet (FOS)

Foscarnet (FOS), generic name phosphonoformic acid, is a pyrophosphate analogue, used as second-line therapy in cases of GCV/vGCV drug-resistance. FOS does not require activation. It non-competitively inhibits HCMV DNA polymerase activity by reversibly blocking the pyrophosphate binding site, thereby inhibiting cleavage of pyrophosphate from deoxynucleoside triphosphates. The main adverse effects associated with FOS are nephrotoxicity, metabolic toxicity, and electrolyte imbalance (hypocalcemia, hypomagnesemia and hypophosphatemia). FOS is particularly useful in myelosuppressed patients, in haematopoietic stem cell transplant (HSCT) recipients with weak grafts, in HIV/AIDS associated HCMV retinitis, and in those with leucopenia. FOS has also been used as maintenance therapy to prevent HCMV relapse or progression (Biron, 2006, Tan, 2014).

Cidofovir (CDV)

CDV ([S]-1-[3-hydroxy-2-phosphonylmethoxypropyl] cytosine) is a monophosphate nucleotide analogue, which is used as third line treatment for HCMV infection. Because it inherently possesses a phosphate-like group, CDV does not require initial phosphorylation by the viral phosphokinase. However, it still requires activation (diphosphorylation) by cellular kinases. Like GCV, the incorporation of CDV-diphosphate

into viral DNA leads to termination of HCMV DNA replication. CDV is approved as a broad-spectrum antiviral for treatment of infections caused by various DNA viruses, including members of all three herpesvirus subfamilies. It has been used effectively to treat HCMV retinitis in HIV/AIDS patients. CDV has a relatively long intracellular half-life compared to GCV and FOS, but similar to FOS, its main adverse effects are nephrotoxicity and myelosuppression (Biron, 2006, Tan, 2014).

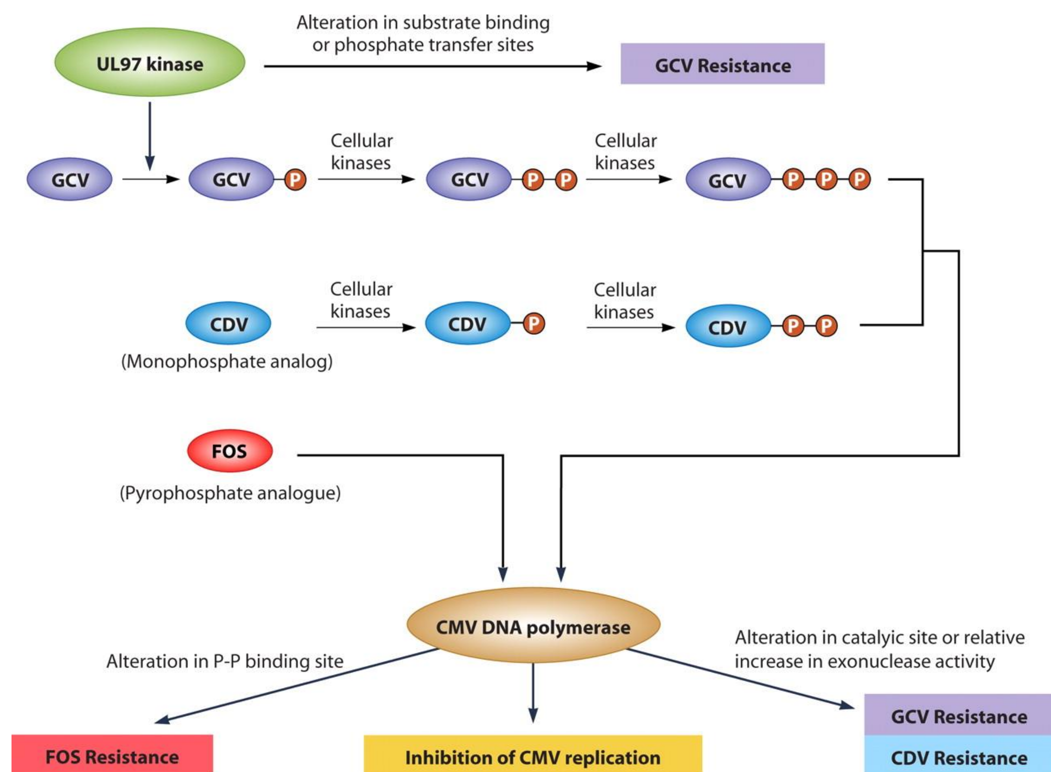


Figure 1.8. Approved anti-HCMV drugs. Addition of the first phosphate to Ganciclovir (GCV) (and its prodrug valganciclovir, vGCV) is performed by the HCMV UL97 serine/threonine phosphokinase. Following addition of two further phosphates by cellular kinases, GCV triphosphate – the active form of the drug – is incorporated into nascent DNA by the viral DNA polymerase (UL54). Resistance results from UL97 and/or UL54 mutations. Cidofovir (CDV), a monophosphate analog, does not require initial viral kinase activity; two phosphates are added by cellular kinases to produce the active form of the drug. Resistance to CDV is conferred only by DNA polymerase mutations. Foscarnet (FOS) is a pyrophosphate analog, which does not require activation. Resistance is conferred only by DNA polymerase mutations. The DNA polymerase is the ultimate target of all four drugs. Image adapted from Lurain and Chou (Lurain and Chou, 2010).

Other drug candidates

The following are among several new drugs (Biron, 2006, Tan, 2014) being actively assessed for treatment of HCMV disease.

Maribavir (MBV), a benzimidazole antiviral agent, targets the viral pUL97 phosphokinase and acts by inhibiting viral encapsidation and nuclear egress of progeny viral particles. MBV has been used with some success as salvage therapy in selected patients with multidrug-resistant HCMV (Avery et al., 2010). However, it has not been successful as prophylaxis in allogeneic stem cell or in liver transplant recipients (Marty et al., 2011), although this has been attributed in part to suboptimal dosing (Marty and Boeckh, 2011). The main side effects reported are taste disturbance and skin eruptions.

CMX001 (hexadecyloxypropyl-cidofovir), a lipid ester of CDV (Williams-Aziz et al., 2005, Tan, 2014), remains intact in plasma and delivers the active compound (CDV) directly to target cells, resulting in enhanced cellular uptake and high intracellular levels of CDV diphosphate. CDV diphosphate competitively inhibits the viral DNA polymerase driven incorporation of deoxycytidine triphosphate into viral DNA, thereby terminating chain elongation. CMX001 has been considered for use in ganciclovir-resistant CMV disease (Price and Prichard, 2011). Side effects are mild and include abdominal discomfort and stomatitis (Hostetler, 2009).

Letermovir (AIC246), a 3,4-dihydroquinazoline, is a terminase inhibitor which targets viral pUL56 (large terminase subunit), thereby inhibiting the cleavage of concatemeric progeny DNA to unit-length genomes that must be packaged into capsids. AIC246 therefore blocks viral replication without inhibiting the synthesis of progeny HCMV DNA or viral proteins (Verghese and Schleiss, 2013). It has shown high efficacy in multidrug-resistant disseminated HCMV infection in a lung transplant patient (Kaul et al., 2011). So far, no toxicities have been reported (Melendez and Razonable, 2015).

Leflunomide, a pyrimidine synthesis inhibitor (Torres-Madriz and Boucher, 2008), is approved for use in some autoimmune conditions, but has also shown anti-HCMV efficacy. It acts by disrupting progeny virion assembly. Successful use has been demonstrated in solid organ transplant recipients, but not in hematopoietic stem cell transplant recipients (Battiwalla et al., 2007). Anaemia and hepatotoxicity are the drug's main adverse effects.

Artesunate (ART), a commonly used antimalarial in endemic countries, has shown some anti-HCMV activity *in vitro*, and is undergoing extensive investigation for use in some

HCMV infection settings. Some research groups have shown that ART enhances activity of commonly used anti-HCMV antivirals when used in combination. However, conflicting outcomes have reported (Gantt et al., 2013, Germi et al., 2014, Kaptein et al., 2006, Lau et al., 2011, Schnepf et al., 2011, Shapira et al., 2008, Wolf et al., 2011, Zeng et al., 2015). The dose and duration of use of artesunate may be contributory factors to the variation in reported results, and need further investigation.

1.9.3 VACCINES

Owing to the significant morbidity and mortality associated with HCMV, and resultant decrease in quality of life, particularly in congenital infection and transplant settings, at the onset of the 21st Century an American Institute of Medicine report designated HCMV a level I (highest priority) vaccine target (Stratton et al., 2000). Although an effective vaccine remains elusive, there has been some good efficacy demonstrated by some formulations such as the gB subunit vaccine (Griffiths et al., 2011, Pass, 2009). Other vaccines are at various stages of development. Further trials have been planned in paediatric groups, or to block maternal transmission (Emery, 2011, Emery et al., 2013, Griffiths, 2002, Griffiths, 2012, Griffiths, 2014, Griffiths et al., 2011, Schleiss and Heineman, 2005, Schleiss, 2013, Sung and Schleiss, 2010, Zhong and Khanna, 2007, Pass, 2009, Pass et al., 2009). Vaccine formulations incorporate different viral immunogens which can elicit adequate humoral and/or cellular responses. This includes HCMV surface glycoproteins, which can be immunogenic and therefore form attractive targets for vaccine development. For instance, gB has been a focus of several anti-HCMV vaccine development efforts because it is relatively conserved, highly immunogenic and an important target for both humoral and cellular immune responses (Rasmussen and Cowan, 2003, Rasmussen et al., 1985, Rasmussen et al., 1988, Gretch et al., 1988a, Gretch et al., 1988b, Chandler and McDougall, 1986). More recent vaccine development efforts also include the gCIII complexes due to their role in host cell tropism and disease, and the highly potent neutralising antibody responses induced (Lilleri et al., 2013, Fouts et al., 2012, Cui et al., 2013, Macagno et al., 2010, Saccoccio et al., 2011). Anti-HCMV vaccines currently being assessed can be broadly categorised as follows (Rieder and Steininger, 2014, McVoy, 2013).

Live attenuated vaccines

Since virtually all antigens can be expressed, live attenuated vaccines have the inherent potential of inducing both humoral and cellular immune responses close to levels induced by natural infection. A major disadvantage however, borders around safety, particularly if the vaccine is administered in pregnancy. One promising live vaccine is based on an isolate of HCMV strain Towne that was attenuated by serial passage in human fibroblasts. It has been reported to induce both cellular and humoral responses in what appears to be a dose-dependent manner (Adler et al., 1995). The vaccine has undergone enhancement, including the generation of chimeras that incorporate portions of the HCMV strain Toledo. Preliminary reports among seropositive volunteers indicate no improvement in baseline immunity; effects in seronegative participants are awaited (Heineman et al., 2006).

Subunit vaccines

These are targeted at inducing robust immune responses to selected immunogens. Commonly, the vaccine is a combination of an immunogenic molecule and an adjuvant. Two subunit vaccines currently in trials are the gB/MF59 vaccine (Sanofi) and the gB/AS01 vaccine (GlaxoSmithKline). Both vaccines are targeted at eliciting antibody responses. The gB/MF59 vaccine has so far gone through phase 2 trials, where it showed 50% efficacy in reducing incidence of primary HCMV infection in one cohort, while in a cohort of renal and liver transplant recipients use of gB/MF59 led to reduced duration of viremia and shorter periods of antiviral treatment (Griffiths et al., 2011, Pass, 2009). There have also been encouraging reports for the gB/AS01 vaccine from phase 1 trials, where its safety and immunogenicity have been confirmed (Schleiss et al., 2014).

Plasmid DNA Vaccines

In this category of vaccines, a portion of DNA encoding an immunogenic molecule is delivered into host cells, which then express the immunogen of interest *in vivo*. Examples are the TransVax vaccine developed for use as a therapeutic vaccine in the transplant setting, and CyMVectin targeting congenital HCMV. TransVax contains pp65 and gB plasmid DNAs with poloxamer adjuvant whereas CyMVectin has pp65 and gB plasmid DNAs with Vaxfectin adjuvant. TransVax has undergone phase 2 trials in HSCT recipients, where it is reported to have reduced post-transplant viremia (Kharfan-Dabaja

et al., 2012, Wloch et al., 2008). Another plasmid DNA vaccine in phase 2 trials is the trivalent CL-CT02 composed of gB/pp65/IE1 (Jacobson et al., 2009).

Virus–Vectored Vaccines

An example in this group is AVX601, an RNA virus–vectored vaccine which utilises replication-defective Venezuelan equine encephalitis virus for expression of a pp65/IE1 and gB fusion protein (Reap et al., 2007a, Reap et al., 2007b).

The quest for an efficacious HCMV vaccine continues to gather momentum as more breakthroughs are made in understanding the virus and its multi-tier arsenal used to evade our immune defences. Vaccine development efforts require multi-pronged, multi-disciplinary approaches encompassing a wide spectrum of expertise from basic science to social science and economics. A particular bottleneck that remains to be overcome is the low public awareness of HCMV and its devastating impact on childhood and quality of life in general.

1.9.4 NON-PHARMACOLOGICAL CONTROL METHODS

Several non-pharmacological strategies have been investigated with the view of limiting infectious HCMV load, particularly in breast milk in the context of preterm and very low birthweight infants. These include freeze-thawing (Hamprecht et al., 2004b, Maschmann et al., 2006), pasteurisation (Hamprecht et al., 2004b, Stock et al., 2015, Ewaschuk et al., 2011a, Ewaschuk et al., 2011b, Hamprecht et al., 2004a), and UV irradiation (Christen et al., 2013, Christen, 2014). These techniques show good promise, although varied outcomes have been obtained with some approaches, particularly freeze-thawing (Maschmann et al., 2006, Chiavarini et al., 2011). Concerns linger, such as the retention of nutritional and immunological qualities in milk following application of treatments; however techniques continue to be improved to address these concerns (Goelz et al., 2009, Czank et al., 2009, Hamprecht et al., 2004a). Other challenges border on logistics for large-scale (community-level) implementation. These include the requirement for specialised equipment and infrastructure such as the pasteurisation equipment or UV source, electricity/energy source, and fridge/freezers for storage of treated milk to ensure that an adequate supply of safe milk is available to the baby ‘on demand’. As technology advances and costs reduce, some of these challenges are being

overcome. In the context of Sub-Saharan Africa, where multiple pathogens can afflict concurrently, efforts that eliminate or reduce pathogen load in breast milk can markedly improve early child health. Apart from HCMV, HIV, Hepatitis B virus, EBV, and various bacteria are transmissible through milk. Therefore milk treatment can potentially limit the transmission of a variety of pathogens.

Public health promotion efforts are now increasingly being made to promote basic personal hygiene practices such as handwashing to limit transmission of HCMV. Hand hygiene has been shown to be a very effective and cost-effective strategy for blocking the HCMV transmission cycle, particularly in hospitals and childcare settings (Stowell et al., 2014, Cannon and Davis, 2005).

1.10 THESIS AIMS

Previous studies in our laboratory showed widespread adverse effects of HCMV infection on growth and development of Zambian infants, particularly the HEU (Gompels et al., 2012, Sanz-Ramos et al., 2013), but the route of HCMV infection associated with these effects was undefined. Building upon these previous studies, the overarching aims of this thesis were therefore to explore congenital and breast milk HCMV transmission in Zambia, and to genetically characterise locally circulating HCMVs, in order to establish the transmission route(s) responsible for the widespread negative effects of HCMV on child wellbeing. Crucially, our studies took into account maternal HIV infection by comparing HIV-infected and uninfected mothers, thereby clarifying HIV influence on HCMV infection burden and transmission dynamics. Our studies sought to utilise non-invasive samples (saliva, umbilical tissue, breast milk) for possible future routine application to HCMV detection in the local setting. It is envisaged that findings presented in this thesis will ultimately help inform clinical practice and add fundamental information, including on virus genetic diversity, to help direct the deployment of anti-HCMV drugs, vaccines, and other interventions in Zambia.

Specific Objectives

To accomplish the above aims, the following objectives were set:

1) To investigate the birth prevalence of HCMV among new-borns in Zambia, using non-invasive samples;
2) To analyse HCMV shedding and viral DNA load kinetics in breast milk from both HIV-infected and uninfected Zambian mothers;
3) To analyse HCMV genotype variants in saliva and breast milk by sequencing the hypervariable gO and gN genes;
4) To assess HCMV infection burden by investigating mixed-genotype infections using NGS deep sequencing and novel *in silico* methods for detecting and enumerating genotypes; and,
5) To assemble and characterise a reference complete HCMV genome from DNA extracted directly from breast milk in Africa.

Chapter 2

METHODS

2.1 ETHICAL CONSIDERATIONS

Studies undertaken as part of this thesis were reviewed and approved by the LSHTM Research Ethics Committee (LSHTM REC No. 6456, appendix 1) and the University of Zambia Biomedical Research Ethics Committee (UNZABREC No. 010-06-13, appendix 2).

Study participants were free to decline or withdraw from their participation in the study at any time and this did not disadvantage them in any way. Confidentiality was maintained at all stages of the study. All samples were coded with a unique study ID and no personal identifiers were used during sample processing or data analysis. In case evidence of congenital HCMV infection was identified during the study, arrangements were in place for referral to locally available specialist assessment and follow-up services including neurological, psychosocial, and audiometry at Zambia's largest tertiary hospital, the University Teaching Hospital (UTH), where part of the study was based.

2.2 STUDY COHORTS AND SAMPLE COLLECTION

HCMV was investigated in two settings:

1) Congenital infection, in a prospective cross sectional study of new-borns at UTH. For this part of the study, we collected saliva and leftover umbilical tissue (which is normally discarded after delivery) as the non-invasive samples for virus diagnosis and analysis. Parents were also given the option of either finger/toe nails or hair follicles if they were uncomfortable with use of umbilical tissue (see participants' information sheet, appendix 5). HCMV analyses focused primarily on use of saliva, while the other sample types were used to screen for congenital HHV-6A/B, which are known to integrate chromosomally and therefore potentially present in every body cell.

2) Postnatal infection transmission, via retrospective study of breast milk, detailed in section 2.2.2.

The roseoloviruses HHV-6A and HHV-6B, which are closely related to HCMV but have unique pathological, clinical and epidemiological characteristics, were also evaluated where appropriate as experimental controls.

2.2.1 NEONATAL COHORT

Study Setting and Participants

This aspect of the study was cross-sectional, and conducted at UTH, in Lusaka, Zambia. As Zambia's largest tertiary hospital, UTH is a 1800-bed government-run referral centre providing specialist health care including clinical, radiological, laboratory, and rehabilitation services, and also serves as the country's main training institution for various levels of healthcare cadres including Doctors, Nurses, Pharmacists, Biomedical Scientists, and Physiotherapists. UTH has an immediate catchment population of over 2 million people, and nationally serves Zambia's population of 15 million. Various healthcare services ranging from primary to tertiary are provided to in- and out-patients under four main clinical departments: Paediatrics & Child Health, Obstetrics & Gynaecology, Internal Medicine, and Surgery, which are complemented by Pathology/Microbiology, Radiology, Pharmacy, and Physiotherapy departments. Participants in our neonatal study were drawn from the Labour Ward (LW) and the Neonatal Intensive Care Unit (NICU). All neonates born in the UTH LW, and all neonates referred to NICU for specialised neonatal care, whether born at UTH or elsewhere, were eligible for inclusion in the study provided they were 14 days or younger, and a parent / guardian consented to participation by signing a consent form (appendix 6). Exclusion criteria were: being above 14 days old, having received a blood transfusion, and refusal of consent by parent / guardian. The study was introduced by a research nurse to parents / guardians of prospective participants who met the inclusion criteria, with detailed explanation and a copy of the study Information Sheet (appendix 5) provided. On obtaining written consent, relevant demographic and medical information was collected on a standardized data capture form (appendix 7).

For this study, our target sample size was 200 neonates. Recruitments were however hampered by protracted industrial action taken by nursing staff during the period of our field study. Between 1st August and 15th September 2013, 117 neonates were recruited to the study: 101 from LW and 16 from NICU. In LW all neonates were recruited within

1 hour of birth (median 10min, IQR 5 – 26min) and in all cases but one, saliva samples were collected before initiation of breastmilk feeding. There were four cases of stillbirths. In NICU, the median postnatal age at recruitment was 4 days (IQR 1.0 – 9.3 days). At recruitment half (8/16) of the participants had already commenced breastmilk feeding.

Specimen Collection and Handling

Saliva was collected by swabbing the neonate's mouth with a sterile polyester-tipped swab (Peel pouch Dryswab™, Medical Wire & Equipment), which was air-dried then placed in a separate, labelled grip seal bag. Additionally, approximately 1cm^3 of umbilical tissue was collected using a sterile surgical blade shortly after birth in LW or from above the umbilical cord clamp in NICU. Where collection of umbilical tissue was declined or not feasible, parents could opt for nail clippings or hair follicles to be collected. Nail clippings were obtained using sterilized nail clippers. Hair follicles were plucked from the scalp. Overall, we collected saliva from 116 participants, umbilical tissue from 115, hair from one, and nail clippings from one participant, as outlined in table 2.1. Prior to DNA extraction, saliva, nail, and hair specimens were stored securely in a locked cabinet at room temperature, while umbilical tissue was frozen at –20°C at the UTH Virology Laboratory (UTHVL) located within the UTH grounds.

Sample Type	LW	NICU	Total
Saliva	100	16	116
Umbilical tissue	101	14	115
Hair	-	01	01
Nails	-	01	01

Table 2.1: Type of sample collected by ward. Saliva and umbilical tissue were the preferred non-invasive sample types by parents/guardians in this study.

2.2.2 BREASTFEEDING COHORT

The second arm of this thesis analysed available breast milk samples remaining from the Breast Feeding and Postpartum Maternal Health (BFPH) study (Collin et al., 2006), and stored at the Institute of Child Health, University College London, London. BFPH was

conducted in Lusaka, Zambia from June 2001 to July 2003 as a longitudinal cohort study that recruited 429 (218 HIV-negative, 211 HIV-positive) women to examine factors associated with postpartum physical and mental morbidity. Maternal and infant health and infant feeding data were collected from 34 weeks gestation to 16 weeks postpartum. At 11 separate postnatal visits (days [D] 3, 7, 10, 14, and weeks [W] 3, 4, 5, 6, 9, 12, 16) spot milk samples were collected for measurement of inflammation markers and HIV-1 viral load. HIV-positive mothers and their new-borns received single-dose nevirapine as per local prevention of mother to child HIV transmission (PMTCT) policy (Collin et al., 2006, Kasonka et al., 2006, Phiri et al., 2006).

Breast milk samples for present study

For this current study, 261 milk samples (118 HIV-positive and 143 HIV-negative) collected at W16, were available in sufficient volume. In order to gain insight into the kinetics of HCMV shedding in breast milk in this population, we also examined all available milk sets with D3, W2, W4, W9, and W12 samples. Forty such sets, from 20 HIV-negative and 20 HIV-positive women, were available for analysis in the kinetics studies. All available milk samples were stored at –80°C.

2.3 DNA EXTRACTION AND PCR

2.3.1 DNA EXTRACTION

DNA utilised in analyses including qualitative and quantitative PCR and genome sequencing assays for HCMV, HHV-6A, HHV-6B, and GAPDH was extracted from saliva, umbilical tissue, nails, hair or breast milk using the QIAamp® DNA Mini kit (QIAGEN®), following manufacturers' instructions, as follows in brief.

Saliva

Here the '*DNA Purification from Buccal Swabs*' spin protocol was used. Each polyester swab tip was separated from the stick and placed in 400µl of PBS in a sterile 2ml microcentrifuge tube. 20µl of proteinase K and 400µl of Buffer AL were then added and immediately mixed thoroughly by vortexing for 15s prior to incubation at 56°C for 10min. The tube was then centrifuged briefly to remove drops from inside the lid, and 400µl of molecular biology grade 100% ethanol added to the sample, mixed by pulse-

vortexing for 15s, followed by brief centrifugation to remove drops from inside the lid. Binding of sample to the QIAamp Mini spin column membrane was achieved in two successive steps, each involving the application of up to 700µl of sample mixture to a QIAamp Mini spin column in a 2ml collection tube followed by centrifugation at 8000rpm for 1min. The filtrate with collection tube was discarded and spin column placed in another clean 2ml collection tube. Wash steps involved addition of 500µl of Buffer AW1, centrifugation at 8000rpm for 1min, followed by 500µl of Buffer AW2 and centrifugation at 14,000rpm for 3min, with change of collection tube at the end of each centrifugation step. Additional centrifugation was performed at 14,000rpm for 1min before the elution step. DNA was eluted into 50µl molecular biology grade water added to the spin column placed in a sterile 1.5ml microcentrifuge tube. Following incubation at room temperature for 5min, final centrifugation at 8000rpm for 1min was performed to recover the DNA. DNA yield and quality were measured on a NanoDrop™ spectrophotometer (Thermo Fisher Scientific Inc), and extracted DNA samples were stored at –20°C at UTHVL and LSHTM, prior to downstream analysis.

Umbilical tissue, Nails, and Hair

For these samples types, the '*DNA Purification from Tissues*' protocol was used. Up to 25mg of each sample was cut up into small pieces using a sterile disposable blade and placed in a sterile 1.5ml microcentrifuge tube. 180µl of Buffer ATL and 20µl of proteinase K were added and the mixture was vortexed for 15s before incubation at 56°C for 3hrs, during which occasional vortexing was done to disperse the sample. At the end of the incubation period, the microcentrifuge tube was centrifuged briefly to remove drops from the inside of the lid. 200µl Buffer AL was then added to the sample, mixed by pulse-vortexing for 15s, and incubated at 70°C for 10min. 200µl of molecular biology grade 100% ethanol was added to the sample, mixed by pulse-vortexing for 15s, and tube briefly centrifuged to remove drops from inside of the lid. The mixture was then carefully applied to a QIAamp Mini spin column in a 2ml collection tube and cap closed before centrifugation at 8000rpm for 1min. The filtrate with collection tube was discarded and spin column placed in another clean 2ml collection tube. Wash, elution, yield quantification, and DNA sample storage steps were the same as for DNA extraction from saliva, outlined in the previous section.

Breast milk

DNA was extracted from breast milk using the '*DNA Purification from Blood or Body Fluids*' spin column protocol. In readiness for DNA extraction, each milk sample was thawed and gently mixed to homogeneity. 20µl of proteinase K was pipetted into the bottom of a sterile 1.5ml microcentrifuge tube followed by 200µl of breast milk sample, and then 200µl of Buffer AL. Tube contents were mixed thoroughly by pulse-vortexing for 15s to yield a homogeneous solution, which was incubated at 56°C for 10min. At the end of incubation, the tube was centrifuged briefly to remove drops from inside of the lid. 200µl of molecular biology grade 100% ethanol was then added to the sample, mixed by pulse-vortexing for 15s, and then the tube was briefly centrifuged to remove drops from inside of the lid. The mixture was then carefully applied into a QIAamp Mini spin column in a 2ml collection tube and cap closed before centrifugation at 8000rpm for 1min. Wash, elution, yield quantification, and DNA sample storage steps were the same as for DNA extraction from saliva, outlined earlier.

2.3.2 QUALITATIVE PCR

All PCR experiments were conducted in three separate dedicated areas to minimise contamination, as follows. Reagents were stored and prepared in a dedicated 'clean' room located on a separate floor of the laboratory building, and using dedicated instruments and laboratory personal protective equipment (PPE, i.e. laboratory coats and gloves). Clinical samples were stored and added to reaction mixtures in another room, again with dedicated instruments and PPE. Amplification and gel electrophoresis were then conducted in a separate third room. The principle of unidirectional workflow was observed, and reaction tubes/plates were sealed prior to loading into the thermocycler. All PCR reactions included appropriate positive controls and two negative controls: reagent control (i.e. no template) and water as template.

Targets and Oligonucleotide Primers

Conserved genes HCMV UL55 and HHV-6A/B U38, encoding HCMV gB and the HHV-6A/B polymerase respectively, were targeted for virus screening, while hypervariable HCMV UL73 and HCMV UL74 genes encoding HCMV gN and gO respectively were used in genotyping assays as defined previously in the Gompels research laboratory (Bates et al., 2008, Mattick et al., 2004, Bates et al., 2009, Mattes et al., 2005). For HHV6-A/B,

additional PCR targets included the viral chemokine encoded by U83 and the chemokine receptor encoded by U51. To validate the DNA extraction protocol, we amplified a 104bp fragment of the human housekeeping gene encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH), as previously described (Asahi-Ozaki et al., 2006, Bates et al., 2009, Bates et al., 2008). Oligonucleotide primers used are detailed in table 2.2. All primers were synthesised by Sigma-Aldrich Co. Ltd., UK. The same primer sets were also used in sequencing PCR reactions (section 2.5.1).

PCR setup and thermal cycling conditions

Qualitative PCR reactions were in 25µl volumes consisting of 12µl of Promega's GoTaq® Green Master Mix (containing 50 units/ml of Taq DNA polymerase, 400µM dATP, 400µM dGTP, 400µM dCTP, 400µM dTTP, 3mM MgCl2), 2.5µl each of forward and reverse primer (10µM), 7µl nuclease-free water (Promega), and 1µl of template DNA. Each reaction had a final primer concentration of 1µM. For HCMV gB, HCMV gO, HCMV gN, and GAPDH amplification, thermal cycling conditions were: 95°C for 2min, then 40 cycles of 95°C for 30s, 59°C for 30s, and 72°C for 2min, followed by final extension at 72°C for 5min, and holding at 4°C. For HHV-6A/B U38 amplification, only annealing temperature differed, at 55°C. PCR was performed using a DYAD™ DNA Engine Dual Block Peltier Thermocycler (MJ research).

Agarose Gel Electrophoresis

Agarose gel electrophoresis used 1X Tris base/Boric acid/EDTA (TBE) buffer to separate PCR amplification products loaded in 1% agarose gels (1g genetic analysis grade agarose (Fisher scientific, UK) in 100ml 1X TBE). Agarose gels were stained with ethidium bromide (5µg/ml) to aid visualisation of DNA bands under UV illumination in a UV box. Because GoTaq® Green Master Mix (Promega) used in our PCR reactions contained loading dye, amplification products were loaded directly onto gels. Electrophoresis was carried out at 100 Volts until the leading edge of dye neared the base of the gel to maximise band separation (approximately 45-60min). Gels were viewed on the GeneGenius SYDR2 Gel Bioimaging Unit (Syngene), with GeneSnap software, which enabled digital images of the DNA bands to be captured.

Target		Oligo Name	Sequence (5' to 3')	Product size (bp)	Reference
HCMV	UL55 (gB)	gB1	GAGGACAACGAAATCCTGTTGGGCA	149	(Bates et al., 2008, Mattes et al., 2005, Mattick et al., 2004)
		gB2	TCGACGGTGGAGATACTGCTGAGG		
		gB-P3	[*FAM*]CAATCATGCGTTTGAAGAGGTAGTCCACG[*TAMRA*]		
	UL73 (gN)	gNup	TGGTGTGATGGAGTGGAAC	411-420	(Bates et al., 2008, Mattick et al., 2004)
		gNlw	TAGCCTTTGGTGGTGGTTGC		
	UL74 (gO)	GO-up	CGACCAGAATCAGCAGTGAG	742	(Bates et al., 2008, Mattick et al., 2004)
		GO-lw	TGTACAGTTGCGTTGTGCGTA		
HHV-6A/B	U38 (Pol)	MaSu38F	AAGACGGGTTATTATGCTGTG	566	Gompels lab, (unpublished)
		MaSu38R	ACAGACATAAAGATGCTATCC		
	U51 (Chemokine receptor)	U51MF1	TATGGTTGGGTATTTATTTTTTCGTC	867	Gompels lab, (unpublished)
		U51MR1	ACTCTCGAGTCATTTTAACATTTTTATTCCAACCTCTAAATCC		
	U83 (Chemokine)	U83FP1	AAGTTAACACGACGGGAACAA	339	Gompels lab, (unpublished)
		U83RP1	TGCCATATCACACATCGAG		
Human	GAPDH	GAPup	GCTCCCTCTTTCTTTGCAGCAAT	104	(Asahi-Ozaki et al., 2006, Bates et al., 2008)
		GAPdown	TACCATGAGTCCTTCCACGATAC		
		GAP Probe	[*FAM*]TCCTGCACCACCAACTGCTTAGCACC[*TAMRA*]		
pGEM-T Vector	Insert of interest	SP6	GCTCCCTCTTTCTTTGCAGCAAT	(Insert size + 46)	Promega Corp., TM042 rev. May 2007
		T7	TACCATGAGTCCTTCCACGATAC		

Table 2.2: Oligonucleotide primers and probes used in our PCR and qPCR assays. The table outlines the sequences of primers and probes used in this study, and size of amplification products. All primers and probes were synthesised and supplied by Sigma-Aldrich Co. Ltd., UK.

2.3.3 QUANTITATIVE REAL-TIME PCR

DNA copy numbers were computed in a TaqMan® Real-Time assay run on the Applied Biosystems® 7500 Fast Real-Time PCR System (Applied Biosystems® Inc.) as described for HCMV gB (Bates et al., 2008, Mattes et al., 2005) and GAPDH (Bates et al., 2008, Asahi-Ozaki et al., 2006). The assay used the 5'-6FAM / 3'-TAMRA – labelled oligonucleotide probes (Sigma-Aldrich Co. Ltd., UK) gB-P3 or GAPprobe with the gB1/gB2 and GAPup/GAPdown primer sets (table 2.2) for HCMV and GAPDH respectively. Standard curves were generated from 10-fold serial dilutions (in triplicate) of plasmid-cloned DNA standards (table 2.3) spanning 10^0 to 10^6 concentrations. Each qPCR reaction was in a 25µl volume consisting of 10µl KAPA PROBE FAST Universal qPCR Master Mix (Kapa Biosystems), 1µl probe (5mM), 1µl each of forward and reverse primers (10 mM), 0.4µl ROX Low, 7µl nuclease-free water, and 5µl template DNA. All standards and samples were run in triplicate. Cycling conditions were 95°C for 10min, followed by 45 cycles of 95°C for 15s and 60°C for 1min. Our in-house plasmid reference standards derived from cloned gB amplicons (see section 2.4) were normalised to the HCMV 09/162 International Clinical Reference Standard (National Institute of Biological Standards, Potters Barr, UK).

2.4 PLASMID CLONING

Generation of plasmid clones

To generate plasmid clones with GAPDH and HCMV gB inserts, the pGEM®-T Easy Vector Systems cloning kit (Promega Corporation) was used, in accordance with manufacturers' instructions (Promega Corporation, TM042 revised May 2007) as follows. PCR product or control DNA insert, pGEM®-T Easy Vector and T4 DNA ligase were added to a microcentrifuge tube at PCR product to vector ratios of 3:1 and 1:1. The PCR product concentration was determined using a NanoDrop™ spectrophotometer (Thermo Scientific™) after purifying DNA bands of interest from agarose gels following electrophoresis. To calculate the amount of PCR product (insert) to include in the ligation reaction, the following equation used:

$$\text{ng of insert} = \frac{\text{ng of vector X kb size of insert}}{\text{kb size of vector}} \text{ X } \quad \text{insert:vector molar ratio}$$

Ligation reactions were performed for one hour at room temperature. After thawing on ice, 50µl of JM109 competent cells were added to microcentrifuge tubes containing 2µl of ligated vector with insert. The tubes were then gently mixed and incubated on ice for 20min, then heat-shocked at 42°C for 45sec in a water bath before being incubated on ice for a further 2min. The tubes were then placed in a shaking pre-warmed 37°C incubator for 90min at 150rpm. Transformations were plated twice onto separate, pre-warmed LB agar plates containing 50µg/ml ampicillin (supplemented with IPTG 0.1M/X-Gal 20µg/ml), one with 100µl and another with 20µl. After 5 minutes, the plates were inverted and placed in a 37°C incubator for overnight growth. Individual white colonies were streaked out on new pre-warmed plates and incubated overnight at 37°C. Individual colonies from these second plates were selected with a sterile pipette tip and added to 3ml of LB broth, supplemented with 50ug/ml ampicillin, and grown overnight in a 37°C shaking incubator. Glycerol stocks were prepared by mixing 700µl of transformed bacteria stock with 700µl of 40% glycerol in a cryovial. These stocks were stored at –80°C. Inserts were verified by PCR and nucleotide sequencing with primers used during production of the insert.

Isolation of Plasmid DNA

Plasmid DNA was purified from overnight cultures of transformed bacteria in LB broth, supplemented with 50ug/ml ampicillin, using the PureLink® HiPure Plasmid Miniprep kit (Invitrogen™/Life Technologies™). For each sample, cells were harvested from 3ml of an overnight LB culture by centrifuging at 4000×g for 5min, followed by removal of all medium. Cells were resuspended in 0.4ml of Resuspension Buffer (R3) with RNase A in a microcentrifuge tube until homogeneous. Thereafter 0.4ml of Lysis Buffer (L7) were added, and contents mixed gently by inverting the capped tube (avoiding vortexing) until the lysate mixture was thoroughly homogenous. The mixture was incubated at room temperature for 5min, followed by addition of 0.4ml of Precipitation Buffer (N3), and mixing immediately by inverting the tube (avoiding vortexing) until the mixture was thoroughly homogeneous. The lysate was then centrifuged at >12,000×g for 10min at room temperature, the clear lysate transferred into another sterile tube and centrifuged at >12,000×g for a further 5min at room temperature to remove any remaining cellular debris. The supernatant was then loaded onto a column and allowed to drain by gravity. The column was washed twice with 2.5ml Wash Buffer (W8), allowing the solution in the

column to drain by gravity and discarding the flow-through after each wash. To elute and precipitate the DNA, a sterile microcentrifuge (elution) tube was placed under the column, and 0.9ml Elution Buffer (E4) added to the column, allowing the solution to drain by gravity. 0.63ml of isopropanol were added to the elution tube and mixed thoroughly before centrifuging the elution tube at >12,000×g for 30min at 4°C. The supernatant was carefully removed and discarded, and the DNA pellet resuspend in 1ml of 70% ethanol before additional centrifugation at >12,000×g for 5min at 4°C. Again, the supernatant was carefully removed and discarded. The DNA pellet was air-dried for 10min before being resuspended in 50µL of nuclease-free water. The DNA yield and quality were measured on a NanoDrop™ spectrophotometer (Thermo Fisher Scientific Inc), and aliquots made for storage at –20°C pending further use. Inserts were verified by PCR amplification across the insertion site using the SP6 and T7 promoter primers (Promega Corporation, TM042 revised May 2007, see table 2.2), followed by nucleotide sequencing with primers used during production of the insert. Serial 10-fold dilutions containing $1X10^{10}$ to $1X10^{0}$ DNA copies/µl were made, and aliquots of each concentration stored at -20°C. Table 2.3 outlines characteristics of the plasmid clones.

	GAPDH	**HCMV gB**
Plasmid designation	G2-4	B2-8
Insert gene source	Human saliva DNA	HCMV strain AD169
Vector	pGEM®-T Easy	pGEM®-T Easy
Vector size (bp)	3015	3015
Insert size (bp)	104	149
Miniprep yield (ng/µl)	73.2	82.5
Stock concentration (copies/µl)	$2.14 X 10^{10}$	$2.38 X 10^{10}$

Table 2.3: Characteristics of plasmid clones. pGEM®-T Easy plasmids with GAPDH and HCMV gB inserts were designated G2-4 and B2-8 respectively as characterised here, for use as in-house standards in qPCR experiments.

2.5 NUCLEOTIDE SEQUENCING

2.5.1 GEL DNA PURIFICATION

Following qualitative PCR runs, DNA was purified from gels using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp.). Bands of interest were excised from gels using

a clean scalpel and transferred to a 1.5ml microcentrifuge tube. To this, buffer ADB was added at a rate of three times the volume of the agarose slice. The tube was incubate at 37-55°C for 5-10min until the gel slice completely dissolved, and the melted agarose solution was then transferred to a Zymo-Spin™ column in a collection tube. This was centrifuged at 14,000rpm for 60s, flow-through discarded, and then washed twice with 200µl of Wash Buffer, each time centrifuging at 14,000rpm for 30s and discarding the filtrate. Finally, 15µl of nuclease-free water were added directly to the column placed into a sterile 1.5ml tube, and incubated for 2-5minbefore centrifugation at 14,000rpm for 60s to recover DNA. The DNA yield and quality were measured on a NanoDrop™ spectrophotometer (Thermo Fisher Scientific Inc.) before samples were sequenced or stored at -20°C.

2.5.2 SANGER SEQUENCING

Nucleotide sequences were determined by capillary electrophoresis using Sanger methods to define the PCR amplicons derived from HCMV and HHV-6A/6B amplified directly from clinical samples as previously described (Bates, 2008 & 2009). Sequencing reactions were performed in-house or outsourced from Source BioScience, Oxford, UK. The in-house method used the BigDye® Terminator v3.1 (Applied Biosystems) kit, following manufacturer's instructions to directly sequence both sense and antisense strands from gel-extracted DNA. Sequencing reactions were in 10µl volumes consisting of 3µl dilution buffer, 1µl primer (1µM), 4µl nuclease-free water, 1µl big dye, and 1µl template (10-20ng of DNA). Amplification reactions were performed in a DYAD™ DNA Engine Dual Block Peltier Thermocycler (MJ Research). Sample tubes were loaded into the thermocycler, preheated to 96°C for 4 minutes, followed by 25 cycles of denaturation at 96°C for 20sec, annealing at 50°C for 10sec, and extension at 60°C for 2min. There was then a final extension step at 60°C for 5min. Samples were then kept at 4°C pending DNA extraction, which involved precipitation with 0.1M sodium acetate / 68% ethanol. After this, samples were incubated on ice for 20 min, and then centrifuged at 3000xg for 30min at 4°C. Supernatant was then removed; ice-cold 70% ethanol was added and the samples were centrifuged at 3000xg for 10min prior to removal of supernatant. Finally, samples were re-suspended in Hi-di formamide (Applied Biosciences). Nucleotide sequences were then determined by capillary electrophoresis on an automated ABI3730 sequencer (Applied Biosystems). In some

cases, Sanger sequencing was outsourced from Source BioScience UK Ltd (http://www.lifesciences.sourcebioscience.com/genomic-services/sanger-sequencing-service/). Nucleotide sequences were compiled using ChromasPro software version 1.7 (Technelysium).

2.5.3 ILLUMINA NEXT GENERATION SEQUENCING

For further genomic analyses using Illumina platforms, six extracted DNA samples with viral loads characterised were selected from HIV-positive as well as HIV-negative W4, 12, and 16 milk samples. The SureSelect system (Agilent Technologies Inc.) (Depledge et al., 2011) was used for HCMV target amplification direct from the clinical samples described above, then genome libraries were prepared by clonal amplification and paired-end sequenced on the MiSeq™ (Illumina®) platform by our collaborators, the Andrew Davison group at the MRC Center for Virus Research, University of Glasgow. Resultant FASTQ reads were delivered using Dropbox (Dropbox Inc., San Francisco) then analysed as described in section 2.6 (Bioinformatics, NGS analyses).

2.6 BIOINFORMATICS

Phylogenetic analysis

Forward and reverse sequence chromatograms were analysed using ChromasPro software version 1.7 (Technelysium) and compiled to generate contiguous sequences which were then checked against archived sequences using the NCBI Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The ExPASy bioinformatics resource portal (http://www.expasy.org/resources) and its suite of tools was also utilised. Multiple alignments were performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and phylogenetic analyses conducted via maximum-likelihood methods in MEGA version 6.06 (Tamura et al., 2013). The Jones-Taylor-Thornton (JTT) model was used, with phylogenies tested by 1,000 bootstrap replications. Tree inference was made by the Nearest-Neighbour-Interchange (NNI) heuristic method.

NGS data analysis

Raw NGS reads were checked for quality using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed to a minimum length of 100 bases using Trimmomatic (Bolger et al., 2014). The NGS reads were optimised by VelvetOptimiser for reference-independent *de novo* assembly with Velvet (Zerbino and Birney, 2008) and ABACAS (Assefa et al., 2009), and contigs were verified by employing the consensus sequence as mapping reference using the Burrows-Wheeler Aligner, BWA (Li and Durbin, 2009, Li and Durbin, 2010), SAMtools / BCFtools (Li et al., 2009) and GATK (McKenna et al., 2010) for alignment, indexing, mapping, and variant calling. Artemis was employed for visualisation (Rutherford et al., 2000). Genome annotations were transferred from the HCMV reference strain Merlin (AY446894.2) using the Rapid Annotation Transfer Tool, RATT (Otto et al., 2011).

Molecular tags for strain identification

Genotype-specific nucleotide sequences ('Molecular tags') were developed (see Chapter 5) for the HCMV hypervariable genes UL73 and UL74 (Dolan et al., 2004, Pignatelli et al., 2004, Puchhammer-Stockl et al., 2006). The molecular tags were validated against all publicly available relevant HCMV sequences in GenBank Flat File Release 211 (National Center for Biotechnology Information (NCBI), 2015). Custom *perl* scripts were then developed to interrogate FASTQ reads in order to quantify mixed infections by enumerating the proportions of individual molecular tags in relation to representative read depths. From this information, Microsoft Excel 2013 was used to model proportions of each represented genotype.

2.7 STATISTICAL ANALYSIS

All data were entered into restricted, password-protected Microsoft Access 2010 databases, and related hard copy information kept in a locked cabinet. Microsoft Excel 2010, SPSS 21.0, and GraphPad Prism version 6.05 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com) were used for data analysis. Nominal data were analysed in contingency tables by Fisher Exact tests, while sample means were compared by the Student's t test and medians analysed by the Mann-Whitney U test. *p*-values were computed with significance threshold set at 0.05.

CHAPTER 3

INVESTIGATION OF HCMV AT BIRTH IN ZAMBIA USING NON-INVASIVE SAMPLES

3.1 INTRODUCTION

This chapter presents findings from studies aimed at assessing congenital HCMV infection in Zambia, Southern Africa. At the start of this thesis, there was limited information on childhood infections with HCMV, particularly congenital infection in Zambia. Previous studies by our research group indicated that HCMV was more prevalent, infection occurred earlier than in other global locations, and these infections were associated with adverse effects on child growth and development (Bates 2008, 2009; Gompels 2012). Studies in this chapter were thus conducted with particular interest to clarify the prevalence of congenital HCMV infection. We also tested for related betaherpesviruses, HHV-6A and HHV6-B, as controls. This was done by studying two neonatal groups i.e. 'normal' newborns delivered in the University Teaching Hospital (UTH) labour ward (LW), and neonates admitted to the UTH Neonatal Intensive Care Unit (NICU). In assessing infections, our approach was to utilise non-invasive samples: saliva, hair, nail clippings, or leftover umbilical tissue (which would otherwise be discarded following delivery), thereby minimising risk of occupational hazards, cross-infection, or distress to the donor – which is of particular importance when dealing with infants, children, and their guardians. These samples have been utilised elsewhere to analyse various infections, including betaherpesviruses (Hall et al., 2008, Hall et al., 2010, Morissette and Flamand, 2010, Mussi-Pinhata et al., 2009, Endo et al., 2009, Flamand et al., 2010, Murthy et al., 2011, Boppana et al., 2011, Arbuckle et al., 2010, Forman et al., 2012, Kainth and Caserta, 2011, Spano et al., 2007, Koyano et al., 2004, Koyano et al., 2009). Bearing in mind local cultural sensitivities associated with some of these sample types, parents in our studies had the option to choose which samples they were comfortable to be collected for testing.

The first part of this chapter describes experimental setup and optimisation, and then results from analysis of clinical samples are presented in latter sections. I wish to acknowledge and thank Dr Nyaxewo Mwaanza (UTH Paediatrics and Child Health

Department), the UTH HERPEZ team, and donors at LSHTM for providing saliva, hair, and nail samples used in assay optimisation experiments.

3.2 ASSAY OPTIMISATION

3.2.1 PCR strategy and primer sensitivity

The strategy employed in our analyses was to first screen samples using conserved gene targets, followed by genotyping hypervariable genes to determine strain variants. PCR assays and oligonucleotide primer sets previously validated and used in our laboratory (Bates et al., 2008, Mattick et al., 2004, Mattes et al., 2005) were adopted for studies in this thesis. For HCMV, screening PCR used the gB assay with the gB1/gB2 primer pair, while the gN and gO assays utilising gNup/gNlw and gOup/gOlw primer sets respectively were used for genotyping. Additionally, we used the GAPDH PCR assay with GAPup/GAPdown primers to validate the DNA extraction procedure (see methods, section 2.3).

The first step was to check assay sensitivity in our hands, after which a pilot assessment using donor samples from the LSHTM and UTH was conducted. Finally, the methods were applied to samples collected from our study cohorts. To check assay sensitivity, serial 10-fold dilutions of plasmid-cloned target genes were used, with PCR runs setup in duplicate. The assay sensitivity was determined as the lowest concentration of template that was detectable in both replicates (figure 3.1).

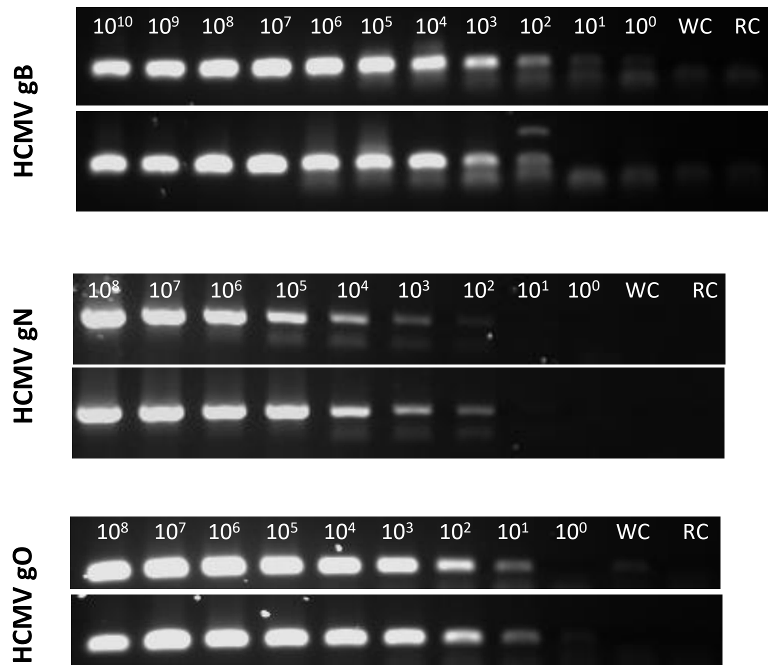


Figure 3.1. PCR primer sensitivity. Serial 10-fold dilutions of were run in duplicate for each gene target. The assay sensitivity was determined as the highest dilution for which the target sequence detection was reproducible, thus in our case 100 copies for gB and gN, and 10 copies for gO.

3.2.2 Real-time quantitative PCR setup and optimisation

Quantitative real time PCR was used to measure the HCMV viral load using the gB assay validated previously by our laboratory (Bates et al., 2008, Mattes et al., 2005). The setup and performance of the qPCR assay were in accordance with best practice as contained in the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin et al., 2010, Bustin et al., 2009). Our TaqMan qPCR assay used the gB1/gB2 primer pair together with the gB3 fluorescent probe. Standard curves were generated by linear regression analysis of the quantification cycle (Cq) measured for each amplification vs. the $\log_{10}$ copy number for each standard dilution, which are given by the relationship:

$$Cq = a*[\log (x)] + b$$

where: *a* is the slope, *b* is the y-intercept, and *x* is the standard quantity

To determine the assay limit of detection (LOD), plasmid-cloned standards were run in triplicate for each load level, covering seven dilution levels (10^6 to 10^0). Three trials were

performed, each on a different day. As per standard practice, the lowest concentration at which at least 95% of positive samples were detected was taken as the assay limit of detection (LOD) (Bustin et al., 2009). For the HCMV gB qPCR assay, this was 600 copies per 5µl template (table 3.1). For each qPCR run, standard curves were constructed and each 'unknown' sample DNA concentration was extrapolated from the regression line based on the sample's Cq value, and output given by the ABI 7500 v.2.0.6 Software.

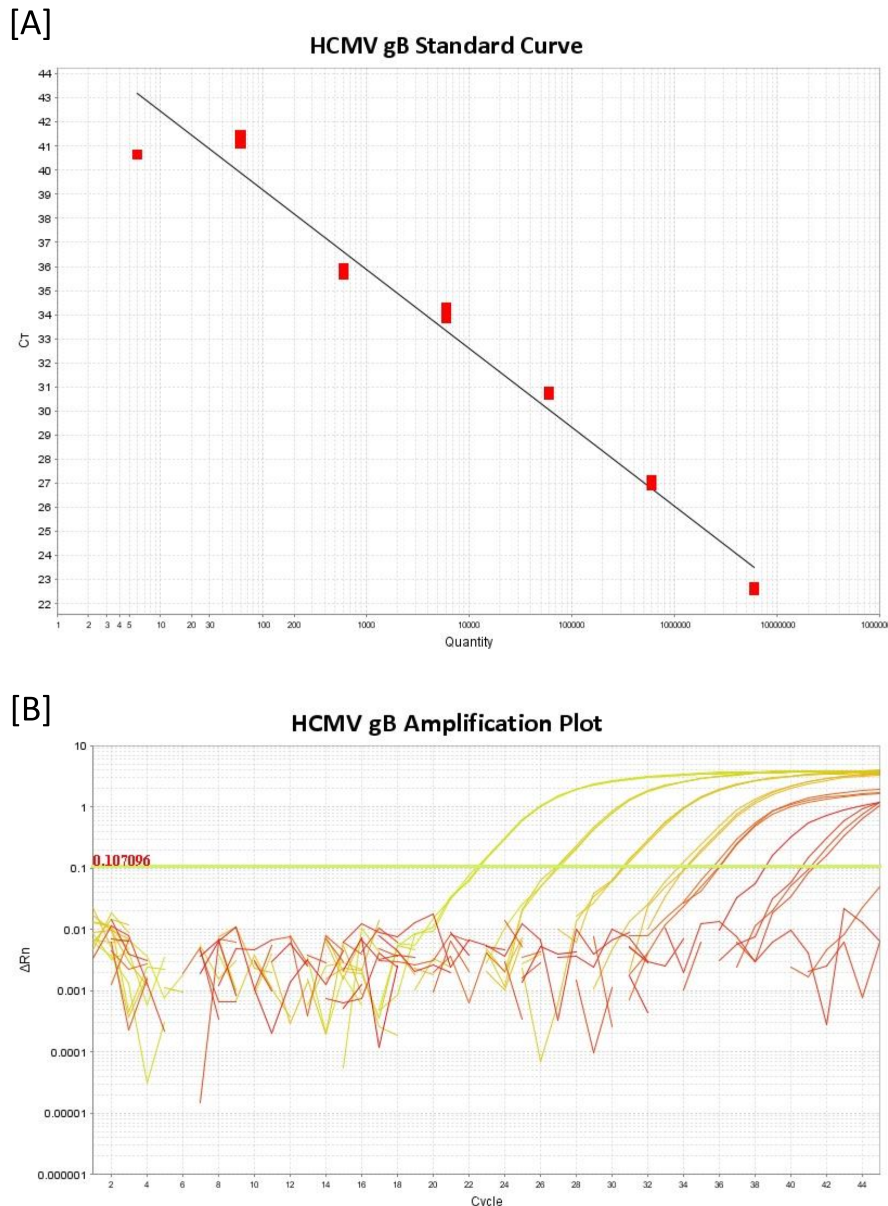


Figure 3.2. HCMV gB qPCR [A] standard curve and [B] amplification plot. Representative standard curve and amplification plot obtained during a gB qPCR run, demonstrating good correlation among replicates, with R^2 of 0.97 and run efficiency (Eff) of 101.7%. Standards were run in triplicate.

Copy number	Positive replicates trial 1	Positive replicates trial 2	Positive replicates trial 3	Total positive replicates	Total tested replicates	Frequency of detection
6000000	3	3	3	9	9	100%
600000	3	3	3	9	9	100%
60000	3	3	3	9	9	100%
6000	3	3	3	9	9	100%
600	**3**	**3**	**3**	**9**	**9**	**100%**
60	3	3	2	8	9	89%
6	1	3	2	6	9	67%

Table 3.1. HCMV gB qPCR Limit of Detection (LOD) determination. Three trials were performed, each consisting of a dilution series of the HCMV gB-containing plasmid. For each load level, three replicates were tested. The Limit of Detection (LOD) was determined to be 600 copies per 5µl template (highlighted).

3.2.3 Normalisation to the HCMV 09/162 International Clinical Reference Standard

The HCMV 09/162 International Clinical Reference Standard (ICRS) (National Institute of Biological Standards, Potters Barr, UK) was used to normalise our in-house plasmid-cloned standards (see Methods section 2.3.3). The ICRS is supplied as a lyophilised powder, and upon reconstitution in 1ml molecular biology grade water, according to manufacturer's instructions, DNA was extracted using the QIAamp® DNA Mini kit (QIAGEN®), following manufacturers' instructions. Serial 10-fold dilutions of reconstituted HCMV ICRS, were run concurrently with our in-house standards and standard curve generated to determine the relationship between the primary and secondary standard. Based on this a conversion factor of 100 was determined. However, since our assay used 5µl of a 1:2 dilution of DNA from a stock that had been eluted in 50µl but obtained from 200µl whole milk, readings per reaction were in fact equivalent to HCMV viral load copies per ml of whole milk.

3.3 PILOT STUDIES USING NON-INVASIVE SAMPLES

A final step in optimising our assays was to apply the assays to evaluate non-invasive tissue for virus diagnosis. A pilot was conducted in which saliva, nails, hair, and urine specimens from anonymous volunteers from LSHTM and UTH were tested. Overall these trials showed that sufficient DNA for virus analysis could be recovered from all these sample types (table 3.2). For saliva collection, the commonly available polyester-tipped

swabs proved effective, and adequate DNA could be recovered even when the swab dried, allowing flexibility for room temperature storage and transportation. During the pilot, we were able to amplify target genes from some of the donor samples, and derive nucleotide sequences (table 3.2). Figures 3.3 and 3.4 show examples of gel pictures demonstrating the amplification in DNA extracted from saliva of HCMV gB and HHV-6A/B U38 respectively.

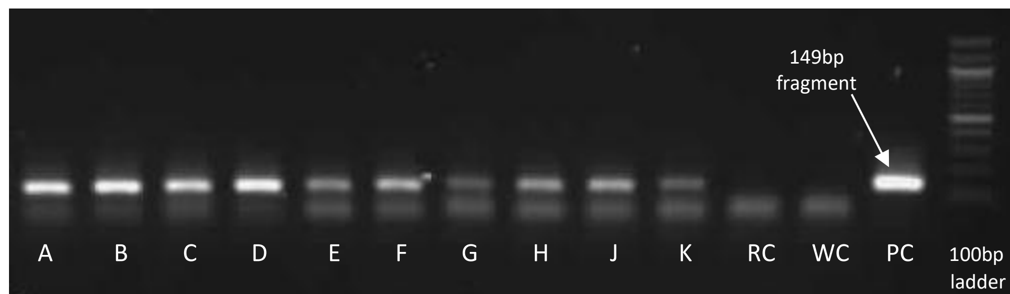


Figure 3.3. PCR amplification of HCMV gB: saliva samples. The 149bp fragment of the HCMV gB gene was amplified by PCR in DNA samples (labelled A-K) extracted from saliva. PCR products were loaded in 1% agarose gels and gel electrophoresis was performed in 1X TBE buffer. RC, Reagent (no-template) Control; WC, Water Control; PC, Positive Control (HCMV strain AD169).

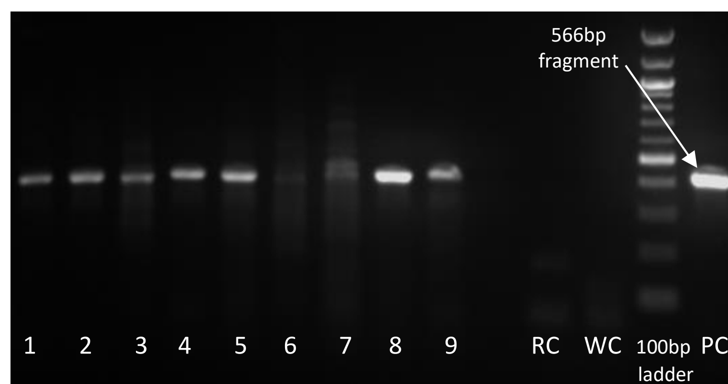


Figure 3.4. PCR amplification of HHV-6A/B U38: saliva samples. Shown is the PCR amplification of the 566bp fragment of the HHV-6A/B U38 gene. Samples tested (1-9) were DNA extracted from saliva. PCR products were loaded in 1% agarose gels and gel electrophoresis was performed in 1X TBE buffer. RC, Reagent (no-template) Control; WC, Water Control; PC, Positive Control (HHV-6A strain U1102).

Sample ID	Sample Type	Donor	DNA yield (ng/µl)	HCMV gB	HHV6-A/B	
					U38	Species
UN-1	Dry Saliva	Neonate	13.6	+	+	-
	Nail clippings	Neonate	9.1	+	-	-
UN-2	Dry Saliva	Neonate	55.3	+	-	-
	Urine	Neonate	15.1	+	+	6A
UN-3	Dry Saliva	Neonate	110.4	+	+	6A
UN-4	Dry Saliva	Neonate	64.7	+	+	6A
UN-5	Dry Saliva	Neonate	94.8	+	+	6A
UN-5	Nail clippings	Neonate	17.4	-	-	-
UN-6	Dry Saliva	Neonate	46.0	+	+	-
UN-7	Dry Saliva	Neonate	11.5	+	+	6B
UN-8	Dry Saliva	Neonate	26.8	+	+	-
SA	Liquid Saliva	Adult	15.2	-	+	6B
SB	Liquid Saliva	Adult	11.22	-	-	-
SC	Liquid Saliva	Adult	24.3	-	-	-
SWA	Dry Saliva	Adult	6.0	-	-	-
SWB	Dry Saliva	Adult	20.9	-	-	-
SWC	Dry Saliva	Adult	12.4	-	-	-
CM	Dry Saliva	Adult	12.9	-	-	-
NM	Dry Saliva	Adult	25.6	-	-	-
MM	Dry Saliva	Adult	44.6	-	-	-
CM	Hair follicles	Adult	1.8	-	-	-
NM	Hair follicles	Adult	5.7	-	-	-
MM	Hair follicles	Adult	1.5	-	-	-
CM	Nail clippings	Adult	13.9	-	-	-
NM	Nail clippings	Adult	1.3	-	-	-
MM	Nail clippings	Adult	3.8	-	-	-

Table 3.2. Donor samples used in assay set-up and optimization. Sufficient DNA could be recovered from all sample types tested. HCMV, HHV-6A and HHV-6B could be detected and sequenced successfully.

3.4 EVALUATION OF HCMV, HHV-6A, AND HHV-6B IN NEONATES

A total of 117 neonates were recruited to the study: 101 from the LW and 16 from the NICU. In LW, all neonates were recruited within 1 hour of birth (median 10min, IQR 5 – 26min) and in all cases but one, saliva samples were collected before initiation of breast milk feeding. There were four cases of stillbirths among participants from LW. In NICU the median postnatal age at recruitment was 4 days (IQR 1.0 – 9.3 days) and half (8/16) of the participants had already commenced breast milk feeding at recruitment. Table 3.3 gives a summary of descriptive characteristics of the study participants.

3.4.1 Samples

For each participant, two samples were collected: Saliva plus either umbilical tissue, hair or nails. Saliva was collected by swabbing the neonate's mouth with a sterile polyester-tipped swab (Peel pouch Dryswab™, Medical Wire & Equipment), which was air-dried then placed in a separate, labelled grip seal bag. Additionally, approximately 1cm^3 of umbilical tissue was collected using a sterile surgical blade shortly after birth in LW or from above the umbilical cord clamp in NICU. Where collection of umbilical tissue was declined or not feasible, parents could opt for nail clippings or hair follicles to be collected. Nail clippings were obtained using sterilized nail clippers. Hair follicles were plucked from the scalp. Overall, we collected saliva from 116 participants, umbilical tissue from 115, hair from 1, and nail clippings from 1 participant, as outlined in Table 3.5.

	Labour Ward		NICU		TOTAL	
	Freq	%	Freq	%	Freq	%
Breastfeeding at recruitment						
Not Breastfed	100	99.0	8	50.0	108	92.3
Breastfed	1	1.0	8	50.0	9	7.7
Total	101	100.0	16	100.0	117	100.0
Gender						
Female	59	58.4	5	31.3	64	54.7
Male	42	41.6	11	68.8	53	45.3
Total	101	100.0	16	100.0	117	100.0
Birth Weight (Kg)						
< 1.0 (ELBW)	4	4.0	3	18.8	7	6.0
< 1.5 (VLBW)	3	3.0	8	50.0	11	9.4
< 2.5 (LBW)	20	19.8	2	12.5	22	18.8
2.5–4.5 (NBW)	74	73.3	2	12.5	76	65.0
> 4.5 (Macrosomia)	0	0.0	1	6.3	1	0.9
Total	101	100.0	16	100.0	117	100.0
Mode of Delivery						
SVD	95	94.1	14	87.5	109	93.2
CS	5	5.0	1	6.3	6	5.1
Instr.	1	1.0	0	0.0	1	0.9
NS	0	0.0	1	6.3	1	0.9
Total	101	100.0	16	100.0	117	100.0
Maternal HIV Serostatus						
Negative	83	82.2	9	56.3	92	78.6
Positive	14	13.9	6	37.5	20	17.1
NK	4	4.0	1	6.3	5	4.3
Total	101	100.0	16	100.0	117	100.0

Table 3.3. Neonatal cohort: descriptive statistics. VLBW, Extremely Low Birthweight; Very Low Birthweight; LBW, Low Birthweight; NBW, Normal Birthweight; SVD, Spontaneous Vaginal Delivery; CS, Caesarean section; Instr., Instrumental; NS, Not stated; NK, Not Known.

[A] REASON FOR ATTENDANCE / DIAGNOSIS – LW	Frequency	%
Normal labour	35	34.7
Multiple pregnancy	8	7.9
Raised blood pressure	8	7.9
Preterm labour	7	6.9
Dangerous presentation (breech / face / hand prolapse)	5	5.0
Pre-Eclampsia	5	5.0
Big baby	4	4.0
Previous C/S	4	4.0
Prolonged labour	4	4.0
Foetal distress	3	3.0
Post-dates	3	3.0
IUFD / BOH	2	2.0
Multiparity	2	2.0
PROM	2	2.0
Other*	7	6.9
Not stated	2	2.0
Total	**101**	**100.0**

[B] DIAGNOSIS – NICU	Frequency	%
Prematurity / Sepsis / RDS	12	75.0
Ichthyosis Valgaris / Sepsis	1	6.3
Macrosomia	1	6.3
Meconium aspiration / Sepsis	1	6.3
Meconium aspiration / SGA	1	6.3
Total	**16**	**100.0**

Table 3.4. [A] Reason for presenting to UTH labour ward (LW) and [B] infant's diagnosis in NICU. C/S, Caesarean Section; IUFD, Intra-uterine foetal demise; BOH, Bad obstetric history; PROM, Premature rapture of membranes; RDS, Respiratory distress syndrome; SGA, Small for gestational age. *'Other' reasons for attendance included: elderly prime gravida, assault in pregnancy, fever, headache, vomiting, and cervical growth.

Sample Type	LW	NICU	Total
Saliva	100	16	116
Umbilical tissue	101	14	115
Hair	-	01	01
Nails	-	01	01

Table 3.5: Type of sample collected by ward. Saliva and Umbilical tissue were accepted and most preferred non-invasive samples in our studies.

3.4.2 HCMV detection

Of 100 saliva samples collected from the LW, one sample was positive for HCMV gB by the screening PCR assay (figure 3.5). The sample was from an HIV-exposed female neonate with normal birthweight (2.8Kg). The HCMV DNA load was quantified using the qPCR assay and showed a high load of 2.3×10^7 copies/ml of saliva. Further, Sanger sequencing for HCMV gO and gN showed gO1c and gN4c (figure 3.6), in keeping with the known genotype gO1c and gN4c linkage (Bates et al., 2008, Mattick et al., 2004) (see table 1.1).

Screening of umbilical tissue showed high background human DNA, and no HCMV DNA was distinguishable. Additional laboratory projects are planned to investigate this further.

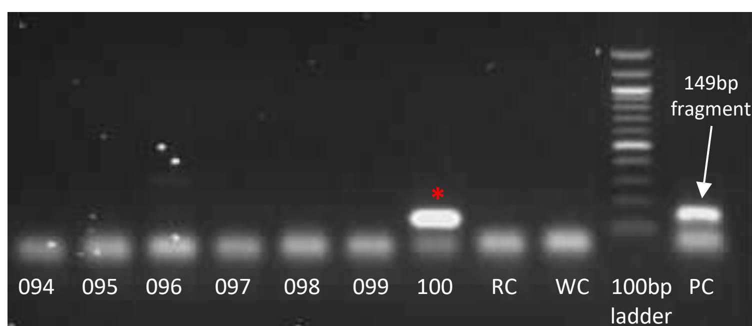
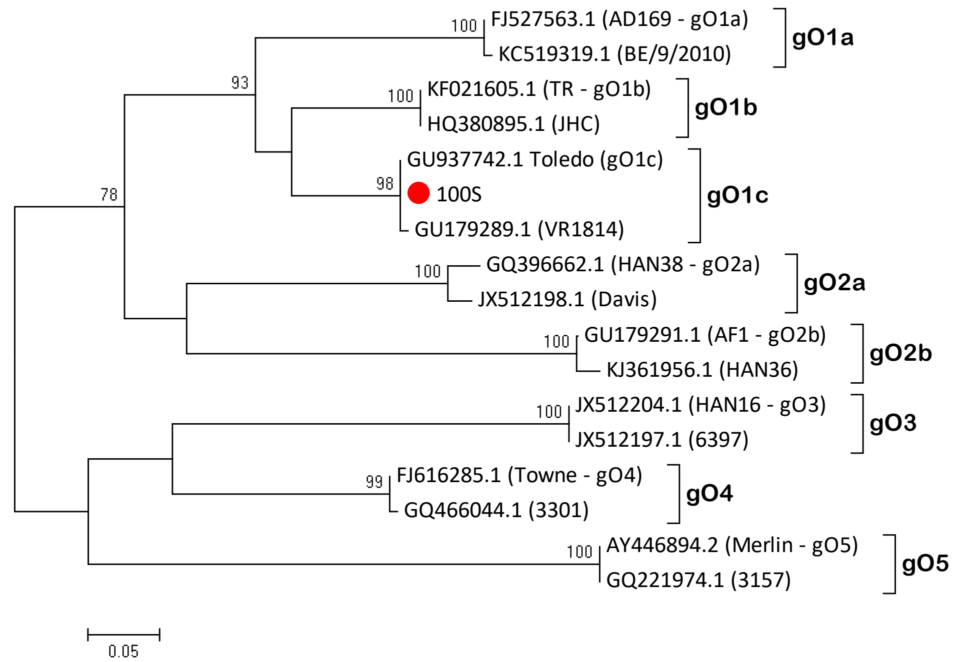


Figure 3.5. PCR amplification of HCMV gB in sample 100S. The 149bp fragment of HCMV gB was amplified in Sample 100S (red asterisk). PCR products were loaded in 1% agarose gels and gel electrophoresis was performed in 1X TBE buffer. RC, Reagent (no-template) Control; WC, Water Control; PC, Positive Control (HCMV strain AD169).

[A]



[B]

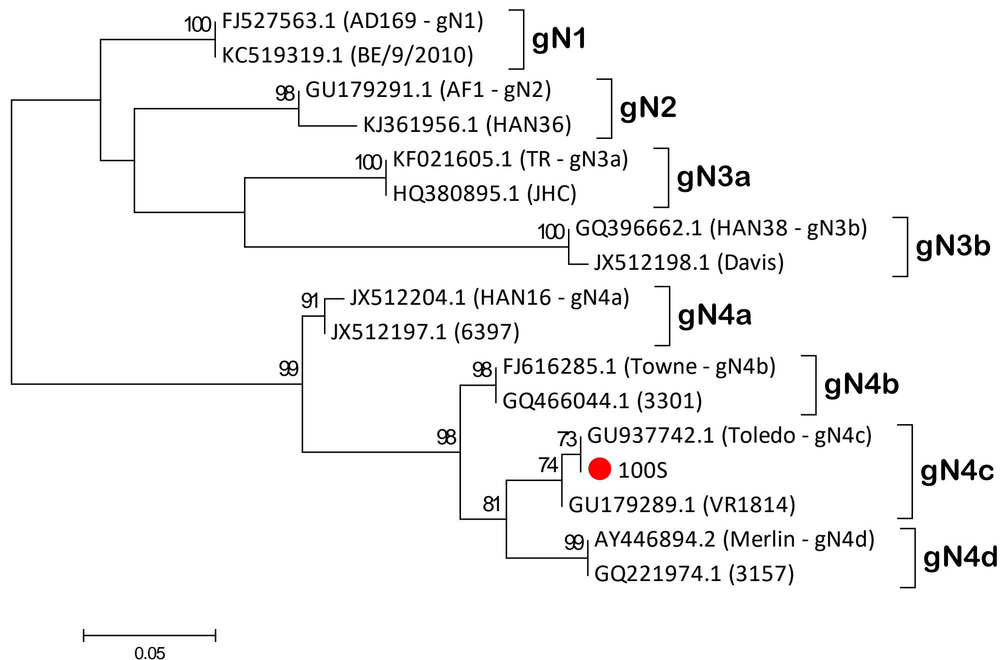


Figure 3.6. Molecular phylogenetic analyses of sample 100S [A] gO and [B] gN. The Maximum Likelihood method based on the JTT matrix-based model was used to infer relationships of the 17 amino acid sequences included in each phylogenetic analysis. Consensus trees were derived after applying 1000 bootstrap replicates. In each tree, sample 100S is indicated by the red circular marker. Analyses were conducted in MEGA 6.06 (Tamura et al., 2013).

3.4.3 HHV-6A/B detection

HHV-6A/B screening of saliva using the HHV6 U38 assay was negative for all samples from both the LW and NICU cohorts. However preliminary results from examination of umbilical tissue showed 54/117 (46.2%) probable amplifications of the U38 target. Given the 1% prevalence of chromosomally integrated HHV6-A/B reported elsewhere (Hall et al., 2004, Daibata and Miyoshi, 1999, Dahl et al., 1999, Adams et al., 1998, Hall et al., 2008, Hall et al., 2010, De Bolle et al., 2005), our prevalence finding seemed unusually high, and we therefore shall follow this up as a separate project.

3.5. DISCUSSION

This chapter set out to investigate prevalence of congenital HCMV as a potential source and route of transmission of HCMV infection associated with the widespread detrimental effects of HCMV on child health and development shown earlier in Zambia (see introduction, chapter 1) (Gompels et al., 2012, Sanz-Ramos et al., 2013). At the start of this thesis, the prevalence of congenital HCMV in Zambia was unknown. Two available reports on congenital HCMV in sub-Saharan Africa showed 5.4% prevalence in the Gambia (van der Sande et al., 2007) and in a small sample, an average of about 10% [2.7% (1/37) among HIV-negative and 29% (4/14) among HIV-infected] in Kenya (Slyker et al., 2009a). Our analysis of new-borns in the normal labour ward showed congenital HCMV was 1% (1/100), which is consistent with rates published from other countries in Europe, North America and Southeast Asia (see section 1.8.1) (Fowler and Boppana, 2006, Fowler et al., 1999, Fowler et al., 2003, Kaye et al., 2008, Yamamoto et al., 2013, Yamamoto et al., 2010).

During the course of this thesis, three studies in sub-Saharan Africa also published information on prevalence of congenital HCMV. The first study analysed high-risk neonates under 3 weeks age at UTH neonatal intensive care unit, NICU, in Lusaka, Zambia, showing a prevalence of 3.8% (Mwaanza et al., 2014), and a link to maternal HIV infection. This was the first ever study to examine and report congenital HCMV in Zambia. However, in that study no measures were in place to distinguish breastmilk-fed from non-breast milk-fed neonates. Given that breast milk feeding is widely practiced in

UTH NICU, HCMV is secreted in breast milk, and infant feeding can affect efforts to diagnose congenital infection (Fujisaki et al., 1998, Dunne and Jevon, 1993), it is not clear whether some of the HCMV infections may have been early post-natal infections. Further, NICU receives births with complications, including infants with HIV-positive mothers, therefore may also contribute to the higher prevalence in this ward compared to the normal labour ward under study in this thesis. Another study published during the course of this thesis was conducted in South Africa among neonates born only to HIV infected mothers who had received prenatal ARV prophylaxis. The birth prevalence of congenital HCMV in that study was 2.6%, and infant HCMV was associated with advanced maternal immunosuppression. However, again, the study did not determine prevalence among 'normal' HIV-unexposed infants (Manicklal et al., 2014). A third study was conducted in Zimbabwe, among treatment-naïve HIV-infected infants, and showed a high rate of 11% congenital HCMV infection. In that study, other pathogens including EBV were also found in an unusually high prevalence congenitally (Gumbo et al., 2014). However, that study did not sample HIV-uninfected neonates for comparison. Unlike these highlighted studies, in our study presented here, both the breastfeeding status and maternal HIV exposure, were taken into account in investigating congenital HCMV infection. The neonate congenitally infected with HCMV in our study was maternally HIV-exposed. As described in other studies, maternal HIV is a risk factor for congenital HCMV transmission (Mwaanza et al., 2014, Gumbo et al., 2014, Slyker et al., 2009a) (see introduction section 1.8) and with the high DNA load of $2.3x10^7$ copies/ml the neonate was at heightened risk of HCMV deleterious effects. Although efforts were made to trace the neonate and mother, in keeping with the ethical principles governing our study, unfortunately they could not be reached as the family was domiciled in another district outside Lusaka. It would have been desirable to link the neonate to available neurological services at UTH for assessment and follow-up clinical management, as well as investigate and link the mother to appropriate HIV management services, which are publicly available in the Zambian healthcare system. At present, therefore, in the maternally HIV negative group on the normal labour ward the prevalence is below 1% as in many other countries.

In this study, it was gratifying to note that non-invasive samples were welcomed by parents for the investigation of infection, particularly during the culturally and clinically

sensitive neonatal period. We have showed that these samples can be used in the local setting to successfully detect, quantify, and sequence viral pathogens. In particular, use of saliva provides a simple, convenient, quick, pain-free, safe, and cost-effective alternative sample for diagnostic use.

CHAPTER 4

ANALYSIS OF HUMAN CYTOMEGALOVIRUS IN BREAST MILK OF HIV-POSITIVE AND NEGATIVE MOTHERS IN ZAMBIA

4.1 INTRODUCTION

This chapter describes results of analyses undertaken to understand the effect of breastfeeding duration and maternal HIV status on HCMV viral load, secretion, and transmission in Zambia. The CIGNIS study showed high seropositivity and early effects of HCMV on growth, therefore it was important to determine sources for this transmission and effects of maternal HIV. To do this, both HIV-positive and negative mothers were compared in assessing the effects of breastfeeding duration and HCMV viral load dynamics. Studies on HIV-negative women had showed that viral load in breast milk can affect infant transmission (Jim et al., 2009, van der Strate et al., 2001, Hamprecht et al., 2003, Yasuda et al., 2003). However, comparisons of HIV-negative to HIV-positive women had not yet been demonstrated. Further, the effect of virus strain or genotype has not yet been analysed. Furthermore, the effect of the longer breastfeeding duration common in the Sub-Saharan Africa region had also not been assessed. In order to address these critical issues, HCMV DNA loads and genotypes were analysed directly in breast milk from both HIV-positive and negative women in the previously described BFPH cohort (Collin et al., 2006, Kasonka et al., 2006) in Zambia. To do this DNA was extracted from all available breast milk samples collected from both HIV-positive and negative Zambian mothers at postpartum week (W) 16, as well as 40 matching sets of milk collected earlier at day (D) 3, D14, W4, W9, and W12 (n=261, see Methods, section 2.3) were analysed. First, the samples were screened by PCR to determine whether in the Zambian population the duration of HCMV shedding in breast milk is any different from European and South East Asian populations. Next, HCMV DNA loads were quantified by a TaqMan® qPCR assay in breast milk samples collected longitudinally at 6 time-points during the first 16 weeks postpartum. Further, HCMV hypervariable genes encoding envelope glycoproteins gO and gN were genotyped in selected W4 and W16 samples so as to evaluate HCMV genotype restriction in the breast milk compartment. Crucially, all these analyses compared HIV-positive to HIV-negative

women, enabling the effect of HIV infection on HCMV viral loads and shedding dynamics to be defined in our study population from an HIV-endemic region of Sub-Saharan Africa.

4.2 BREAST MILK DNA FROM THE BFPH COHORT

Of the 261 W16 breast milk samples available for this study, 118 (45.2%) were from HIV-positive while 143 (54.8%) from HIV-negative mothers. For the HCMV kinetics evaluation, we additionally sought out all available longitudinal sample sets collected at 5 earlier time points (D3, D14, W4, W9, W12), which comprised 20 sets from HIV-positive and 20 from HIV-negative mothers. DNA was extracted directly from the milk, as detailed in Methods section 2.3.1. Overall, DNA was extracted from a total of 461 milk samples for subsequent analysis. Prior to screening for HCMV, DNA extractions were validated (see Methods, section 2.3.2) and all extractions had good quality, with 280nm/260nm absorbance ratios averaging 1.8 (data not shown). The 104 bp fragment of the housekeeping gene encoding human GAPDH (Methods section 2.3.2) was used to PCR-screen all DNA samples extracted from the milk, as exemplified in figure 4.1. All samples were positive for the GAPDH gene. Negative and positive controls were included with every batch of 11 samples in all assay runs, and there was no evidence for contamination using our unidirectional three-room PCR containment procedures (see Methods).

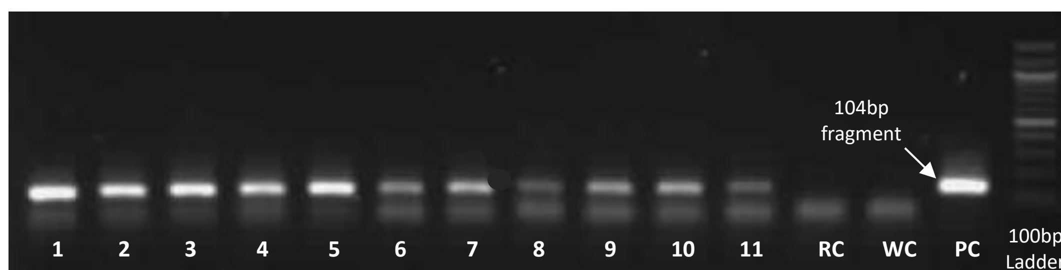


Figure 4.1. PCR amplification of GAPDH in breast milk. The picture shows PCR amplification of the 104bp GAPDH fragment in DNA samples (here labelled 1–11) extracted from breast milk. PCR products were loaded in 1% agarose gels, and gel electrophoresis was performed in 1X TBE buffer at 100V for 45 minutes. RC, Reagent (no-template) Control; WC, Water Control; PC, Positive Control (DNA extracted from human saliva).

4.3 HCMV DETECTION IN BREAST MILK

The breast milk DNA samples were then screened qualitatively for HCMV using our validated HCMV gB assay. As demonstrated in section 3.1, this assay had a sensitivity of 100 copies per reaction and amplified a 149 bp fragment of the conserved HCMV gB gene. Each PCR run included 2 negative controls i.e. a no-template (reagent only) control and a water-as-template control, in addition to the positive control (HCMV strain AD169). The 40 longitudinal milk sample sets (D3, D14, W4, W9, and W12) were 100% positive by the HCMV gB qualitative PCR assay. Among the 261 W16 samples overall HCMV prevalence was 76.3% (199/261). Stratifying the complete W16 set for HIV showed that among HIV-positive women HCMV prevalence was 83.9% (99/118), compared to 63.6% (91/143) among HIV-negatives. This difference in HCMV prevalence in the two groups was statistically significant, $p<0.001$. Thus in this population of Sub-Saharan African mothers, we detected HCMV shedding in breast milk from day 3 to week 16 postpartum, with a significantly larger proportion of HIV-infected women having DNAlactia at week 16 compared to their HIV-negative counterparts. Figure 4.2 depicts some of the HCMV gB screening PCR results for breast milk samples in our study.

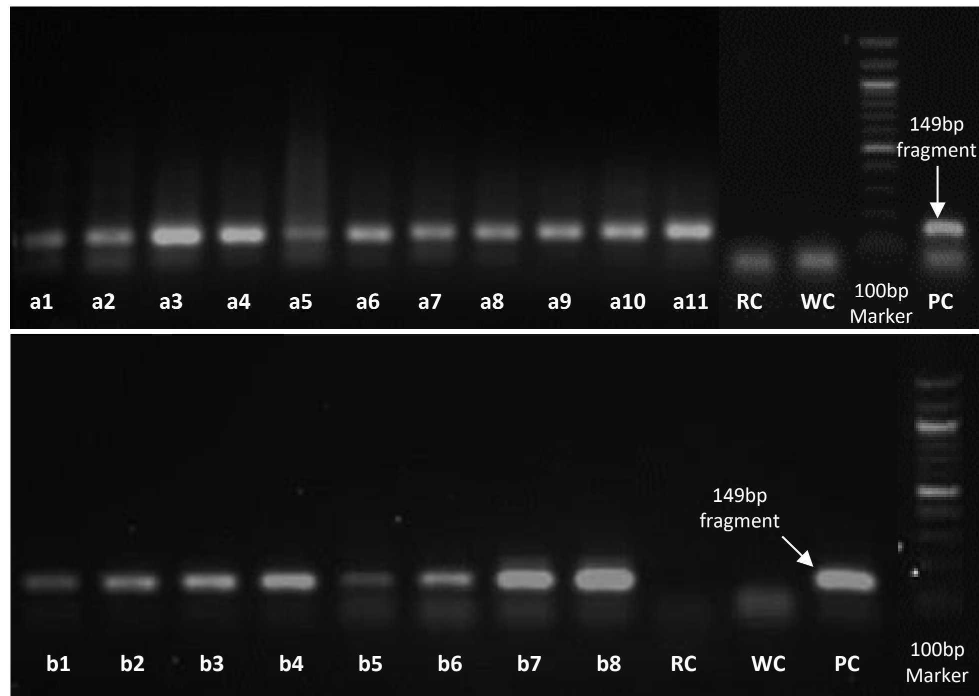


Figure 4.2. HCMV gB amplification in breast milk. Panels show examples of the 149bp HCMV gB gene fragment amplified in DNA extracted from breast milk (here labelled a1–a11 and b1–b8). PCR products were loaded in 1% agarose gels and gel electrophoresis performed in 1X TBE buffer at 100V for 45 minutes. RC, Reagent (no-template) Control; WC, Water Control; PC, Positive Control (HCMV strain AD169).

4.4 HCMV VIRAL LOADS IN BREAST MILK

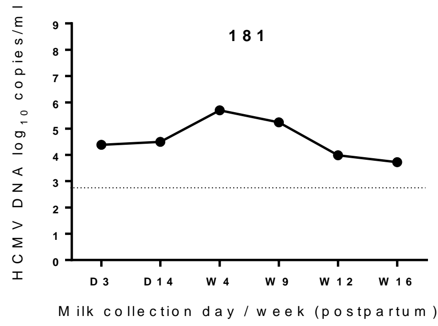
The qualitative assay showed differences in detection of HCMV in breast milk of HIV-positive compared to HIV-negative mothers. In order to compare HCMV viral loads in breast milk from the two groups, real time qPCR was applied. This used the TaqMan® qPCR assay described in section 2.3.3 (Bates et al., 2008, Mattick et al., 2004) as adapted from Mattes (Mattes et al., 2005) and validated in section 3.2.2 on an Applied Biosystems® 7500 Fast Real-Time PCR System (Applied Biosystems® Inc.). Each of the qPCR runs were calibrated by including plasmid-cloned HCMV gB standards covering 7 logs concentrations as well as negative (no template) controls. Samples were diluted 1:2 in nuclease-free molecular biology grade water. Each reaction contained 5µl of sample or control and was run in triplicate (see Methods, sections 2.3.3 and 2.4). The in-house plasmid-cloned HCMV gB standards were normalised to the HCMV 09/162 International Clinical Reference Standard (National Institute of Biological Standards, Potters Barr, UK) (see methods). Using these calibration controls and standards, the assay had a limit of detection (LOD) of 600 copies per reaction, with 100% reproducibility in triplicate and in data analyses samples with viral loads below the LOD were assigned a value set at half the LOD. Data were log transformed and plotted using GraphPad Prism version 6.0, then transformed back; thus medians and geometric means are presented in comparisons between HIV-positive and negative groups.

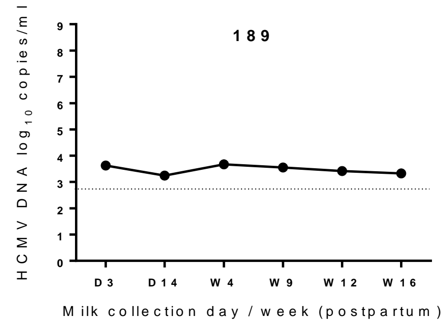
HCMV Viral Load Kinetics in breast milk

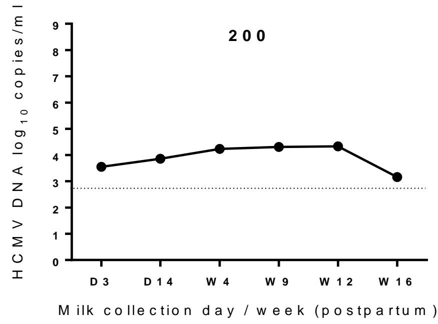
The 40 sets of longitudinal breast milk samples collected at six time-points, namely D3, D14, W4, W9, W12, and W16, were used to investigate HCMV secretion kinetics by examining DNA loads over the 16-week postpartum period covered. TaqMan® realtime qPCR was performed for each sample (in triplicate) at each of these six time-points and standard curves were generated and used to compute HCMV DNA copies per reaction. Since each reaction contained 5µl of 1:2 diluted sample and for each sample 200µl whole milk were used to extract DNA, which was eluted in 50µl of nuclease-free water, the quantity obtained by qPCR represented HCMV DNA copies in 10µl of whole milk. This value was multiplied by a factor of 100 to obtain the HCMV DNA copies/ml of milk. The HCMV DNA copies/ml values were log10 transformed and plotted on a log linear scale as depicted in figures 4.3 and 4.4 for HIV-positive and HIV-negative mothers respectively. These data were then amalgamated and the geometric mean (gMean) at

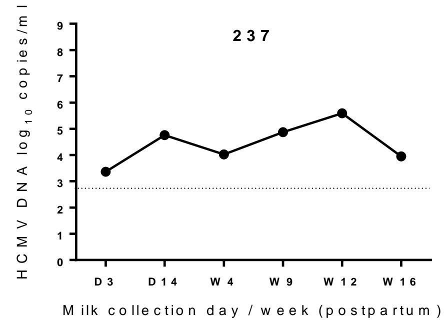
each time-point was computed for each of the two groups of mothers (figures 4.5). At five of the six time-points, more milk samples from HIV-positive mothers had viral loads above the assay sensitivity cut-off compared to milk from HIV-negative women and by W16, 90% (18/20) of milk from HIV-positive mothers was above assay sensitivity cut-off, compared to 55% (11/20) from HIV-negative mothers, p = 0.031. At each of the 6 time-points, gMean viral loads were higher among HIV-infected mothers (figure 4.5).

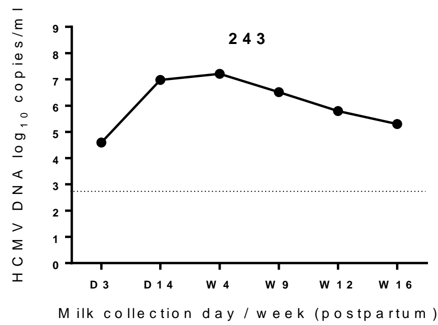
[A]

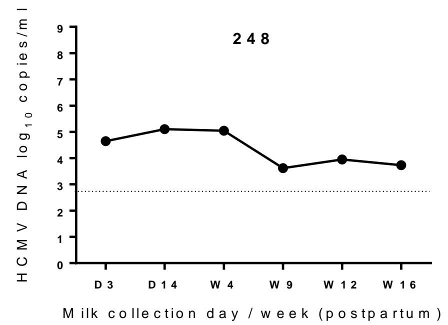


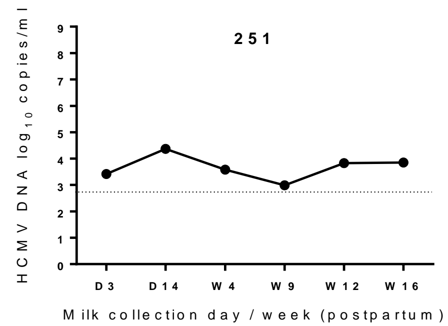




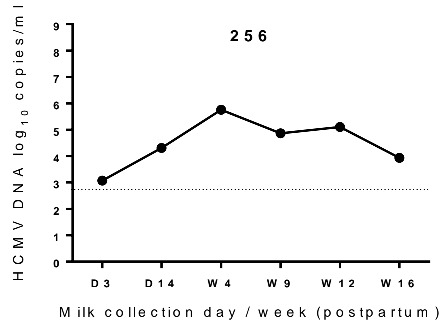


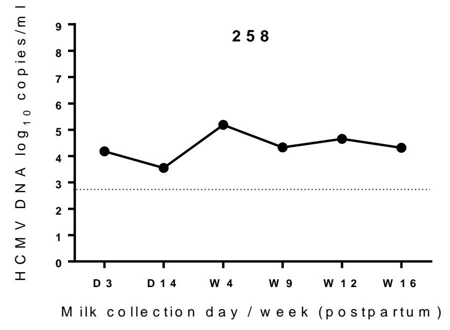


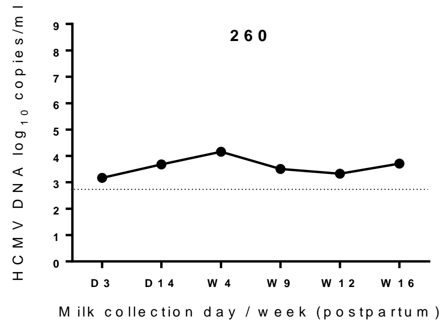


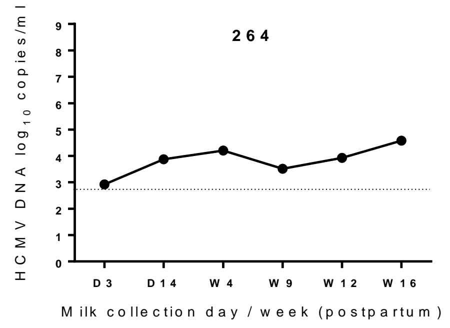


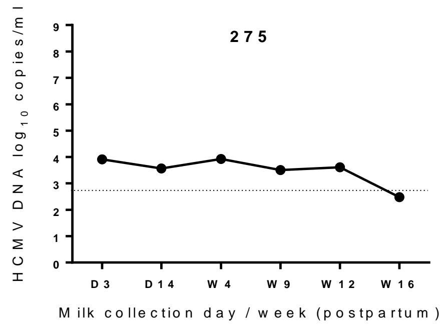
[B]

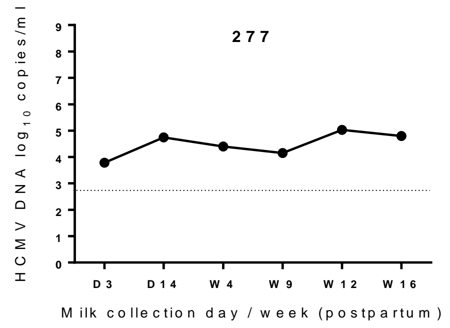


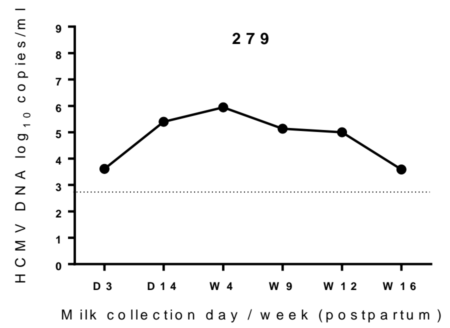












[C]

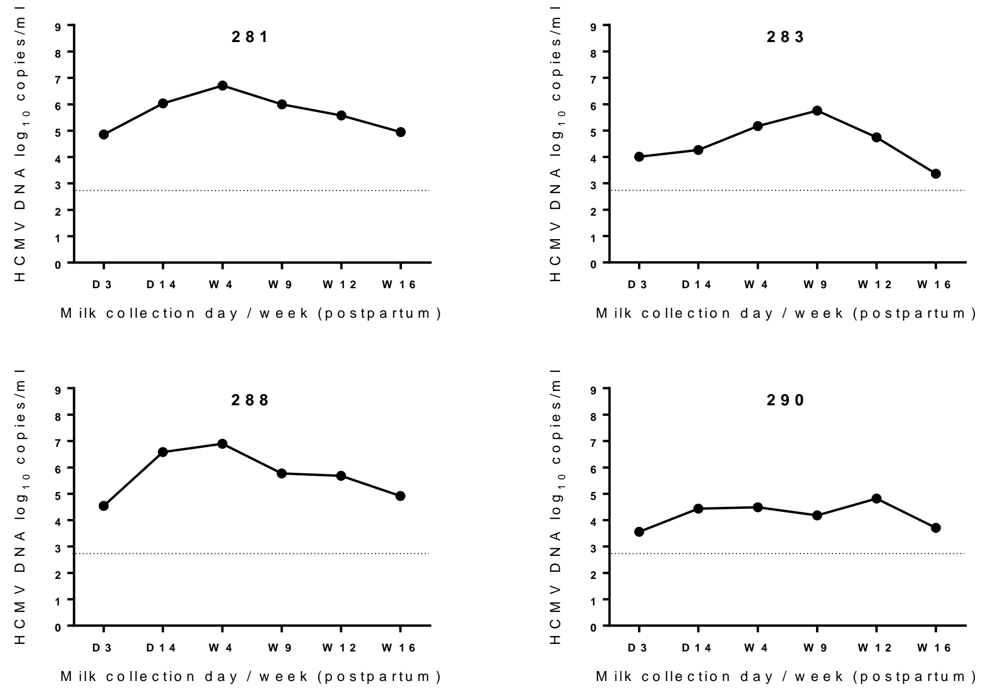
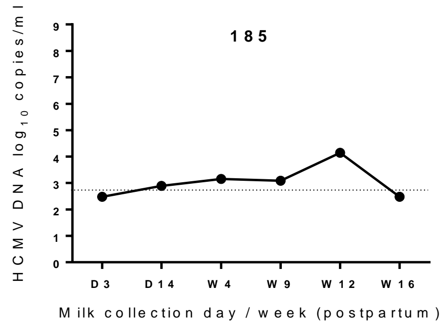
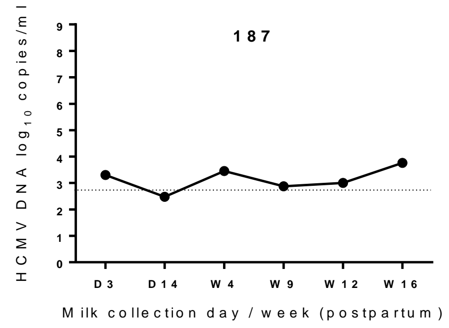
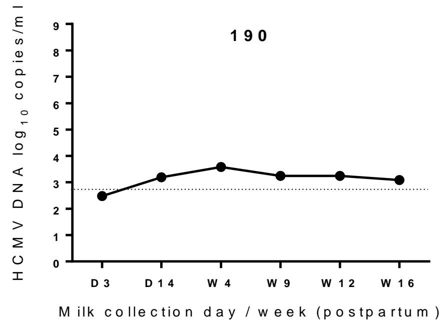


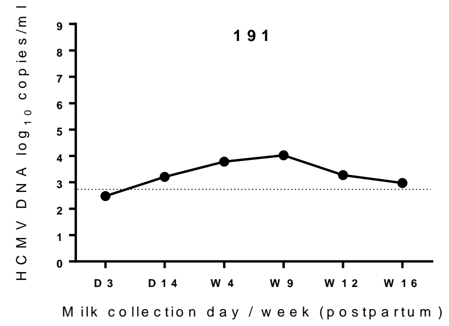
Figure 4.3. HCMV viral load kinetics in breast milk from HIV-positive mothers. Panels [A] to [C] show $\log_{10}$ HCMV DNA loads in breast milk for each of 20 HIV-positive women at 6 time-points during the first 16 weeks postpartum. The assay sensitivity cut-off is shown by the lower horizontal dotted line. Bold numbers at the top centre of each graph are laboratory identifiers.

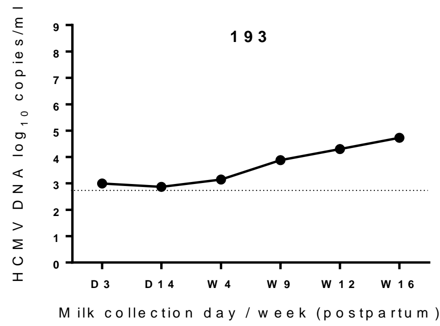
[A]

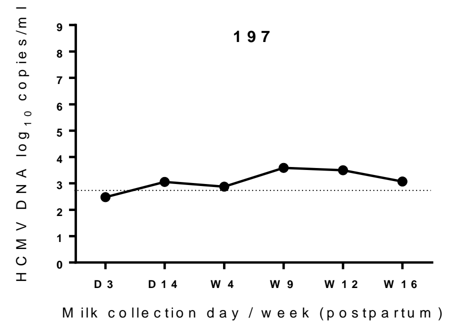


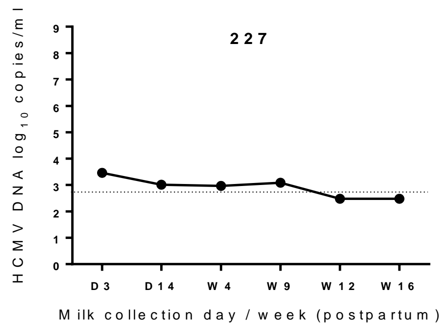


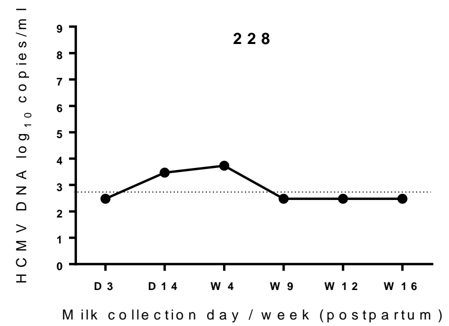




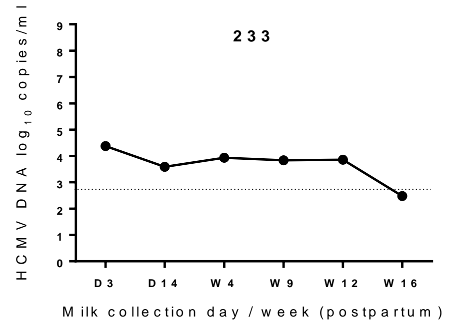


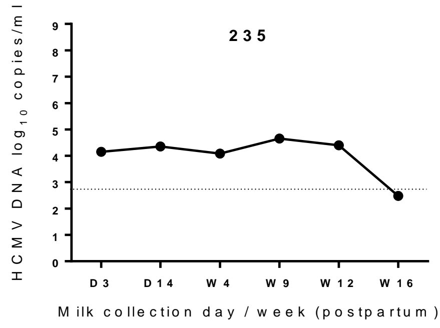


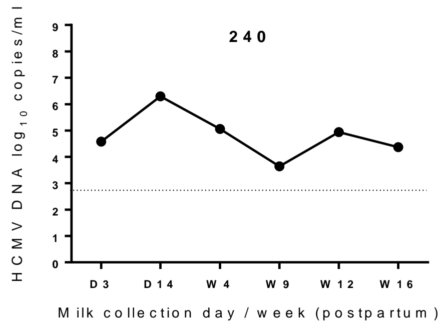


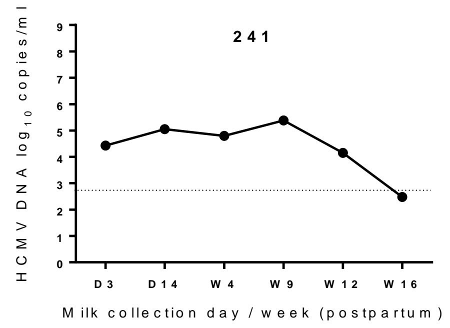


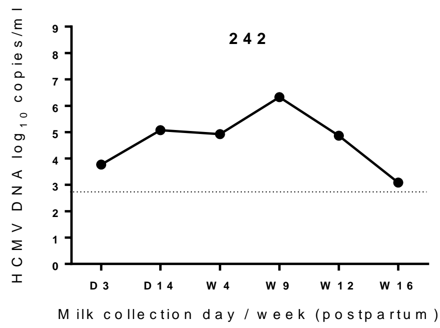
[B]

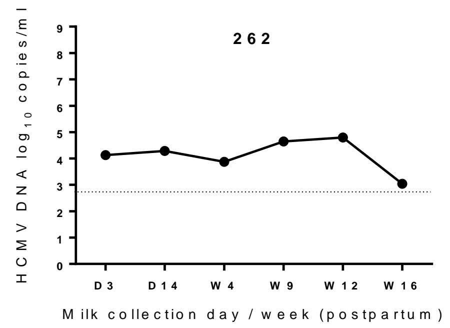












[C]

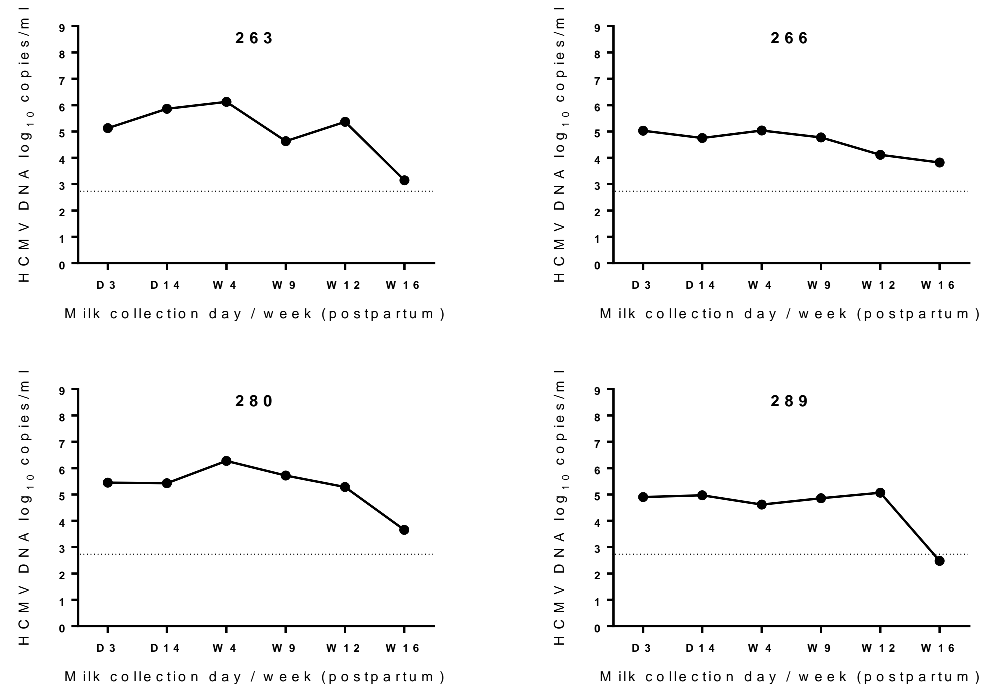


Figure 4.4. HCMV viral load kinetics in breast milk from HIV-negative mothers. Panels [A] to [C] show $\log_{10}$ HCMV DNA loads in breast milk for each of 20 HIV-negative women at 6 time-points during the first 16 weeks postpartum. The assay sensitivity cut-off is shown by the lower horizontal dotted line. Bold numbers at the top centre of each graph are laboratory identifiers.

Comparisons of HCMV viral load and rates of secretion in breast milk from HIV-positive and negative women

Overall, HCMV levels were similar at day 3, rising by W4, and then declined or persisted to W16. In D3 breast milk, gMean HCMV viral loads in HIV-positive and HIV-negative mothers were 3.88 log10 copies/ml (95% CI 3.61 – 4.15) and 3.71 $\log_{10}$ copies/ml (95% CI 3.27 – 4.15) respectively, translating into 7,568 copies/ml and 5,116 copies/ml respectively. These viral loads were not significantly different, p=.89. After D3, gMean viral loads significantly increased in both HIV-positive and negative groups and, interestingly, peak levels were reached at week 4 (figures 4.6 and 4.7). From D3 to W4 there was an overall increase of 1.2 $\log_{10}$ copies/ml in gMean viral loads in HIV-positive mothers compared to only 0.5 $\log_{10}$ copies/ml in HIV-negative mothers. At W4 the gMean viral load was significantly higher in HIV-positive mothers, with 5.04 $\log_{10}$ copies/ml (95% CI 4.49 – 5.58) versus 4.21 $\log_{10}$ copies/ml (95% CI 3.74 – 4.67) among HIV-negatives, p=.026 (figure 4.6). From W4 to W9 the gMean viral load declined by 0.5 $\log_{10}$ copies/ml in HIV-positive mothers, and remained stable at this level until W12, after which there was a further decline of 0.6 $\log_{10}$ copies/ml to reach 3.90 $\log_{10}$ copies/ml (95% CI 3.51 – 4.29), the W3 gMean level, by W16. In contrast, in HIV-negative mothers the gMean HCMV load plateaued from W4 to W12, averaging 4.05 $\log_{10}$ copies/ml, then from W12 declined sharply by 0.7 $\log_{10}$ copies/ml to 2.99 $\log_{10}$ copies/ml (95% CI 2.69 – 3.29) at W16. By the last time point, W16, in HIV-negative women, the gMean HCMV DNA was markedly below the D3 level of 3.71 $\log_{10}$ copies/ml (95% CI 3.27 – 4.15). Hence, at week 16 the gMean HCMV viral load was 0.91 $\log_{10}$ copies/ml higher in HIV-positive mothers compared to HIV-negative mothers and statistically significant, with p<.001 (figure 4.5).

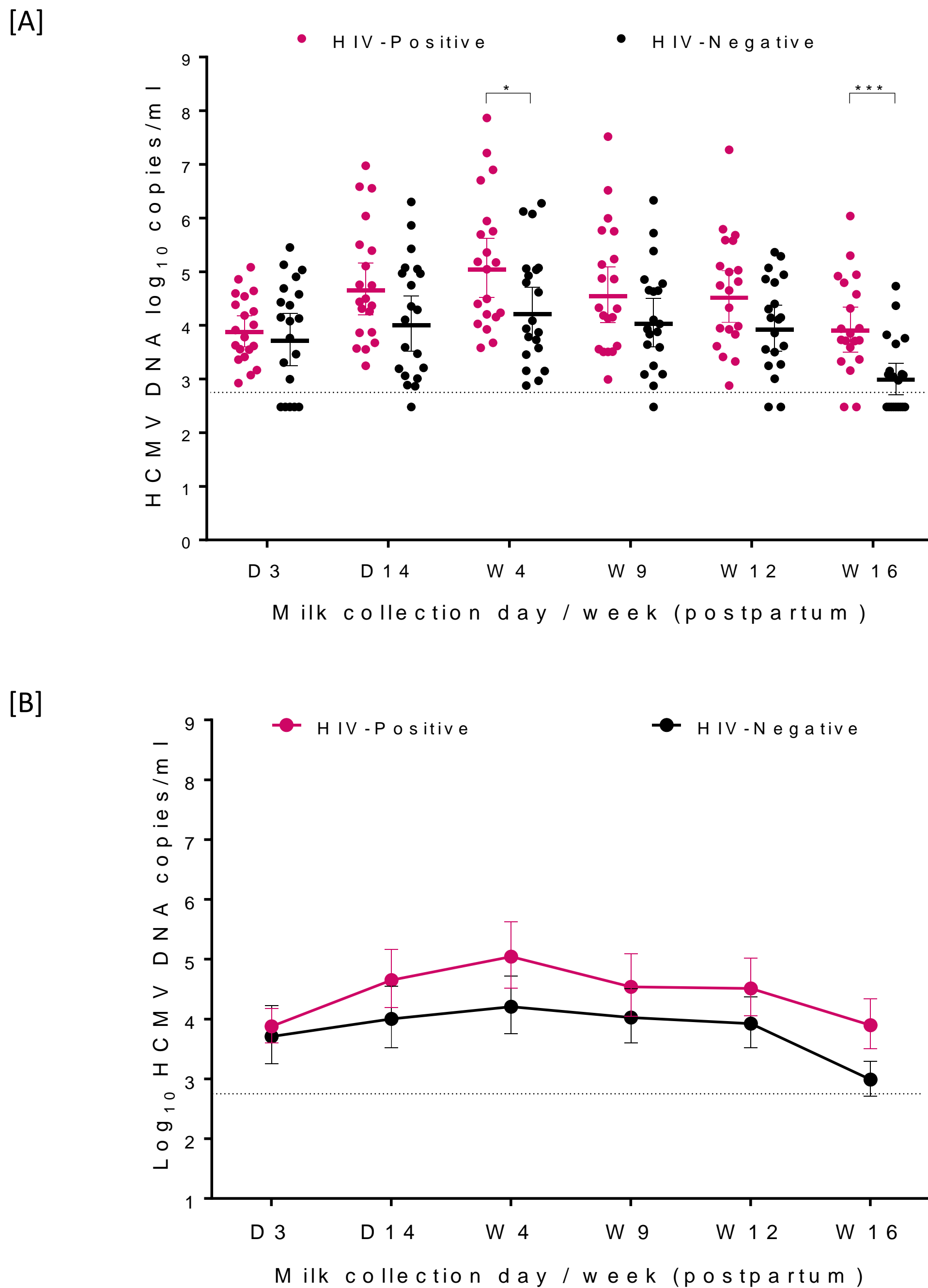


Figure 4.5. Kinetics of HCMV shedding in breast milk of HIV-positive and negative mothers. Comparison of breast milk HCMV DNA load kinetics between HIV-positive (pink dots/line) and negative (black dots/line) mothers during the first 16 postpartum weeks. Each group consisted of 20 milk samples. [A] Scatter plot shows spread of HCMV viral loads at each postpartum time-point. Thick horizontal lines indicate the geometric means, with 95% confidence intervals indicated by the error bars. Sensitivity cut-off is indicated by the lower dotted line. From W4, HCMV DNA loads in milk from HIV-positive women were raised compared to HIV-negative women. Significant differences in geometric means between HIV-positive and negative mothers were at W4 and W16, **p=.026 and ***p<.001 respectively. [B] Shows trend of geometric means over the 16-week time course. Error bars indicate 95% confidence intervals and the assay sensitivity cut-off is indicated by the lower dotted line.

Rate of HCMV viral load change

To infer the average rates of change in HCMV viral loads in the two groups of mothers, three key time-points were considered: D3 (earliest detection of HCMV shedding), W4 (peak HCMV viral loads), and W16 (latest time-point), and viral load kinetics taken to be in steady state, as depicted in figure 4.6. For HIV-positive mothers the gMean HCMV viral load increased at an average rate of 0.4 $\log_{10}$ copies/ml per week from baseline (D3) to peak (W4), and then declined at -0.1 $\log_{10}$ copies/ml per week from W4 to W16. In HIV-negative mothers, gMean HCMV viral load increased at a lower rate of 0.2 $\log_{10}$ copies/ml per week from baseline (D3) to peak (W4), and then, similar to HIV-positive mothers, declined at -0.1 $\log_{10}$ copies/ml per week from W4 to W16. Notably, the rate of HCMV viral load increase from D3 to W4 was two times higher in HIV-positive mothers, who correspondingly also had significantly higher HCMV loads at W4 (figure 4.5). In contrast, the overall rate of HCMV viral load decline from peak levels at W4 was similar in the two groups at -0.1 $\log_{10}$ copies/ml per week. By W16 the HIV-negative group had gMean lower than the D3 level (figure 4.5). Given the higher peak loads at W4 combined with the same rate of decline as in HIV negative women, the HIV-positive group would take at least 26 weeks (6 months) to lower below detection limit, therefore exposing their infants to higher loads for significantly longer duration.

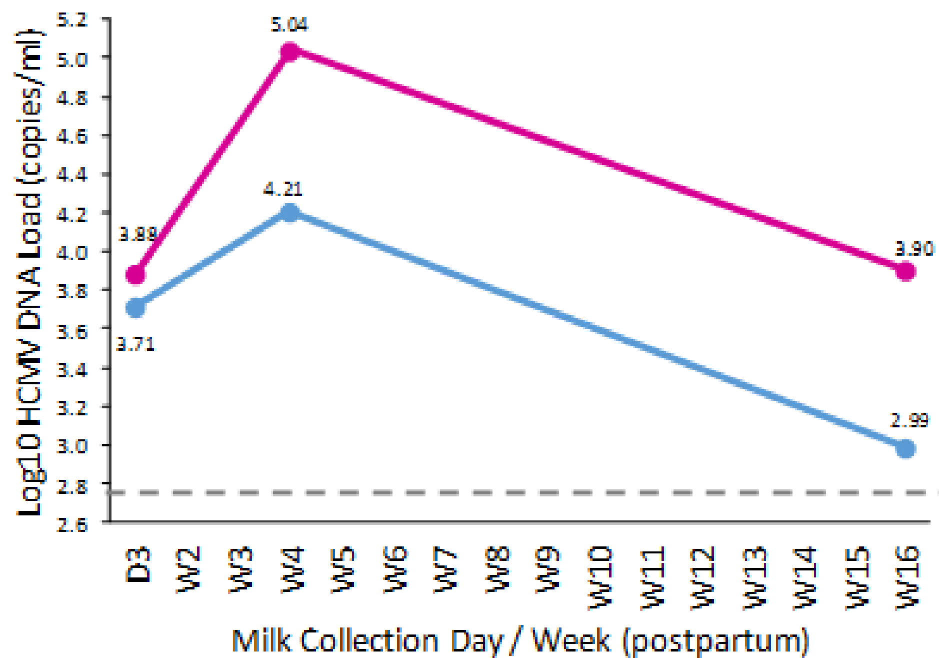


Figure 4.6. Rate of HCMV viral load change. Geometric mean HCMV viral loads at three critical time-points: D3 (the earliest detection of HCMV shedding), W4 (peak HCMV viral loads) and W16 (latest time-point) are compared in HIV-positive (pink line) versus HIV-negative (black line) mothers. The rate of HCMV viral load increase from D3 to W4 was steeper in HIV-positive women, but rate of decrease from peak W4 levels to W16 was comparable in the two groups of women. At D3, gMeans were comparable, but at both W4 and W16 were significantly higher in the HIV-positive group. The qPCR assay sensitivity cut-off is indicated by the lower horizontal dashed line.

HCMV Viral Loads in W16 breast milk

The longitudinal analyses presented in the preceding sections were from available samples for 40 women. However, at W16 samples were collected for the entire cohort. There were 205 samples at W16 with sufficient remaining volume for further analyses. Thus overall at W16, viral loads were measured in 205 samples: 92 (44.9%) HIV-positive and 113 (55.1%) HIV-negative. Of the HIV-positive mothers, 88.0% (81/92) had detectable HCMV DNA, compared to 59.3% (67/113) among the HIV-negative, p<.001. This was consistent with the earlier observation in the kinetics subset. Similar to the kinetics subset, gMean viral load in this larger W16 sample was significantly higher in the HIV-positive group, 3.9 log_{10} copies/ml (95% CI 3.74 – 4.13) compared to the HIV-negatives, 3.2 log_{10} copies/ml (95% CI 3.04 – 3.33), p<.001 (figure 4.7).

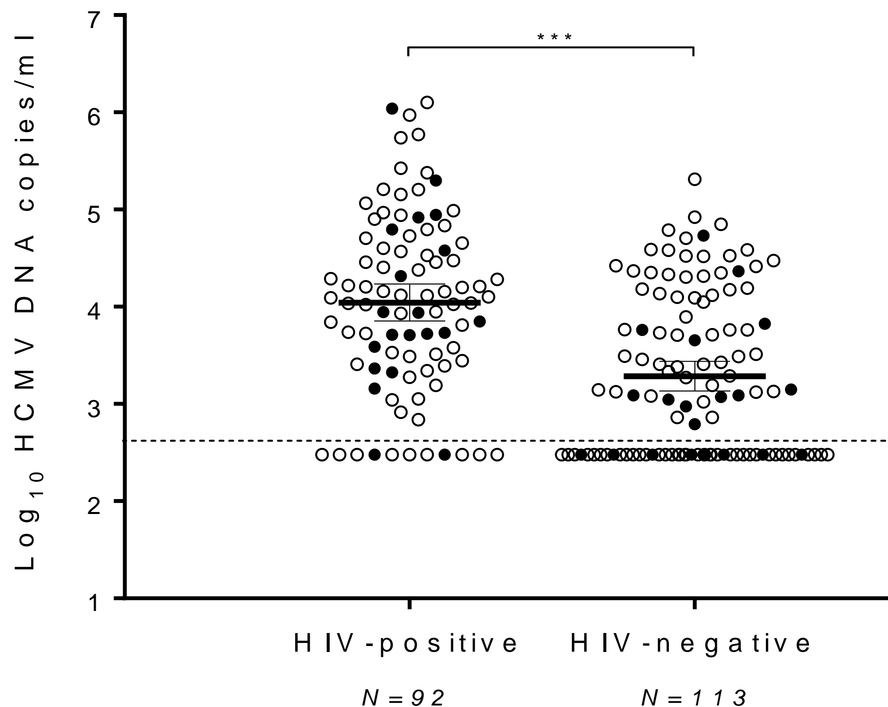


Figure 4.7. HCMV viral Loads in week 16 breast milk: HIV-positive vs. HIV-negative mothers. Scatter plot showing HCMV viral loads in 205 available breast milk samples at W16 (92 HIV-positive and 113 HIV-negative). A higher proportion of breast milk from the HIV-positive group remained with detectable HCMV (88.0% versus 59.3%, ***p<.001), and HIV-positive women also had nearly one log higher gMean HCMV DNA load than their HIV-negative counterparts, 3.93 $\log_{10}$ copies/ml versus 3.19 $\log_{10}$ copies/ml, p<.001. These comparisons in W16 values were similar to observations in samples analysed in the kinetics subset (black circles). Adapted from Musonda and colleagues (Musonda et al., 2016).

4.5 HCMV gO AND gN GENOTYPES IN BREAST MILK

Having established that there are significantly different kinetics of HCMV secretion in HIV-positive versus HIV-negative women, the next step was to examine whether there is any strain restriction in the breast milk compartment or with HIV exposure among the Sub-Saharan African mothers under study. At the time of studies in this thesis there were no data on HCMV strains or isolates from breast milk. To explore the question of HCMV strain restriction in breast milk, the hypervariable UL74 gene, encoding gO, was targeted for sequencing. Additionally, UL73, which encodes gN, was also investigated. Previous studies from our laboratory showed that HCMV gO and gN genotypes form linkage groups and this was further analysed here with potential to validate the gO genotyping within these linkage groups. gO and gN are located adjacent to each other at the centre

of the linear HCMV genome, and encode glycoproteins that have a critical role in virus entry, spread and tropism. Because they are capable of eliciting potent neutralising antibodies, they have recently been targets for HCMV vaccine design studies (see Introduction, section 1.9) (Paterson et al., 2002, Mach et al., 2005, Mach et al., 2000, Dunn et al., 2003, Hobom et al., 2000, Kabanova et al., 2016, Ciferri et al., 2015a, Lemmermann et al., 2015, Ciferri et al., 2015b, Chiuppesi et al., 2015, Wussow et al., 2014). Previous studies in our laboratory have demonstrated that the two genes each form eight genotypes, and unlike some other hypervariable HCMV genes, gO and gN form linkage groups (Bates et al., 2008, Mattick et al., 2004). These are represented by the following full-genome HCMV strains: gO1a/gN1 (AD169, FJ527563.1), gO1b/gN3a (TR, KF021605.1), gO1c/gN4c (Toledo, GU937742.1), gO2a/gN3b (HAN38, GQ396662.1), gO2b/gN2 (AF1, GU179291.1), gO3/gN4a (HAN16, JX512204.1), gO4/gN4b (Towne, FJ616285.1), and gO5/gN4d (Merlin, AY446894.2). In this section, results are presented of HCMV gO and gN sequencing in milk with sufficient remaining DNA sample for analysis.

HCMV gO genotypes in breast milk

HCMV gO analyses were performed for 43 available samples, of which 27 (64.3%) were from HIV-positive and 16 (35.7%) from HIV-negative mothers. Sequence was derived using Sanger methods on PCR amplification products from the variable HCMV gO N-terminal region as described (Mattick, 2004; Bates, 2008) (see Methods section 2.3.2). This amplified the predominant strain in a 741bp N-terminal fragment, as shown in figure 4.8. Sequence analysis showed all eight gO genotypes detectable in breast milk (figure 4.9, table 4.1, and figure 4.10). Overall, the three most prevalent genotypes were gO1a, gO1b, and gO5. With our sample size, gO2a was only detected among HIV-infected mothers, while gO1c and gO3 were detected only among HIV-uninfected mothers (table 4.1, and figure 4.10. Seven samples were paired at W4 and W16, and so these were examined to assess genotype stability at the two time-points (table 4.2). Of these, three pairs – all HIV-positive – had different gO genotypes detected at the two time points, suggesting reinfection and/or infection with mixed strains. The three samples were among those with high HCMV DNA loads (figure 4.3).

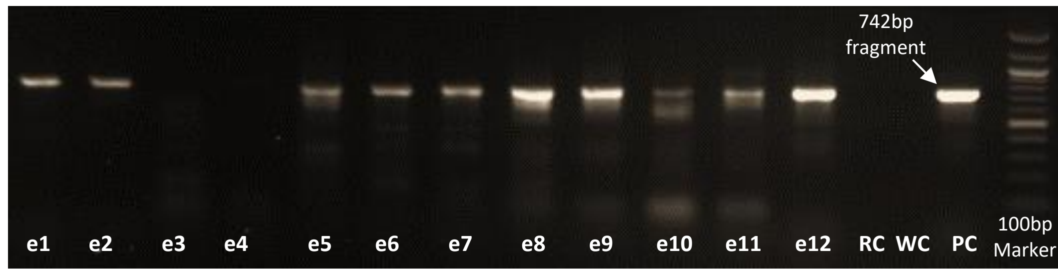


Figure 4.8. PCR amplification of HCMV gO: breast milk. Picture shows PCR amplification of the 741bp HCMV gO gene fragment in DNA extracted from breast milk (here labelled e1–e12). PCR products were loaded in 1% agarose gels and gel electrophoresis was performed in 1X TBE buffer. RC, Reagent (no-template) Control; WC, Water Control; PC, Positive Control (HCMV strain AD169).

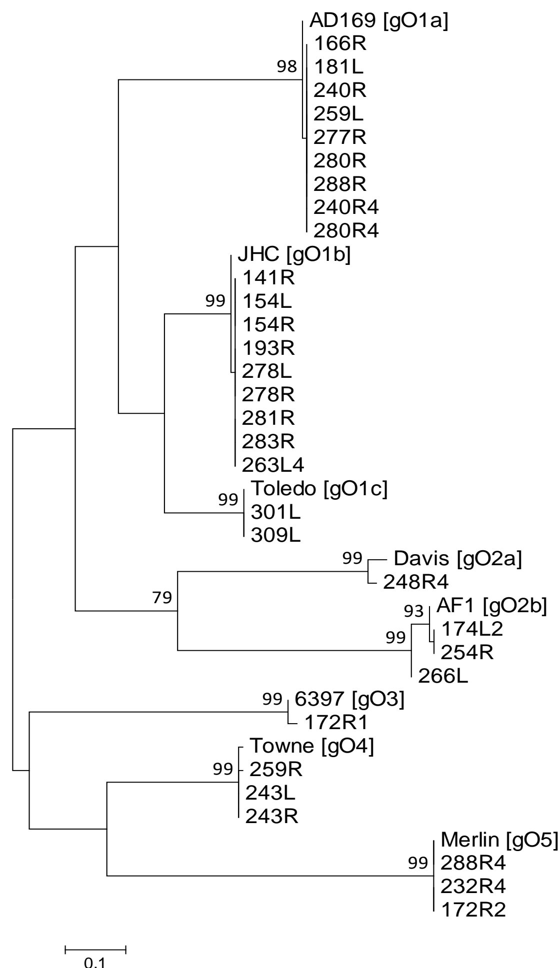
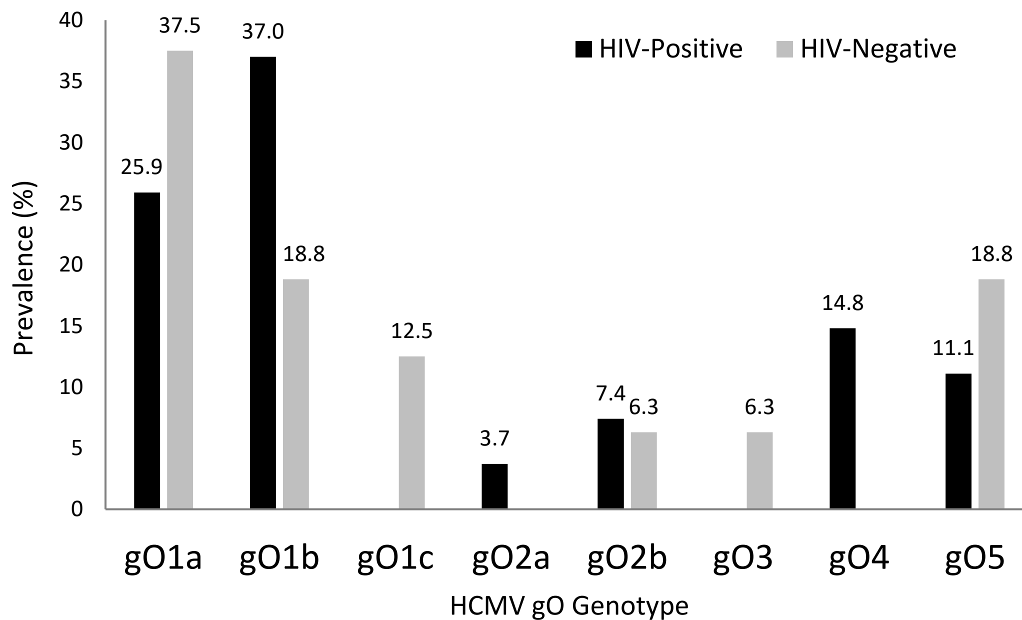


Figure 4.9. Phylogenetic analysis of gO genotypes in breast milk. Representatives of genotype groups defined here were analysed in comparisons to reference strains described previously (Bates et al., 2008, Mattick et al., 2004). Multiple alignments were performed using CLUSTAL in MEGA 6.06 (Tamura et al., 2013), followed by phylogenetic constructions inferred using the Maximum Likelihood method based on the JTT matrix-based model. Bootstrapping values are indicated, supporting major nodes. Figure adapted from Musonda and co-workers (Musonda et al., 2016).

[A]

gO Genotype	HIV-Negative		HIV-Positive		Total	
	n	*%*	*n*	*%*	*n*	*%*
1a	6	37.5	7	25.9	13	30.2
1b	3	18.8	10	37.0	13	30.2
1c	2	12.5	0	0	2	4.7
2a	0	0	1	3.7	1	2.3
2b	1	6.3	2	7.4	3	7.0
3	1	6.3	0	0	1	2.3
4	0	0	4	14.8	4	9.3
5	3	18.8	3	11.1	6	14.0
All	16	100	27	100.0	43	100.0

[B]



[A] Table 4.1 and [B] Figure 4.10. Prevalence of HCMV gO genotypes in breast milk. Overall, the three most prevalent genotypes were gO1a, gO1b, and gO5. In the HIV-infected group, gO4 was higher than gO5.

S/N	ID	HIV	HCMV gO genotype	
			wk4	wk16
1	278R	Positive	1b	1b
2	243R	Positive	1a	4
3	281R	Positive	5	1b
4	288R	Positive	5	1a
5	181L	Positive	1a	1a
6	240R	Negative	1a	1a
7	280R	Negative	1a	1a

Table 4.2. HCMV gO genotype in paired W4/W16 breast milk samples. Three (highlighted in grey shading) of seven W4/W16 paired samples had different gO genotypes at the two time-points. All three were from HIV-infected mothers.

HCMV DNA Load vs. gO Genotype

For the 43 samples for which gO was sequenced, we next investigated whether any particular genotype was associated with high viral load. Figure 4.11 shows the HCMV DNA load for each sample, and highlights the three most prevalent gO genotypes (gO1a, gO1b, and gO5). Although all genotypes appeared increased in HIV-positive compared to HIV-negative mothers, no particular genotype was associated with higher DNA loads among either the HIV-infected or uninfected mothers.

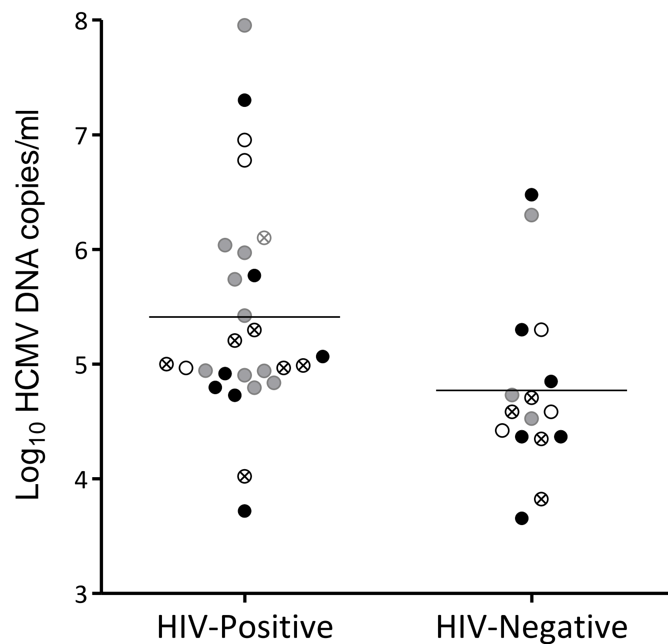


Figure 4.11. Genotype independent increases in HCMV load in breast milk from HIV-positive and negative women. Viral loads per genotype were examined for HIV-positive and negative women. The main three gO genotypes gO1a, gO1b and gO5 were plotted and all the remaining genotypes grouped together. Genotype gO1a (black circle), gO1b (grey circle), gO5 (white circle), other gO genotypes (crossed circle). Geometric means were 5.4 and 4.8 $\log_{10}$ copies/ml for HIV-positive and negative women at both maximum and minimum secreted levels, 4 and 16 weeks *postpartum* respectively. Figure adapted from Musonda and co-workers (Musonda et al., 2016).

HCMV gN PCR amplification in DNA extracted from breast milk

In order to further validate the results from the gO analyses, and to examine linkage groups the samples were analysed with HCMV gN. Initial PCR analyses followed by Sanger sequencing were performed by an MSc student supervised during these studies, (Margaret Njenga, MSc MM, LSHTM 2015). A total of 77 samples were available for HCMV gN analyses. Of these, 42 (54.5%) were from HIV-positive and 35 (45.5%) from HIV-negative mothers. The samples included 22 collected at W4 and 56 collected at W16, with 21 samples paired at both time-points. The 411bp N-terminal fragment targeted in the HCMV gN PCR assay as described (Bates et al., 2008) (see Methods section 2.3.2) was amplified in all the 77 samples, similar to the results shown for gB (figure 4.12).

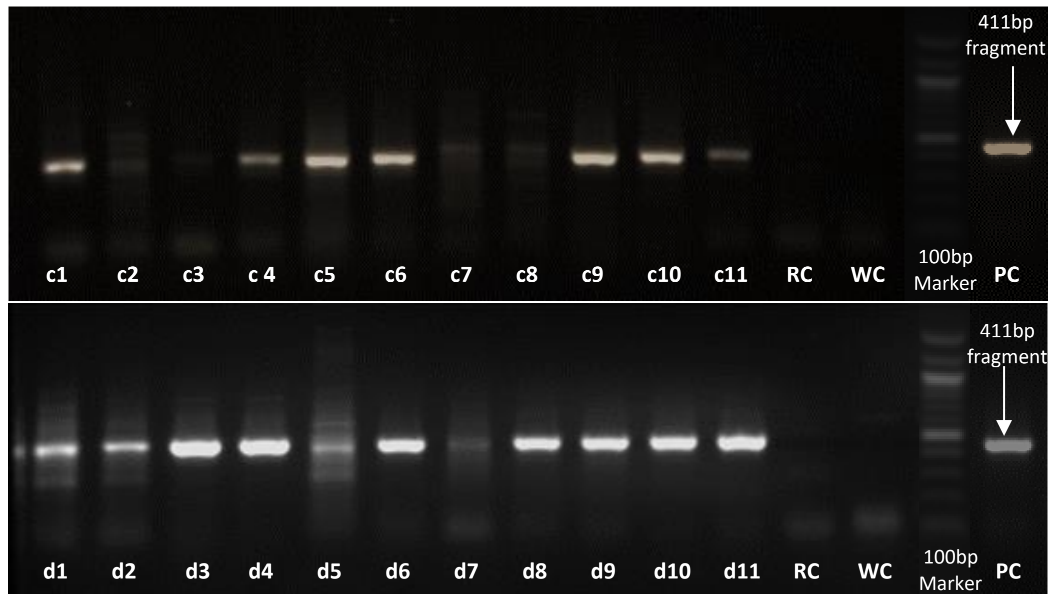


Figure 4.12. PCR amplification of HCMV gN: breast milk. Representation PCR amplification of the 411bp HCMV gN gene fragment in DNA extracted from the breast milk cohort (here labelled c1–c11 and d1–d11). PCR products were loaded in 1% agarose gels and gel electrophoresis performed in 1X TBE buffer. RC, Reagent (no-template) Control; WC, Water Control; PC, Positive Control (HCMV strain AD169).

HCMV gN genotypes recovered from breast milk

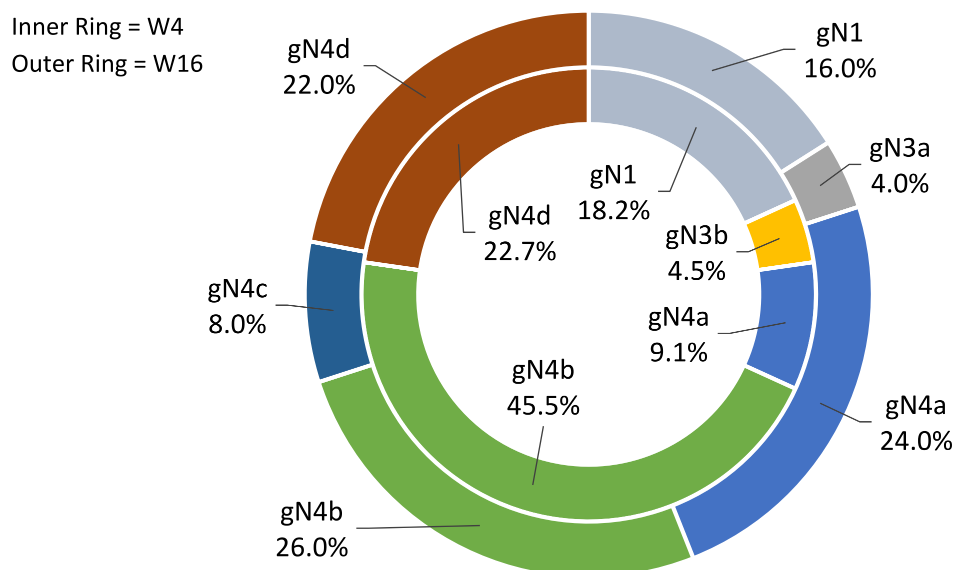
The amplification products were purified from gels using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp.) (see Methods, section 2.5.1) and sequences derived by Sanger methods using the BigDye® Terminator v3.1 (Applied Biosystems) kit (see Methods, section 2.5.2) for 72 of the 77 samples. This included all 42 HIV-positives and 30/35 (85.7%) of the HIV-negatives. The five unresolved HIV negative samples only yielded human DNA sequence and corresponded to the lowest amplification products. Therefore, these could be considered borderline negative reactions. Overall, we were able to identify seven of the eight gN genotypes in breast milk samples in our study, with the exception being gN2, as shown in table 4.3 and figure 4.13. There were seven compared to five gN genotypes identified in the HIV-positive compared to negative samples. Both gN4a and 4d were predominant for HIV-positive and negative samples. Of the 21 W4-W16 paired samples, 12 pairs were HIV-positive and nine HIV-negative. As shown in table 4.5, there were mixed genotypes detected in both the HIV-positive and negative samples, four and two respectively. Overall, there were more genotypes detected in HIV-positive than HIV-negative mothers. Further, more genotypes were

detected at W16 compared to W4, possibly indicating effects of superinfection or reduced immunity. In comparisons at this sample size, only gN4a appeared significantly greater in HIV-positive women compared to negative (table 4.4 and figure 4.14) (p= 03).

[A]

HCMV gN Genotype	W4		W16		
	Freq	%	Freq	%	p value
gN1	4	18.2	8	16.0	1.0
gN2	0	0.0	0	0.0	
gN3a	0	0.0	2	4.0	
gN3b	1	4.5	0	0.0	
gN4a	2	9.1	12	24.0	.04
gN4b	10	45.5	13	26.0	.17
gN4c	0	0.0	4	8.0	
gN4d	5	22.7	11	22.0	1.0
Totals	22	100	50	100	

[B]

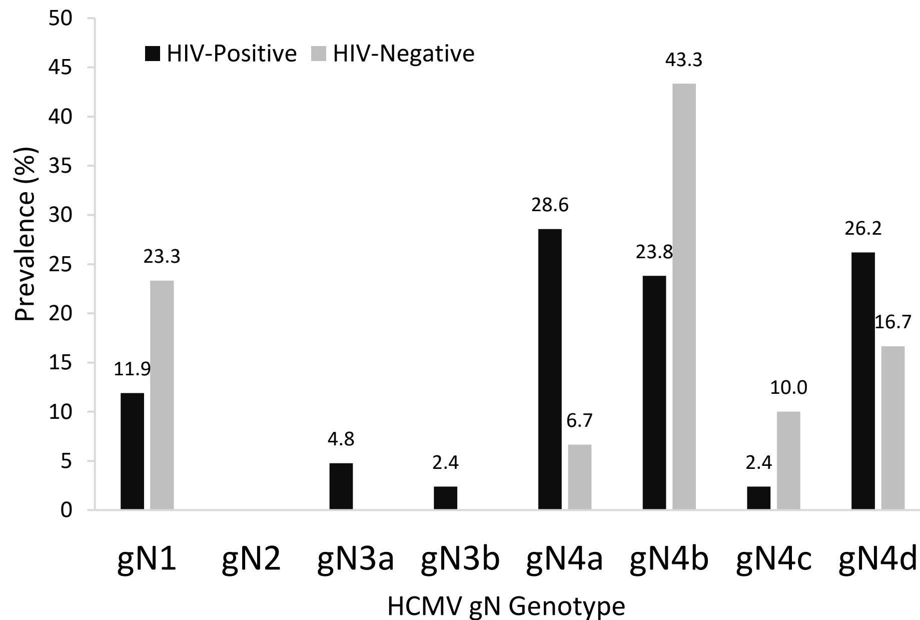


[A] Table 4.3 and [B] Figure 4.13. HCMV gN genotype prevalence in W4 vs. W16 breast milk. Compared to W4, there was loss of gN3b but additional detection of gN3a and gN4c at W16, bringing the overall number of genotypes detected to seven.

[A]

	HIV-POSITIVE		HIV-NEGATIVE		TOTAL		p value
	Freq.	%	Freq.	%	Freq.	%	
gN1	5	11.9	7	23.3	12	16.7	.34
gN2	0	0	0	0	0	0	
gN3a	2	4.8	0	0	2	2.8	
gN3b	1	2.4	0	0	1	1.4	
gN4a	12	28.6	2	6.7	14	19.4	**.03**
gN4b	10	23.8	13	43.3	23	31.9	.07
gN4c	1	2.4	3	10.0	4	5.6	.30
gN4d	11	26.2	5	16.7	16	22.2	.40
TOTAL	42	100	30	100	72	100	

[B]



[A] Table 4.4 and [B] Figure 4.14. HCMV gN genotype prevalence in breast milk of HIV-positive and negative mothers. HIV-positive mothers exhibited greater diversity of HCMV gN genotypes, with seven of the eight genotypes detected. gN4a, gN4b, gN4c being the top three most predominant. Among HIV-negative mothers, only 5 of the eight genotypes were present, with gN4b, gN1, and gN4d being the top three most prevalent genotypes. There were significant differences shown in the prevalence of gN4a (p=.03) but not in the others analysed.

	Sample ID	HIV status	gN Genotype W4	gN Genotype W16
1	181L	Positive	gN1	gN1
2	237L	Positive	gN4d	gN4d
3	243R	Positive	gN4b	gN4b
4	248R	Positive	gN3b	gN4b
5	258R	Positive	gN4a	gN1
6	264R	Positive	gN4b	gN4b
7	277R	Positive	gN4d	gN4c
8	278R	Positive	gN4d	gN4d
9	281R	Positive	gN4d	gN4d
10	283L	Positive	gN4a	gN4a
11	288R	Positive	gN4d	gN4a
12	290L	Positive	gN4b	gN4d
1	185R	Negative	gN4b	gN4b
2	190L	Negative	gN4b	gN4b
3	197L	Negative	gN4b	gN4d
4	228L	Negative	gN4b	gN4d
5	233R	Negative	gN4b	gN4b
6	239R	Negative	gN4b	gN4b
7	240R	Negative	gN1	gN1
8	263L	Negative	gN4b	gN4b
9	280R	Negative	gN1	gN1

Table 4.5. HCMV gN genotypes in paired W4/W16 breast milk samples. Samples with gN genotype changes from W4 to W16 are highlighted in grey. There were four such samples in the HIV-positive group and two among HIV-negatives, p=.58.

4.6 DISCUSSION

In this chapter we examined HCMV shedding in breast milk, comparing both HIV-positive and negative mothers in Zambia, a Sub-Saharan African country. With additional epidemiological analyses which we conducted in the CIGNIS cohort, a childhood cohort from the same community as the breastfeeding cohort examined here, part of these findings were published (Musonda et al., 2016, Gompels et al., 2012, Sanz-Ramos et al., 2013). Our studies show that HCMV infections are widespread in both HIV-positive and negative mothers. However, HIV-positive mothers had markedly higher HCMV levels secreted in their breast milk, and these higher HCMV loads persisted for longer compared to HIV-negative mothers. Initial HCMV levels were equal in the two groups in

early milk at day 3, then were higher in HIV-positive than negative mothers from 2 to 16 weeks postpartum. These findings are in agreement with reports during the course of this thesis study of prolonged (>24 weeks) HCMV loads in breast milk from a cohort of Kenyan HIV-infected mothers (Slyker et al., 2014), although that study did not make comparisons to HIV-uninfected mothers. Our analysis shows peak HCMV DNA levels were at 4 weeks postpartum, which is similar to findings in European and South East Asian surveys. However, there was a difference in shedding duration, with our cohort continuing to shed virus in milk beyond 16 weeks postpartum compared to less than 12 weeks in the European and Asian studies (Hamprecht et al., 2008, Hamprecht et al., 2004b, Hamprecht et al., 2001, Jim et al., 2009, Jim et al., 2004, Yasuda et al., 2003, Hamprecht et al., 2003). The main difference between HIV-positive and negative mothers was the increased rate of HCMV secretion from reactivation at D3 to peak loads at W4. It has recently been shown that CD14+ leukocytes, that are susceptible to HCMV, increase in breast milk around three months postpartum (Maschmann et al., 2015). It is possible that in our cohort there may be more of these susceptible cells in the mammary glands of HIV-infected mothers, and coupled with higher immune dysregulation, this may favour HCMV proliferation in these mothers. Immune dysregulation, can be induced by HIV and HCMV independently, and worsens when both viruses act concurrently. There may also be effects from increased inflammation, given the higher levels of mastitis (including sub-clinical) shown among HIV-positive mothers in earlier analyses of this cohort (Kasonka et al., 2006). In the midst of the high HCMV levels demonstrated in our present study, mastitis would be an aggravating factor for breast milk transmission of HCMV to infants of HIV-infected mothers.

Studies in Europe and Asia have shown HCMV secretion in breast milk for up to 3 months postpartum, with higher HCMV viral loads in milk and prolonged secretion linked to infant transmission (Hamprecht et al., 2003, Jim et al., 2009, Yasuda et al., 2003, van der Strate et al., 2001). In contrast to these regions, we have now shown prolonged HCMV secretion in breast milk beyond 4 months in Zambia. Furthermore, epidemiological analyses performed on the CIGNIS cohort show that breastfeeding for over 6 months among HIV-positive women, or beyond 18 months among HIV-negative women increased risk of infant HCMV infection (table 4.6). As described above, risk of transmission to infants of HIV-positive mothers may be enhanced by immune

impairment, a hallmark of both HIV and HCMV, and amplified by prolonged breastfeeding duration. In the case of HIV-negative mothers, the risk of HCMV transmission to their infants may be related to protracted local reactivation and secretion of HCMV, sustained by the much longer duration of lactation. Similar to other studies (Bate et al., 2010, Boppana and Fowler, 2007, Gompels et al., 2012), we found that low socioeconomic status and education level were associated with HCMV seroprevalence. However, even after controlling for these risk factors, breastfeeding up to 18 months remained significantly associated with HCMV infection (table 4.6).

	Infant HCMV Infection (Serum Antibody)		
Months Breastfeeding	**Antibody Positive N (%)**	**Adjusted OR[a] (95% CI)**	***P* Value**
HIV-Negative Mothers			
<12[b]	25/32 (78.1%)	1	
12–17	128/161 (79.5%)	0.94 (0.35–2.53)	
18+	110/119 (92.4%)	2.69 (0.84–8.59)	.03
HIV-Positive Mothers			
Never	13/26 (50.0%)	1	
<6	31/35 (88.6%)	6.83 (1.69–27.6)	
6+	42/44 (95.5%)	20.37 (3.71–111.7)	<.001

Table 4.6. Effects of maternal HIV and breastfeeding duration on infant HCMV infection. [a] Adjusted for socioeconomic status and maternal education. [b] Only 3 HIV-negative mothers never breastfed and only 6 for <6 months. CI, confidence interval; HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; OR, odds ratio. (Musonda et al., 2016)

Slyker and co-workers showed in Kenya that mothers who transmitted HCMV had nearly one log higher median HCMV loads at 2 weeks postpartum compared to non-transmitters (5.4 log copies/ml versus 4.5 log copies/ml). Further, the probability of infant CMV infection was modelled to be 5.5 log copies/ml at a maternal CD4 count of 1000 cells/mm^3 (Slyker et al., 2014). In our Zambian cohort, at W2 HIV-positive mothers also had a 1.1 log higher median HCMV load compared to their HIV-negative counterparts (5.1 versus 4.0 log copies/ml). Although it is difficult to directly apply the thresholds from the Kenyan study to ours, a basic comparison indicates that HIV-positive mothers would be more likely to be transmitting HCMV at this early postpartum time-point compared to HIV-negative mothers. Between day 3 and week 16, HIV-positive women had 23 measurements above 5.5 log copies/ml compared with only seven for

HIV-negative mothers (table 4.7). However, it is likely that a lower transmission threshold applies and therefore many more mothers would have had readings above the transmission threshold, since the 5.5 log copies/ml cut-off assumes a strong maternal cellular immune compartment (1000 cells/mm^3 CD4 cell count), which is rare in HIV-infected individuals, even with HAART.

	HIV-positive (N=20)		HIV-negative (N=20)	
	Freq	*%*	*Freq*	*%*
D3	0	0	0	0
W2	5	25	2	10
W4	7	35	3	15
W9	5	25	2	10
W12	5	25	0	0
W16	1	5	0	0

Table 4.7. Proportion of breast milk samples with HCMV DNA load above 5.5 $\log_{10}$ copies/ml. HCMV DNA loads above this transmission threshold persisted beyond W9 in HIV-positive mothers.

We also assessed overall rates of change in gMean HCMV loads by considering three key reference points (D3 – the earliest detection of HCMV shedding; W4 – the peak HCMV viral loads; and W16 – the latest available time-point) (figure 4.6), and showed that while both HIV-infected and uninfected mothers had similar baseline levels at D3, the average rate of viral load increase to peak at W4 was twice as much in HIV-infected mothers as in the HIV-uninfected, resulting in higher peak levels at W4 in the HIV-positive group. Interestingly, from W4 to W16, the average rate of decline was the same in the two groups of mothers, with the result that HIV-negative mothers returned to baseline levels much quicker than the HIV-infected mothers. Assuming steady state kinetics, we thus estimated that it would take up to 26 weeks for HIV-positive mothers to achieve DNA loads below the detection limit. This gives a broader opportunity of HCMV exposure to their infants. The first month of life is therefore a crucial window for targeting interventions to limit the DNA load attained at W4, akin to the well-established concept of 'viral set-point' in HIV-1 infection (de Wolf et al., 1997, Geskus et al., 2007). Since an efficacious vaccine is yet to be formulated, drug administration to mothers to lower

HCMV load in breast milk is a possible option. However, no anti-HCMV drug is yet approved for use during lactation. Other measures could include heat treatment or irradiation of milk, which have shown success particularly in milk banks in the setting of very low birthweight and premature babies (Hamprecht et al., 2004b, Christen, 2014, Christen et al., 2013).

Our analysis of HCMV genotypes found all 8 gO genotypes could be detected in breast milk, showing that there was no genotype restriction in the mammary gland. Mixed infections, suggestive of maternal reinfection, were also noted and confirmed by gN genotyping. However, we could not find any particular genotype linking with high HCMV DNA loads. Overall, the main gO genotypes were gO1a, gO1b, and gO5. This is similar to gO genotypes analysed in blood and respiratory compartments of HIV-infected Zambian children, and in various tissues from different global sources (Bates et al., 2008, Mattick et al., 2004). Differences in prevalence of the genotypes were not significant statistically; however, more gO3 was reported in the Bates and Mattick studies (figure 4.15).

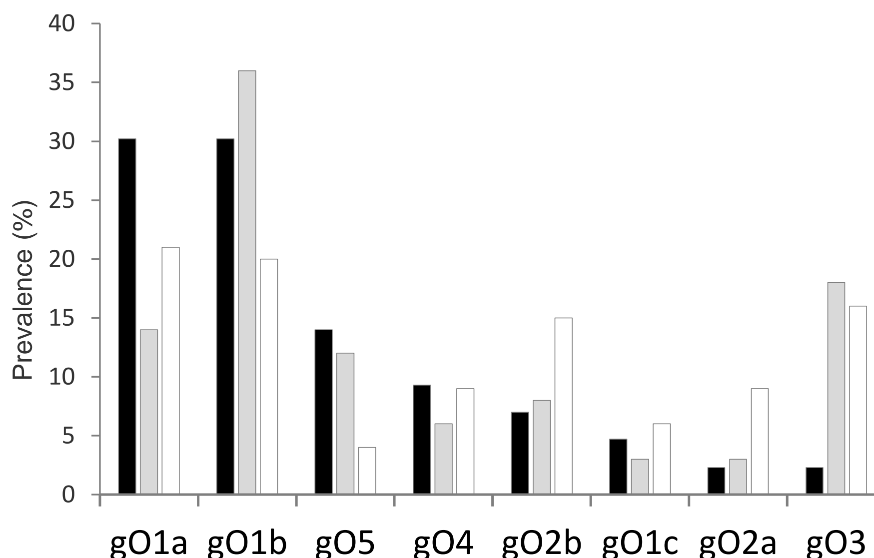


Figure 4.15. Comparison of HCMV gO genotypes prevalence in breast milk and other tissues. The prevalence of the eight gO genotypes in milk was similar to studies of genotypes from different compartments. The most predominant genotypes in the current study (black bars) were 1a and 1b followed by gO5. Grey bars depict gO genotype prevalence from our previous study of HIV-positive blood and lung tissue (Bates et al., 2008) in the same region, and white bars show gO prevalence in various tissues sourced from multiple global regions (Mattick et al., 2004).

Further study of genotype ratios is essential, since the trimeric gH/gL/gO complex and alternate pentameric gH/gL/pUL128-131 complex affect cellular tropism, host transmission, as well as candidate vaccines (Paterson et al., 2002, Mach et al., 2005, Mach et al., 2000, Dunn et al., 2003, Hobom et al., 2000, Kabanova et al., 2016, Ciferri et al., 2015a, Lemmermann et al., 2015, Ciferri et al., 2015b, Chiuppesi et al., 2015, Wussow et al., 2014).

Our study of HCMV in breast milk had three main limitations. First, the HCMV serostatus of the mothers was not known. Nevertheless, our HCMV DNA PCR screen showed that all mothers were HCMV positive. Secondly, maternal plasma CD4 cell counts were not available. The CD4 level can affect HCMV secretion in breast milk, as demonstrated in the cohort from Kenya (Slyker et al., 2014). Thirdly, we did not control for the contribution of other HCMV transmission routes, particularly saliva and urine from other children (Grosjean et al., 2014, Cannon et al., 2014) and congenital HCMV. However, we expect that the general background exposure would have been similar in families of both HIV-infected and uninfected mothers, since most mothers had children under the age of 5 years. Based on other studies and our own investigation of congenital infection (Chapter 3), the contribution from congenital HCMV infection would have been very low (1%.) Strengths of our study are the sizes of the maternal and infant cohorts analysed, and the crucial comparisons made between HIV-positive and negative groups, vividly showing raised levels and duration of HCMV secretion in breast milk, with correspondingly increased risks for infant infection in HIV-positive compared with HIV-negative mothers.

CHAPTER 5:

GENOMIC ANALYSIS OF HCMV DIVERSITY IN BREAST MILK

5.1 INTRODUCTION

In the previous chapter, hypervariable gO and gN genotypes were identified together with their use to identify mixed infections with different strains of HCMV in breast milk. This could be used as a marker of burden of infection in addition to the virus loads evaluated in the previous chapters. In order to investigate this further and understand the relationship to the entire HCMV genetic complement, in this chapter new methods of next generation deep sequencing were used to analyse these infections. Having established that in breast milk there is a higher HCMV viral load associated with maternal HIV-infection, there was preliminary evidence that some HCMV infections were composed of mixed genotypes, indicating repeated infections. Therefore, the high viral load could be compounded by a high burden of infection from repeated or mixed infections. The Sanger sequencing used previously has a low sensitivity to detect mixed infections as basically the dominant consensus sequence is identified and mixed infections can be detected either by multiple PCR reactions or by cloning. To investigate this further, new methodologies were required to sensitively detect and quantitate mixed infections. The hypervariable gO and gN genes served as key determinants to investigate this as not only could they be used to identify mixed infection, they themselves could also affect the infection since they are mediators of infection. Therefore, in this chapter analyses were undertaken using NGS deep sequencing to shed light on how the genotype proportions relate to viral load and HIV status. Deep sequencing resolves the target region typically thousands of times, enabling sensitivities of mixtures comprising as little as 1% of the total (Margeridon-Thermet et al., 2009, Mitsuya et al., 2008, Radford et al., 2012, Varghese et al., 2009). Where detected previously mixed-genotype infections are markers for high disease burden, severe infection, or persistent HCMV infection (Humar et al., 2003, Coaquette et al., 2004, Pang et al., 2008, Sarcinella et al., 2002, Sowmya and Madhavan, 2009, Puchhammer-Stockl et al., 2006). Various research groups have demonstrated occurrence of mixed-genotype infections from study of selected HCMV genes including gB (UL55), gN (UL73), gO (UL74), gH UL75), gL (UL115), UL139, and UL146 (Puchhammer-Stockl et al., 2006, Sarcinella et

al., 2002, Coaquette et al., 2004, Humar et al., 2003, Sowmya and Madhavan, 2009, Pang et al., 2008, Peek et al., 1998, Gorzer et al., 2008, Bradley et al., 2008, Hassan-Walker et al., 2004, Stanton et al., 2005, Bates et al., 2008, Mattick et al., 2004). These studies suggest that mixed HCMV genotypes exist within an infected individual, and following transmission, a dominant genotype may be selected in the new host. In addition, there is also some initial evidence for *de novo* mutations that may arise in each host, giving rise to a unique dominant genotype (Renzette et al., 2011, Renzette et al., 2013). However, without sufficient deep sequencing analyses, the origins of these mutations are not clear. With previous technology, in general, it follows that this dominant genotype is the one detectable by Sanger sequencing methods, leaving co-infecting genotypes undetected.

Next Generation Sequencing (NGS)

The original Sanger dideoxynucleotide chain termination sequencing method (Sanger and Coulson, 1975) revolutionised genetic studies. HCMV was one of the first complete genome sequences determined using the technology (Chee et al., 1990, Bankier et al., 1991). This method was eventually automated, incorporating fluorescent markers and enabled the first finished-grade human genome sequence to be compiled (International Human Genome Sequencing Consortium, 2004). However, the original method had limitations including laborious bacterial/bacteriophage cloning (which may introduce bias) and electrophoretic separation of chain-terminated fragments, low sample throughput, and high cost. New technologies include 'Next Generation Sequencing' (NGS), which has transformed nucleotide sequencing by making it more accurate, reliable, and cheaper (and therefore more accessible), while reducing amounts of template DNA required. Because NGS allows combination of chain synthesis steps with signal detection, reactions are several fold quicker than Sanger methods. Further, the massively parallel sequencing setup used in many NGS platforms allows hundreds of gigabases of DNA to be read in a single run on a single chip, with capacity increased by as much as 1,000 times (Kircher and Kelso, 2010). However, as NGS technologies have become more advanced, challenges include storing the large quantities of data generated, the relatively high error rates, estimated at around 0.1–15% in different systems, and shorter read lengths of around 50-700 bp (Liu et al., 2012). This could compound analyses of, for example, mixed infection or emerging drug resistance, and

highlights the need for careful analysis of results, especially for variant discovery and clinical applications. Among the commercially available technologies, that developed by Illumina is currently the most widely available (Goodwin et al., 2016, Investec, 2015), and involves Library preparation, sequencing, and data analysis.

To further address the questions of infectious burden which was limited by Sanger sequencing sensitivity as outlined above, this study identified 'Molecular tags', a type of barcode technology combined with computer-based customised Scripts. This facilitated interrogation and quantification of deeply sequenced genomes using next generation sequencing methods. This devised technique enabled detection and quantification of multiple infections using identification of HCMV gO and gN genotype variants as biomarkers in a sample. The method was validated on sequences publicly available in GenBank, and then applied to analyse Illumina NGS reads derived from DNA extracted from breast milk from a cohort of both HIV-infected and uninfected women who had been studied for post-partum health in Zambia; this cohort was also analysed in the previous chapter and in a publication arising from this thesis (Collin et al., 2006, Kasonka et al., 2006, Musonda et al., 2016). This enabled analyses of burden of infection over the first 16 weeks postpartum. The second part of this analyses identified samples with a dominant or single genotype. This would then allow unambiguous compilation of a complete genome, without the risk of an artefactual 'mixed genome' derived from genome compilation from a source with mixed infections. Samples with verified single or dominant (>90%) genotypes were then evaluated for assembly of NGS FASTQ reads into complete HCMV genomes. Here we characterise ZMB240, a complete HCMV genome similar to the reference strain Merlin, but assembled from DNA directly extracted from breast milk, without bias introduced from isolation or tissue culture. This is, to our knowledge, the first complete HCMV genome from Africa.

5.2 IDENTIFICATION OF 'MOLECULAR TAGS' AND DEVELOPMENT OF *PERL* SCRIPTS FOR DETECTION AND ENUMERATION OF HCMV gO AND gN GENOTYPES

5.2.1 Identification of HCMV gO and gN 'Molecular Tags'

The first step in addressing the question of mixed-genotype infections was to develop tools to enable both the detection and enumeration of genotypes of a given hypervariable HCMV gene present in a sample. For this thesis, the focus is on the hypervariable HCMV gO. However, since gO forms linkage groups with gN, and both are hypervariable loci (Mattick et al., 2004, Bates et al., 2008), use of both gO and gN allowed systematic testing of this linkage, some verification of our methods, and identification of rare recombination at this locus. To start with, GenBank release 211 was systematically searched for all available complete HCMV genome sequences, and 163 complete genomes were identified and downloaded. Following extraction of the gO and gN nucleotide sequences from each of these genomes, MEGA 6.06 (Tamura et al., 2013) was used for analysis, including amino acid translation and alignment using the in-built ClustalW tool, with default settings. The aligned amino acid sequences and corresponding nucleotide sequence alignments (*.meg* files) were saved and exported to Microsoft Excel format for convenient visualisation. Phylogenetic trees were also generated based on aligned amino acid sequences, as shown in figures 5.1 and 5.4. Following systematic analyses of the amino acid alignments, short stretches of amino acids unique to each genotype at particular loci were identified as highlighted in figures 5.2 and 5.5, and then the corresponding back-translated nucleotide region identified in the nucleotide alignment (figures 5.3 and 5.6). At a nucleotide level, these 'molecular tags' were subsequently refined, taking into account the redundancy of the standard genetic code, to obtain sequences of 12-14 bases as optimum length. The process of identifying and fine-tuning the genotype-specific molecular tags was also guided by DNA polymorphism plots generated in DnaSP (Librado and Rozas, 2009) from the alignment of the 163 gO and gN nucleotide sequences (figure 5.7). For gO, three sets of tags were identified at three different loci: two within the hypervariable N-terminus and one within the C-terminus. Table 5.1 details the sequences and coordinates of each identified molecular tag relative to the reference strain for each gO and gN genotype. For HCMV gN, initial molecular tag determinations were done by Margaret Njenga

(MSc.MM, LSHTM 2015) using an earlier GenBank dataset with 136 complete HCMV genomes. This was extended to include the set of 163 as analysed for gO as available in GenBank release 211 (National Center for Biotechnology Information (NCBI), 2015).

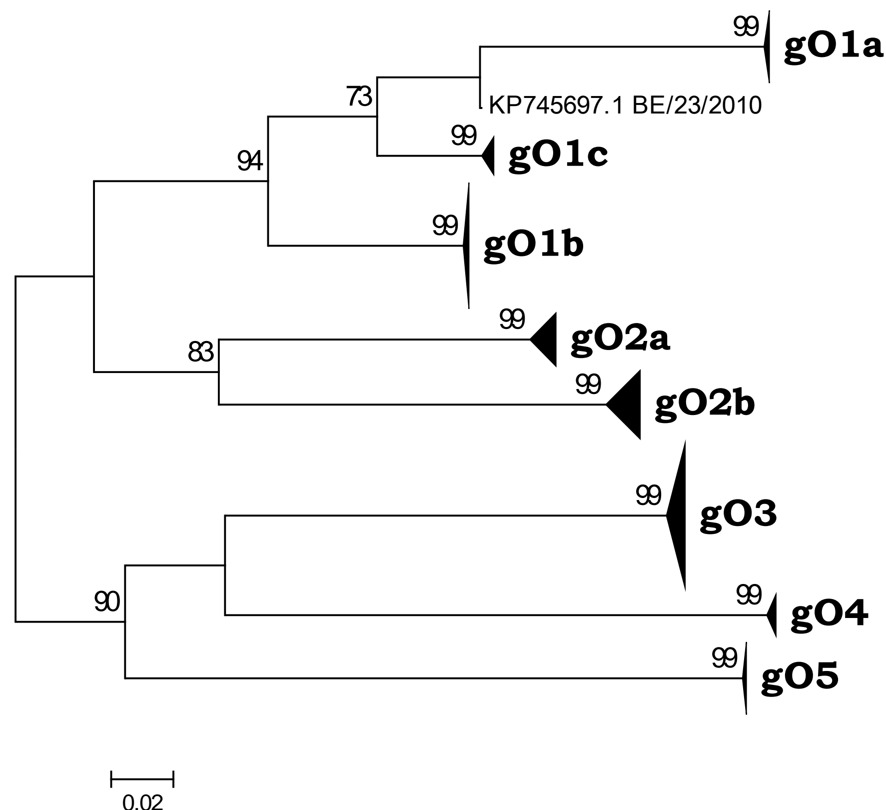


Figure 5.1. Molecular phylogenetic relationship of the eight HCMV gO genotypes. This unrooted consensus tree was inferred from 1000 bootstrap replicates to represent the evolutionary history of 163 full-length HCMV gO amino acid sequences derived from complete HCMV genomes available in GenBank database release 211 (National Center for Biotechnology Information (NCBI), 2015). The evolutionary history was inferred by the Maximum Likelihood method based on the JTT matrix-based model and a discrete Gamma distribution with five categories (+*G*, parameter = 0.6555). Positions with less than 95% site coverage were eliminated, leaving 434 amino acid positions in the final dataset. Due to space constraints, the branches are compressed to highlight the eight gO genotypes (Bates et al., 2008, Mattick et al., 2004). All sequences grouped to one of the eight genotypes, except for BE/23.2010 (KP745697.1) discussed in the text. Analyses were conducted in MEGA 6.06 (Tamura et al., 2013).

Accession No. & Strain ID	Genotype	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
FJ527563.1 AD169 gO1a.	gO1a	I	L	R	Q	L	E	T	T	I	-	-	S	T	K	Y	-	-	-	-	N
KP745640.1 BE/22/2010.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745641.1 BE/31/2011.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KJ872542.1 PAV21.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
GU179290.1 U11.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745699.1 BE/1/2012.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KC519319.1 BE/9/2010.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745657.1 BE/13/2011.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745653.1 BE/22/2011.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KC519323.1 BE/27/2010.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KT634296.1 UKNEQAS2.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KJ361966.1 PAV12.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745674.1 BE/33/2011.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KJ361958.1 HAN40.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KJ361959.1 PAV1.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KJ361950.1 HAN11.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KJ872541.1 PAV20.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745722.1 BE/40/2011.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745709.1 BE/48/2011.	gO1a	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	-	-	-	-	.
KP745697.1 BE/23/2010.	gO1a/1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
KP745652.1 BE/2/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745649.1 BE/10/2012.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745634.1 BE/32/2010.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KJ361967.1 PAV23.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745635.1 BE/5/2012.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KJ361962.1 PAV6.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745639.1 BE/10/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745701.1 BE/6/2010.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745703.1 BE/26/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745696.1 BE/27/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745698.1 BE/20/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745704.1 BE/32/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745725.1 BE/49/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745726.1 BE/30/2010.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745713.1 BE/35/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745715.1 BE/44/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745691.1 CZ/1/2013.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745663.1 BE/5/2010.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745668.1 BE/18/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745658.1 BE/14/2012.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745660.1 BE/6/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745673.1 BE/42/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745687.1 BE/36/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745689.1 BE/17/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745675.1 BE/23/2011.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KP745676.1 BE/28/2010.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KF021605.1 TR gO1b.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KJ361947.1 2CEN5.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KJ361948.1 2CEN15.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
JX512207.1 HAN28.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KC519321.1 BE/11/2010.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
HQ380895.1 JHC.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KJ361954.1 HAN32.	gO1b	.	.	K	.	.	.	K	D	.	-	-	Y	.	.	.	F	-	-	-	.
KJ872540.1 PAV18.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
KJ361955.1 HAN33.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
JX512201.1 HAN3.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
KF297339.1 TB40/E Lisa.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
GU179289.1 VR1814.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
GU937742.1 Toledo gO1c.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
KP745710.1 BE/2/2012.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
GQ396663.1 HAN20.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
KP745721.1 BE/14/2010.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
KP745685.1 CZ/3/2012.	gO1c	.	.	K	.	.	.	R	E	P	-	-	Y	.	.	.	F	-	-	-	.
KP745724.1 BE/4/2012.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KP745646.1 BE/8/2012.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KP745638.1 BE/15/2010.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
GQ221973.1 HAN13.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KP745643.1 CZ/2/2012.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q

KC519322.1 BE/21/2010.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
JX512202.1 HAN8.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KP745677.1 BE/1/2010.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
JX512199.1 HAN1.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
JX512198.1 Davis.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KP745712.1 BE/19/2010.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KJ361968.1 PAV24.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
GQ396662.1 HAN38 gO2a.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KP745683.1 BE/12/2011.	gO2a	.	.	.	.	.	.	P	I	P	-	-	Y	I	.	.	P	-	-	-	Q
KP745656.1 BE/2/2013.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KJ361956.1 HAN36.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
JX512208.1 HAN31.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745667.1 BE/5/2011.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745686.1 BE/39/2011.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745716.1 BE/16/2010.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745654.1 BE/19/2011.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745705.1 BE/38/2011.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
GU179288.1 U8.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KJ361951.1 HAN21.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745678.1 BE/25/2010.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KJ361965.1 PAV11.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
JX512200.1 HAN2.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
JX512203.1 HAN12.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745672.1 BE/29/2011.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745679.1 BE/24/2010.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
GU179291.1 AF1 gO2b.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KJ361946.1 2CEN2.	gO2b	.	.	K	.	.	Q	S	D	.	-	-	Y	.	.	.	P	-	-	-	Q
KP745680.1 BE/11/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745719.1 BE/26/2010.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745681.1 BE/43/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745702.1 BE/21/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745700.1 BE/4/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745714.1 BE/29/2010.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745706.1 BE/41/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745711.1 BE/24/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745708.1 BE/8/2010.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745707.1 BE/13/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
JX512197.1 6397.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745688.1 BE/12/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745684.1 BE/11/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745692.1 BE/3/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745718.1 CZ/1/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745694.1 BE/12/2010.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745693.1 BE/15/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745651.1 BE/9/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745650.1 BE/1/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KJ361952.1 HAN27.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KJ361964.1 PAV8.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745655.1 BE/3/2010.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745633.1 BE/45/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KJ361949.1 2CEN30.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745636.1 BE/7/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745645.1 BE/13/2010.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745642.1 CZ/1/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745659.1 BE/3/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745666.1 BE/7/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745665.1 BE/16/2012.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
JX512204.1 HAN16 gO3.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KJ361957.1 HAN39.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745670.1 BE/30/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745723.1 BE/37/2011.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745661.1 BE/33/2010.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KJ361961.1 PAV5.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KJ361960.1 PAV4.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KP745664.1 CZ/2/2013.	gO3	L	.	K	.	I	G	A	S	Q	-	D	Y	Y	.	F	F	T	-	-	I
KJ361969.1 PAV25.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F
GQ466044.1 3301.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F
KJ426589.1 HAN.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F
FJ616285.1 Towne gO4.	gO4	.	.	Q	E	.	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F
KP745690.1 BE/34/2011.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F

KC519320.1 BE/10/2010.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F
KP745720.1 BE/15/2011.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F
JX512206.1 HAN22.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	L	T	-	-	F
KP745648.1 BE/8/2011.	gO4	.	.	Q	E	.	A	S	K	T	G	D	Y	Y	.	F	F	T	-	-	F
KP745717.1 BE/2/2010.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	L	T	-	-	F
KP745637.1 BE/9/2011.	gO4	.	.	Q	E	I	A	S	K	T	G	D	Y	Y	.	F	L	T	-	-	F
GQ221974.1 3157.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KT959235.1 DB.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745728.1 BE/4/2010.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
AY446894.2 Merlin gO5.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745727.1 BE/17/2010.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KJ872539.1 PAV16.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745644.1 BE/31/2010.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745647.1 BE/18/2010.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KJ361971.1 UKNEQAS1.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KJ361963.1 PAV7.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KJ361953.1 HAN30.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KJ361970.1 PAV26.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745662.1 BE/20/2010.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745682.1 BE/46/2011.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745695.1 BE/6/2012.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
GQ221975.1 JP.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
JX512205.1 HAN19.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745669.1 BE/28/2011.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y
KP745671.1 BE/14/2011.	gO5	.	.	K	R	.	M	S	.	S	-	-	.	D	G	.	R	F	L	M	Y

Figure 5.2. Amino acid sequence alignment of HCMV gO from 163 complete HCMV genomes. The eight gO genotypes are colour-coded, with the region showing unique amino acids corresponding to each genotype highlighted. The peculiar case of strain BE/23/2010 (KP745697.1) is discussed in section 5.2.3 'Unique Cases'.

Accession No. & Strain ID	Genotype	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228
FJ527563.1 AD169 gO1a.	gO1a	A	G	A	A	A	C	G	A	C	T	A	T	T	-	-	-	-	-	-	T	C	T
KP745640.1 BE/22/2010.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KP745641.1 BE/31/2011.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KJ872542.1 PAV21.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
GU179290.1 U11.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KP745699.1 BE/1/2012.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KC519319.1 BE/9/2010.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KP745657.1 BE/13/2011.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KP745653.1 BE/22/2011.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KC519323.1 BE/27/2010.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KT634296.1 UKNEQAS2.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KJ361966.1 PAV12.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KP745674.1 BE/33/2011.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KJ361958.1 HAN40.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KJ361959.1 PAV1.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KJ361950.1 HAN11.	gO1a	.	.	.	.	.	.	.	.	.	A	.	.	.	-	-	-	-	-	-	.	.	.
KJ872541.1 PAV20.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KP745722.1 BE/40/2011.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KP745709.1 BE/48/2011.	gO1a	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	.	.	.
KP745697.1 BE/23/2010.	gO1a/1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
KP745652.1 BE/2/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745649.1 BE/10/2012.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745634.1 BE/32/2010.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KJ361967.1 PAV23.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745635.1 BE/5/2012.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KJ361962.1 PAV6.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745639.1 BE/10/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745701.1 BE/6/2010.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745703.1 BE/26/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745696.1 BE/27/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745698.1 BE/20/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745704.1 BE/32/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745725.1 BE/49/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745726.1 BE/30/2010.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745713.1 BE/35/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745715.1 BE/44/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745691.1 CZ/1/2013.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745663.1 BE/5/2010.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745668.1 BE/18/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745658.1 BE/14/2012.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745660.1 BE/6/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745673.1 BE/42/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745687.1 BE/36/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745689.1 BE/17/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745675.1 BE/23/2011.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KP745676.1 BE/28/2010.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KF021605.1 TR gO1b.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KJ361947.1 2CEN5.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KJ361948.1 2CEN15.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
JX512207.1 HAN28.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KC519321.1 BE/11/2010.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
HQ380895.1 JHC.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KJ361954.1 HAN32.	gO1b	.	.	.	.	.	A	.	G	A	.	.	.	C	-	-	-	-	-	-	.	A	.
KJ872540.1 PAV18.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
KJ361955.1 HAN33.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
JX512201.1 HAN3.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
KF297339.1 TB40/E Lisa.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
GU179289.1 VR1814.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
GU937742.1 Toledo gO1c.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
KP745710.1 BE/2/2012.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
GQ396663.1 HAN20.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
KP745721.1 BE/14/2010.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
KP745685.1 CZ/3/2012.	gO1c	.	.	.	.	.	G	.	G	A	A	C	C	.	-	-	-	-	-	-	.	A	.
KP745724.1 BE/4/2012.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KP745646.1 BE/8/2012.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KP745638.1 BE/15/2010.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
GQ221973.1 HAN13.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KP745643.1 CZ/2/2012.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C

KC519322.1 BE/21/2010.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
JX512202.1 HAN8.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KP745677.1 BE/1/2010.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
JX512199.1 HAN1.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
JX512198.1 Davis.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KP745712.1 BE/19/2010.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KJ361968.1 PAV24.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
GQ396662.1 HAN38 gO2a.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KP745683.1 BE/12/2011.	gO2a	.	.	.	.	C	.	T	.	T	.	C	C	.	-	-	-	-	-	-	.	A	C
KP745656.1 BE/2/2013.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KJ361956.1 HAN36.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
JX512208.1 HAN31.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745667.1 BE/5/2011.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745686.1 BE/39/2011.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745716.1 BE/16/2010.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745654.1 BE/19/2011.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745705.1 BE/38/2011.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
GU179288.1 U8.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KJ361951.1 HAN21.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745678.1 BE/25/2010.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KJ361965.1 PAV11.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
JX512200.1 HAN2.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
JX512203.1 HAN12.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745672.1 BE/29/2011.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745679.1 BE/24/2010.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
GU179291.1 AF1 gO2b.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KJ361946.1 2CEN2.	gO2b	.	C	.	G	.	G	C	G	A	C	.	.	A	-	-	-	-	-	-	.	A	C
KP745680.1 BE/11/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745719.1 BE/26/2010.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745681.1 BE/43/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745702.1 BE/21/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745700.1 BE/4/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745714.1 BE/29/2010.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745706.1 BE/41/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745711.1 BE/24/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745708.1 BE/8/2010.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745707.1 BE/13/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
JX512197.1 6397.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745694.1 BE/12/2010.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745684.1 BE/11/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745692.1 BE/3/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745718.1 CZ/1/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745688.1 BE/12/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745693.1 BE/15/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745651.1 BE/9/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745650.1 BE/1/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KJ361952.1 HAN27.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KJ361964.1 PAV8.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745655.1 BE/3/2010.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745633.1 BE/45/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KJ361949.1 2CEN30.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745636.1 BE/7/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745645.1 BE/13/2010.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745642.1 CZ/1/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745659.1 BE/3/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745666.1 BE/7/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745665.1 BE/16/2012.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
JX512204.1 HAN16 gO3.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KJ361957.1 HAN39.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745670.1 BE/30/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745723.1 BE/37/2011.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745661.1 BE/33/2010.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KJ361961.1 PAV5.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KJ361960.1 PAV4.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KP745664.1 CZ/2/2013.	gO3	.	.	G	T	G	.	.	.	G	C	C	A	G	-	-	-	G	A	T	.	A	C
KJ361969.1 PAV25.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
GQ466044.1 3301.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
KJ426589.1 HAN.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
FJ616285.1 Towne gO4.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
KP745690.1 BE/34/2011.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.

KC519320.1 BE/10/2010.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
KP745720.1 BE/15/2011.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
JX512206.1 HAN22.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
KP745648.1 BE/8/2011.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
KP745717.1 BE/2/2010.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
KP745637.1 BE/9/2011.	gO4	.	.	C	G	T	.	A	.	A	A	.	C	A	G	G	T	G	A	T	.	A	.
GQ221974.1 3157.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KT959235.1 DB.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745728.1 BE/4/2010.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
AY446894.2 Merlin gO5.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745727.1 BE/17/2010.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KJ872539.1 PAV16.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745644.1 BE/31/2010.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745647.1 BE/18/2010.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KJ361971.1 UKNEQAS1.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KJ361963.1 PAV7.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KJ361953.1 HAN30.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KJ361970.1 PAV26.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745662.1 BE/20/2010.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745682.1 BE/46/2011.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745695.1 BE/6/2012.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
GQ221975.1 JP.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
JX512205.1 HAN19.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745669.1 BE/28/2011.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.
KP745671.1 BE/14/2011.	gO5	G	A	T	G	T	.	T	.	.	A	T	C	A	-	-	-	-	-	-	.	.	.

Figure 5.3. Nucleotide sequence alignment of HCMV gO from 163 complete HCMV genomes. The nucleotide sequences correspond to the amino acid region depicted in figure 5.2. The region showing 'molecular tags' unique to each genotype is highlighted. The peculiar case of strain BE/23/2010 (KP745697.1) is discussed in section 5.2.3 'Unique Cases'.

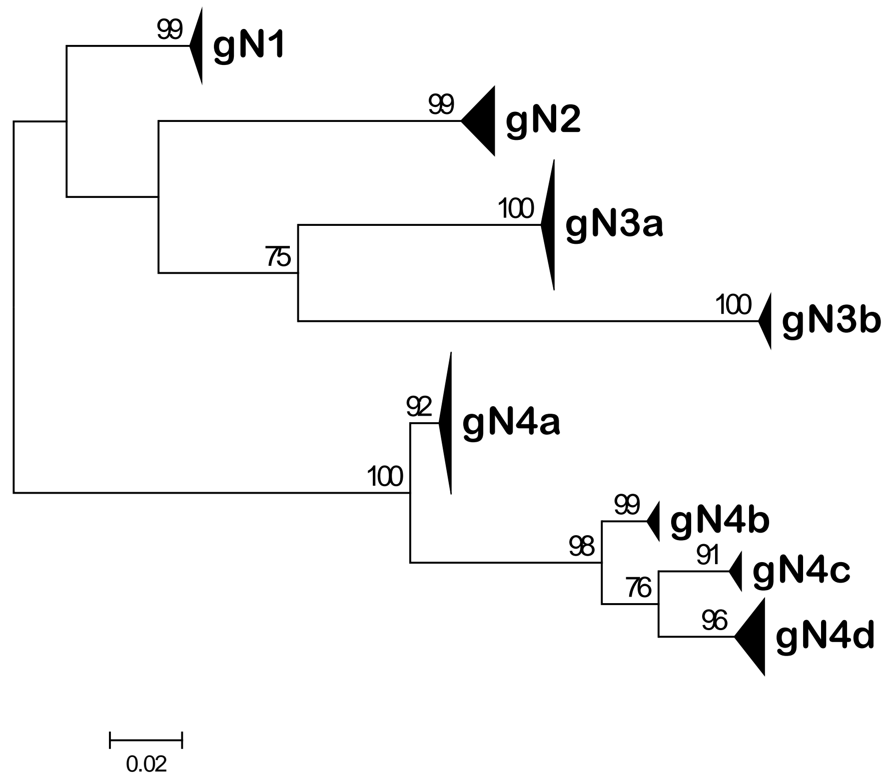


Figure 5.4. Molecular phylogenetic analysis of HCMV gN genotypes. This unrooted consensus tree, showing the eight gN genotypes, was inferred from 1000 bootstrap replicates to represent the evolutionary history of 163 full-length HCMV gN amino acid sequences from complete HCMV genomes available in GenBank database release 211 (National Center for Biotechnology Information (NCBI), 2015). The evolutionary history was inferred by the Maximum Likelihood method based on the JTT matrix-based model and a discrete Gamma distribution with five categories (+*G*, parameter = 0.8361). All positions with less than 95% site coverage were eliminated, leaving 135 amino acid positions in the final dataset. Due to space constraints, the branches are compressed. All sequences grouped to one of the eight HCMV gN genotypes. Analyses were conducted in MEGA 6.06 (Tamura et al., 2013).

Accession No. & Strain ID	Genotype	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
FJ527563.1 AD169 gN1.	gN1	T	S	K	-	S	S	A	S	V	S	T	T	K	L	T	T	V	A	T	T
KP745697.1 BE/23/2010.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.
KP745641.1 BE/31/2011.	gN1	.	.	.	-	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745653.1 BE/22/2011.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745640.1 BE/22/2010.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ361966.1 PAV12.	gN1	.	.	.	-	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745699.1 BE/1/2012.	gN1	.	.	.	-	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745674.1 BE/33/2011.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ361958.1 HAN40.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745657.1 BE/13/2011.	gN1	.	.	.	-	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
KC519319.1 BE/9/2010.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KC519323.1 BE/27/2010.	gN1	.	.	.	-	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ361959.1 PAV1.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745709.1 BE/48/2011.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ872541.1 PAV20.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745722.1 BE/40/2011.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ872542.1 PAV21.	gN1	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GU179290.1 U11.	gN1	.	.	.	-	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
KT634296.1 UKNEQAS2.	gN1	.	.	.	-	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
GU179291.1 AF1 gN2.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745654.1 BE/19/2011.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745656.1 BE/2/2013.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745679.1 BE/24/2010.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745678.1 BE/25/2010.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KJ361956.1 HAN36.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KJ361951.1 HAN21.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745667.1 BE/5/2011.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745672.1 BE/29/2011.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KJ361946.1 2CEN2.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KJ361965.1 PAV11.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
JX512200.1 HAN2.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
JX512203.1 HAN12.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745705.1 BE/38/2011.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
GU179288.1 U8.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
JX512208.1 HAN31.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745686.1 BE/39/2011.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KP745716.1 BE/16/2010.	gN2	.	P	-	-	.	P	S	.	.	.	.	S	.	P	.	.	S	V	.	.
KF021605.1 TR gN3a.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745639.1 BE/10/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745635.1 BE/5/2012.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745649.1 BE/10/2012.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745634.1 BE/32/2010.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745652.1 BE/2/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745658.1 BE/14/2012.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745701.1 BE/6/2010.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745703.1 BE/26/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745696.1 BE/27/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745698.1 BE/20/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745704.1 BE/32/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745725.1 BE/49/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745726.1 BE/30/2010.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745713.1 BE/35/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745715.1 BE/44/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745668.1 BE/18/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745673.1 BE/42/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745660.1 BE/6/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.

KP745663.1 BE/5/2010.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745675.1 BE/23/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745689.1 BE/17/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745691.1 CZ/1/2013.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745676.1 BE/28/2010.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KP745687.1 BE/36/2011.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KC519321.1 BE/11/2010.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KJ361962.1 PAV6.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KJ361967.1 PAV23.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
JX512207.1 HAN28.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KJ361948.1 2CEN15.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
HQ380895.1 JHC.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KJ361950.1 HAN11.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KJ361947.1 2CEN5.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
KJ361954.1 HAN32.	gN3a	.	L	.	S	.	.	S	.	.	.	.	S	.	.	.	.	T	.	.	.
GQ396662.1 HAN38 gN3b.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KP745646.1 BE/8/2012.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KP745683.1 BE/12/2011.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KC519322.1 BE/21/2010.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KP745677.1 BE/1/2010.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KP745712.1 BE/19/2010.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
JX512202.1 HAN8.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
GQ221973.1 HAN13.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KP745638.1 BE/15/2010.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KP745724.1 BE/4/2012.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KP745643.1 CZ/2/2012.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
JX512199.1 HAN1.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
KJ361968.1 PAV24.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
JX512198.1 Davis.	gN3b	.	P	S	P	P	.	S	.	.	.	.	V	.	S	.	.	S	V	.	.
JX512204.1 HAN16 gN4a.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745702.1 BE/21/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745692.1 BE/3/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KJ361949.1 2CEN30.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745693.1 BE/15/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745670.1 BE/30/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745706.1 BE/41/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745681.1 BE/43/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745680.1 BE/11/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745684.1 BE/11/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745700.1 BE/4/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745694.1 BE/12/2010.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
JX512197.1 6397.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745688.1 BE/12/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745711.1 BE/24/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745714.1 BE/29/2010.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745645.1 BE/13/2010.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KJ361961.1 PAV5.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745651.1 BE/9/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745650.1 BE/1/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745718.1 CZ/1/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745719.1 BE/26/2010.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745723.1 BE/37/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745642.1 CZ/1/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KJ361964.1 PAV8.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745636.1 BE/7/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KJ361952.1 HAN27.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745707.1 BE/13/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.

Sequence	Genotype																				
KP745661.1 BE/33/2010.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745666.1 BE/7/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745665.1 BE/16/2012.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745664.1 CZ/2/2013.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KJ361957.1 HAN39.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745655.1 BE/3/2010.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KJ361960.1 PAV4.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745659.1 BE/3/2011.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
KP745708.1 BE/8/2010.	gN4a	S	.	-	-	.	T	R	T	L	T	.	V	.	A	.	.	T	.	.	.
FJ616285.1 Towne gN4b.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
GQ466044.1 3301.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KP745717.1 BE/2/2010.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KP745720.1 BE/15/2011.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KP745648.1 BE/8/2011.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KC519320.1 BE/10/2010.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KP745637.1 BE/9/2011.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KJ426589.1 HAN.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KJ361969.1 PAV25.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
JX512206.1 HAN22.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
KP745690.1 BE/34/2011.	gN4b	S	.	-	-	.	T	H	T	S	T	.	V	.	A	.	.	T	.	.	.
GU937742.1 Toledo gN4c.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
KJ872540.1 PAV18.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
JX512201.1 HAN3.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
KJ361955.1 HAN33.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
KP745710.1 BE/2/2012.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
GU179289.1 VR1814.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
GQ396663.1 HAN20.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
KP745685.1 CZ/3/2012.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
KF297339.1 TB40/E Lisa.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
KP745721.1 BE/14/2010.	gN4c	S	.	-	-	.	T	R	T	S	T	.	V	.	S	.	A	.	.	.	.
AY446894.2 Merlin gN4d.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745633.1 BE/45/2011.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
GQ221974.1 3157.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KJ361971.1 UKNEQAS1.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KT959235.1 DB.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745728.1 BE/4/2010.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KJ872539.1 PAV16.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745727.1 BE/17/2010.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KJ361970.1 PAV26.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745662.1 BE/20/2010.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KJ361953.1 HAN30.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745669.1 BE/28/2011.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745682.1 BE/46/2011.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745671.1 BE/14/2011.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KJ361963.1 PAV7.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
JX512205.1 HAN19.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745695.1 BE/6/2012.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
GQ221975.1 JP.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745647.1 BE/18/2010.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.
KP745644.1 BE/31/2010.	gN4d	S	.	-	-	.	T	H	A	S	T	.	V	.	A	.	.	.	.	.	.

Figure 5.5. Amino acid sequence alignment of HCMV gN from 163 complete HCMV genomes. The eight gN genotypes are colour-coded, with the region showing unique amino acids corresponding to each genotype highlighted.

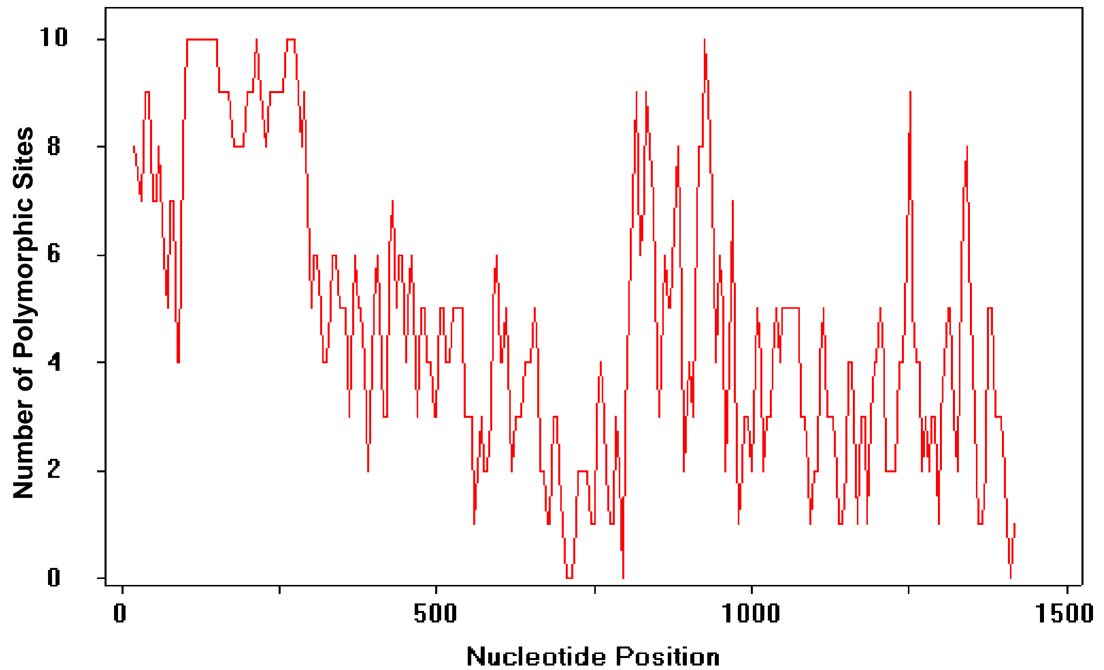
Accession No. & Strain ID	Genotype	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129
FJ527563.1 AD169 gN1.	gN1	C	T	G	C	T	A	G	C	G	T	A	T	C	A	A	C	T	A	C	C
KP745697.1 BE/23/2010.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745641.1 BE/31/2011.	gN1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745653.1 BE/22/2011.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745640.1 BE/22/2010.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ361966.1 PAV12.	gN1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745699.1 BE/1/2012.	gN1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745674.1 BE/33/2011.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ361958.1 HAN40.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745657.1 BE/13/2011.	gN1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KC519319.1 BE/9/2010.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KC519323.1 BE/27/2010.	gN1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ361959.1 PAV1.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745709.1 BE/48/2011.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ872541.1 PAV20.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP745722.1 BE/40/2011.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KJ872542.1 PAV21.	gN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GU179290.1 U11.	gN1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KT634296.1 UKNEQAS2.	gN1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GU179291.1 AF1 gN2.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745654.1 BE/19/2011.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745656.1 BE/2/2013.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745679.1 BE/24/2010.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745678.1 BE/25/2010.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KJ361956.1 HAN36.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KJ361951.1 HAN21.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745667.1 BE/5/2011.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745672.1 BE/29/2011.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KJ361946.1 2CEN2.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KJ361965.1 PAV11.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
JX512200.1 HAN2.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
JX512203.1 HAN12.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745705.1 BE/38/2011.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
GU179288.1 U8.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
JX512208.1 HAN31.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745686.1 BE/39/2011.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KP745716.1 BE/16/2010.	gN2	.	.	T	.	.	.	.	T	.	.	G	.	.	G	.	.	G	.	G	T
KF021605.1 TR gN3a.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745639.1 BE/10/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745635.1 BE/5/2012.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745649.1 BE/10/2012.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745634.1 BE/32/2010.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745652.1 BE/2/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745658.1 BE/14/2012.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745701.1 BE/6/2010.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745703.1 BE/26/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745696.1 BE/27/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745698.1 BE/20/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745704.1 BE/32/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745725.1 BE/49/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745726.1 BE/30/2010.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745713.1 BE/35/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745715.1 BE/44/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745668.1 BE/18/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745673.1 BE/42/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745660.1 BE/6/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.

KP745663.1 BE/5/2010.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745675.1 BE/23/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745689.1 BE/17/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745691.1 CZ/1/2013.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745676.1 BE/28/2010.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KP745687.1 BE/36/2011.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KC519321.1 BE/11/2010.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KJ361962.1 PAV6.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KJ361967.1 PAV23.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
JX512207.1 HAN28.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KJ361948.1 2CEN15.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
HQ380895.1 JHC.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KJ361950.1 HAN11.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KJ361947.1 2CEN5.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
KJ361954.1 HAN32.	gN3a	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	G	.
GQ396662.1 HAN38 gN3b.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KP745646.1 BE/8/2012.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KP745683.1 BE/12/2011.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KC519322.1 BE/21/2010.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KP745677.1 BE/1/2010.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KP745712.1 BE/19/2010.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
JX512202.1 HAN8.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
GQ221973.1 HAN13.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KP745638.1 BE/15/2010.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KP745724.1 BE/4/2012.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KP745643.1 CZ/2/2012.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
JX512199.1 HAN1.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
KJ361968.1 PAV24.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
JX512198.1 Davis.	gN3b	G	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	G	G	T	A
JX512204.1 HAN16 gN4a.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745702.1 BE/21/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745692.1 BE/3/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KJ361949.1 2CEN30.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745693.1 BE/15/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745670.1 BE/30/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745706.1 BE/41/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745681.1 BE/43/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745680.1 BE/11/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745684.1 BE/11/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745700.1 BE/4/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745694.1 BE/12/2010.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
JX512197.1 6397.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745688.1 BE/12/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745711.1 BE/24/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745714.1 BE/29/2010.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745645.1 BE/13/2010.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KJ361961.1 PAV5.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745651.1 BE/9/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745650.1 BE/1/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745718.1 CZ/1/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745719.1 BE/26/2010.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745723.1 BE/37/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745642.1 CZ/1/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KJ361964.1 PAV8.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745636.1 BE/7/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KJ361952.1 HAN27.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745707.1 BE/13/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G

KP745661.1 BE/33/2010.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745666.1 BE/7/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745665.1 BE/16/2012.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745664.1 CZ/2/2013.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KJ361957.1 HAN39.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745655.1 BE/3/2010.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KJ361960.1 PAV4.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745659.1 BE/3/2011.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
KP745708.1 BE/8/2010.	gN4a	.	.	C	G	C	.	C	.	T	.	.	A	.	.	.	.	C	G	T	G
FJ616285.1 Towne gN4b.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
GQ466044.1 3301.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KP745717.1 BE/2/2010.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KP745720.1 BE/15/2011.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KP745648.1 BE/8/2011.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KC519320.1 BE/10/2010.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KP745637.1 BE/9/2011.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KJ426589.1 HAN.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KJ361969.1 PAV25.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
JX512206.1 HAN22.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
KP745690.1 BE/34/2011.	gN4b	.	.	C	A	C	.	C	.	T	C	.	A	.	G	.	.	C	G	T	G
GU937742.1 Toledo gN4c.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KJ872540.1 PAV18.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
JX512201.1 HAN3.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KJ361955.1 HAN33.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745710.1 BE/2/2012.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
GU179289.1 VR1814.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
GQ396663.1 HAN20.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745685.1 CZ/3/2012.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KF297339.1 TB40/E Lisa.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745721.1 BE/14/2010.	gN4c	.	.	C	G	C	.	C	.	T	C	.	A	.	.	.	.	C	G	T	G
AY446894.2 Merlin gN4d.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745633.1 BE/45/2011.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
GQ221974.1 3157.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KJ361971.1 UKNEQAS1.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KT959235.1 DB.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745728.1 BE/4/2010.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KJ872539.1 PAV16.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745727.1 BE/17/2010.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KJ361970.1 PAV26.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745662.1 BE/20/2010.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KJ361953.1 HAN30.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745669.1 BE/28/2011.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745682.1 BE/46/2011.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745671.1 BE/14/2011.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KJ361963.1 PAV7.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
JX512205.1 HAN19.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745695.1 BE/6/2012.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
GQ221975.1 JP.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745647.1 BE/18/2010.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G
KP745644.1 BE/31/2010.	gN4d	.	.	C	A	C	G	C	.	T	C	.	A	.	.	.	.	C	G	T	G

Figure 5.6. Nucleotide sequence alignment of HCMV gN from 163 complete HCMV genomes. The nucleotide sequences correspond to the amino acid region depicted in figure 5.5. The region showing 'molecular tags' unique to each genotype is highlighted.

[A]



[B]

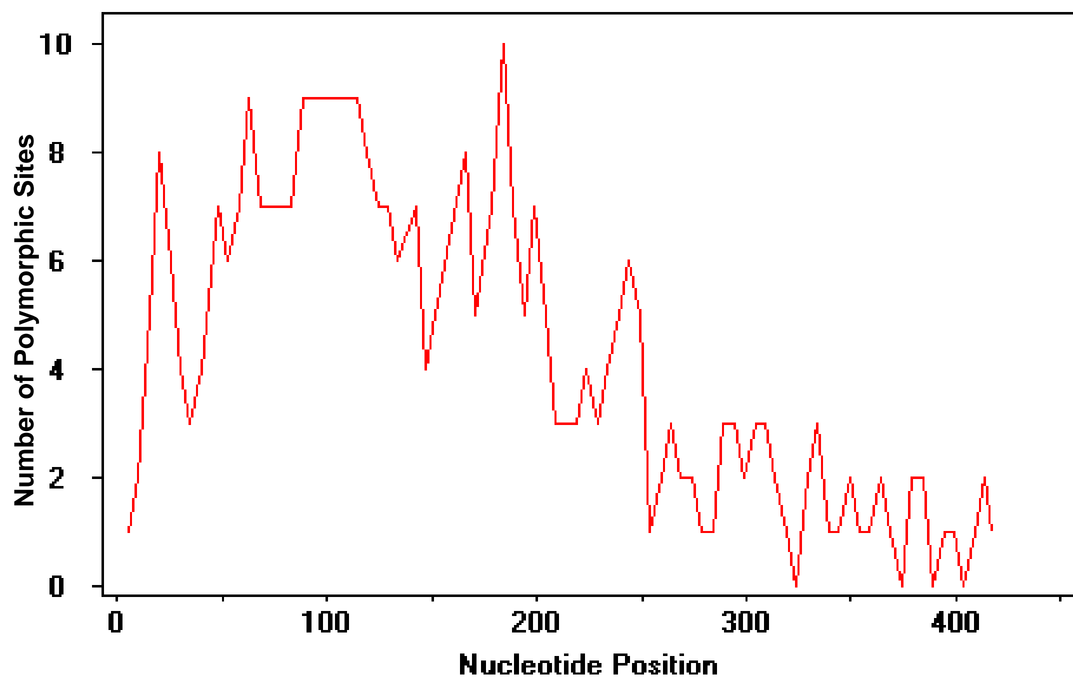


Figure 5.7. Polymorphic sites across 163 HCMV gO and gN sequences. [A] gO and [B] gN nucleotide sequences of 163 complete HCMV genomes available in GenBank release 211 (National Center for Biotechnology Information (NCBI), 2015) were aligned and polymorphic sites plotted in DnaSP 5.10.01 (Librado and Rozas, 2009) with a sliding window of length 10 and step size 5.

	Genotype		Reference Strain	Nucleotide sequence	Coordinates
[A] UL74 (gO)					
N-terminus tags	gO1a	:1	FJ527563.1 [AD169]	AAACGACTATTT	c108270-108281
		:2	KC519323.1 [BE/27/2010]	AAACGACAATTT	c108441-108452
	gO1b		KF021605.1 [TR]	AAAAGGATATCT	c108658-108669
	gO1c		GU937742.1 [Toledo]	AAAGGGAACCTT	c108248-108259
	gO2a		JX512198.1 [Davis]	AACCTATTCCTT	c103114-103125
	gO2b		GU179291.1 [AF1]	AGAGCGACATAT	c108478-108489
	gO3		JX512197.1 [6397]	GTAGCCAGGATT	c108267-108278
	gO4		FJ616285.1 [Towne]	CGACAGGTGATT	c107847-107858
	gO5		AY446894.2 [Merlin]	TGTCTACATCAT	c108632-108643
'mid'-N-terminus tags	gO1a		FJ527563.1 [AD169]	CCTTGTGGTACTG	c108045-108057
	gO1b		KF021605.1 [TR]	TCTTGCGGTACGG	c108430-108442
	gO1c		GU937742.1 [Toledo]	TCTTGTGGTACAG	c108020-108032
	gO2a		JX512198.1 [Davis]	TCGTGTGGCGCAG	c102886-102898
	gO2b		GU179291.1 [AF1]	CCTTGCGGTACAG	c108250-108262
	gO3		JX512197.1 [6397]	TCTTGTGGCACTG	c108039-108051
	gO4	:a	FJ616285.1 [Towne]	TCCTGTGGTACGA	c107616-107628
		:b	JX512206.1 [HAN22]	TCCTGTGGCACGA	c108469-108481
	gO5	:a	AY446894.2 [Merlin]	CCTTGCGGCACAG	c108395-108407
		:b	GQ221974.1 [3157]	CCTTGTGGCACAG	c108218-108230
C-terminus tags	gO1a		FJ527563.1 [AD169]	TATTACTACCGCC	c107581-107593
	gO1b		KF021605.1 [TR]	TGTTACTACCACC	c107966-107978
	gO1c		GU937742.1 [Toledo]	GGTTACCACCAGC	c107556-107568
	gO2a		JX512198.1 [Davis]	TGTTACCACCACC	c102422-102434
	gO2b		GU179291.1 [AF1]	TGTTACAACCACC	C107786-107798
	gO3		JX512197.1 [6397]	TGCTACCACCACT	c107584-107596
	gO4		FJ616285.1 [Towne]	TCCTATTGTCCCA	c107164-107176
	gO5		AY446894.2 [Merlin]	TGCTACCGCTGCT	c107931-107943
[B] UL73 (gN)					
N-terminus tags	gN1		FJ527563.1 [AD169]	GCGTATCAACTACC	106804-106817
	gN2		GU179291.1 [AF1]	GTGTGTCGACGAGT	107012-107025
	gN3a		KF021605.1 [TR]	GCGTGTCAACAAGC	107201-107214
	gN3b		JX512198.1 [Davis]	GTGTATCAACGGTA	101651-101664
	gN4a		JX512204.1 [HAN16]	GCACCTTAACAACC	106966-106979
	gN4b		FJ616285.1 [Towne]	ACACCTCAACGACC	106387-106400
	gN4c		GU937742.1 [Toledo]	GCACCTCAACAACC	106779-106792
	gN4d		AY446894.2 [Merlin]	ACGCCTCAACAACC	107154-107167

Table 5.1. Molecular tags developed for detection of HCMV gO and gN genotypes. These tags were determined from amino acid and nucleotide sequence alignments of respective genes for 163 complete HCMV genomes available in GenBank release 211 (National Center for Biotechnology Information (NCBI), 2015). The coordinates indicate the location of each tag with respect to the reference strain per respective genotype. c, complement.

5.2.2 Verification of Tags and Development of *Perl* Scripts

NCBI's nucleotide collection (nr/nt) database, organism Human Herpesvirus 5 (taxid 10359), was queried using each molecular tag sequence via NCBI BLASTn program (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) (Altschul et al., 1990) with default parameters in order to verify that each tag was both unique to a given HCMV gO or gN genotype and also that the tag occurred only once within a given full HCMV genome sequence. All tags fulfilled these two critical criteria. To aid quick determination of which tag was present, and therefore which genotype a sequence belonged to, *perl* scripts were developed for application to the two commonly used sequence formats: FASTA (for archived NCBI database sequences) and FASTQ (for NGS reads from Illumina sequencing). The scripts search for multiple tags simultaneously in each FASTA sequence or set of FASTQ reads in both the forward and complementary / reverse orientation and return the genotype, tag sequence, and number of times the tag occurs in the interrogated file. The output is conveniently tab delimited and can easily be copied directly from a UNIX terminal or piped to a plain text file for easy manipulation across various word processing and data analysis programs including Microsoft Excel. Figure 5.8 shows screenshots of typical outputs. The 163 complete HCMV genomes were rechecked using the script and results compared to the gO genotype assignment from the phylogenetic tree construction (figure 5.1), with which they were completely consistent (table 5.2). As an additional step to validate the tags, the FASTA script was applied to all available gO sequences (partial or complete CDS). 195 such sequences were identified and downloaded from GenBank release 211 (National Center for Biotechnology Information (NCBI), 2015). The script correctly assigned gO genotypes to all but two sequences (table 5.3): KT988000.1 and KT988001.1, which are incomplete gene sequences falling outside the three gO regions targeted by our tags. These are discussed further in section 5.2.3 'Unique Cases'. The gN tags were equally checked against all available gN sequences in GenBank release 211, and of 465 partial and complete CDSs, the Script correctly identified gN genotypes for 456 sequences (table 5.4). The remaining 9 sequences could not be assigned to a genotype for various reasons related to the quality/nature of published sequence, as discussed in section 5.2.3 'Unique Cases'.

[A]

```
kunda@plum-p3:~/2016bm/174R
[kunda@plum-p3 174R]$ ./fq.typetriON.pl 174R1_trimmed_paired.fastq
gO1a1    6      AAACGACTATTT
CgO3     433    TGCTACCACCACT
gN1      7      GCGTATCAACTACC
gO4      6      AAACAGGTGATT
MgO4a    15     TCCTGTGGTACGA
gN2      78     GTGTGTCGACGAGT
MgO1c    2      TCTTGTGGTACAG
CgO4     14     TCCTATTGTCCCA
gO3a     490    CGAGCCAGGATT
MgO4b    1      TCCTGTGGCACGA
gO1b     22     AAAAGGATATCT
gN3a     6      GCGTGTCAACAAGC
gO3b     6      CAAGCCAGGATT
MgO2b    286    CCTTGCGGTACAG
MgO5b    3      CCTTGTGGCACAG
gO1c     4      AAAGGGAACCTT
gO2b     219    AGAGCGACATAT
CgO1a    5      TATTACTACCGCC
gO5      3      TGTCTACATCAT
CgO2a    7      TGTTACCACCACC
MgO1a    50     CCTTGTGGTACTG
gN4c     1      GCACCTCAACAACC
```

[B]

```
kunda@plum-p3:~/genbankseqs/ul74/CDSgenomes
  GNU nano 2.0.9              File: gOgenotypes.txt                    Modified

AY582475.1.fasta    gO5     1    TGTCTACATCAT
AY582477.1.fasta    gO5     1    TGTCTACATCAT
AY582479.1.fasta    gO5     1    TGTCTACATCAT
AY582480.1.fasta    gO2b    1    AGAGCGACATAT
AY582481.1.fasta    gO2b    1    AGAGCGACATAT
AY582482.1.fasta    gO3a    1    CGAGCCAGGATT
EU348351.1.fasta    CgO1b   1    TGTTACTACCACC    gO1b    1    AAAAGGATATCT
EU348352.1.fasta    CgO1b   1    TGTTACTACCACC    gO1b    1    AAAAGGATATCT
EU348353.1.fasta    CgO1b   1    TGTTACTACCACC    gO1b    1    AAAAGGATATCT
EU348354.1.fasta    CgO2a   1    TGTTACCACCACC    gO2a    1    AACCTATTCCTT
EU348355.1.fasta    CgO2b   1    TGTTACAACCACC    gO2b    1    AGAGCGACATAT
EU348356.1.fasta    CgO5    1    TGCTACCGCTGCT    gO5     1    TGTCTACATCAT
EU348357.1.fasta    CgO5    1    TGCTACCGCTGCT    gO5     1    TGTCTACATCAT
EU348358.1.fasta    CgO5    1    TGCTACCGCTGCT    gO5     1    TGTCTACATCAT
EU348359.1.fasta    CgO5    1    TGCTACCGCTGCT    gO5     1    TGTCTACATCAT
EU348360.1.fasta    CgO5    1    TGCTACCGCTGCT    gO5     1    TGTCTACATCAT
EU348361.1.fasta    CgO1a   1    TATTACTACCGCC
EU348362.1.fasta    gO1a2   1    AAACGACAATTT     CgO1a   1    TATTACTACCGCC
EU348363.1.fasta    gO1a2   1    AAACGACAATTT     CgO1a   1    TATTACTACCGCC
EU348364.1.fasta    CgO3    1    TGCTACCACCACT    gO3a    1    CGAGCCAGGATT
EU377488.1.fasta
EU377490.1.fasta
EU686463.1.fasta    gO1a1   1    AAACGACTATTT
```

Figure 5.8. Screenshots of search results returned by *perl* scripts. [A]shows the typical output from searching FASTQ reads, while [B] depicts the typical tab delimited format when output from several FASTA files is piped to a plain text file.

	GenBank Accession No. & Strain ID	gO Genotype			gN Genotype
		N-terminus	mid-N- terminus	C- terminus	
1	FJ527563.1_AD169	gO1a	gO1a	gO1a	gN1
2	GU179290.1_U11	gO1a	gO1a	gO1a	gN1
3	KC519319.1_BE/9/2010	gO1a	gO1a	gO1a	gN1
4	KC519323.1_BE/27/2010	gO1a	gO1a	gO1a	gN1
5	KJ361950.1_HAN11	gO1a	gO1a	gO1a	gN3a
6	KJ361958.1_HAN40	gO1a	gO1a	gO1a	gN1
7	KJ361959.1_PAV1	gO1a	gO1a	gO1a	gN1
8	KJ361966.1_PAV12	gO1a	gO1a	gO1a	gN1
9	KJ872541.1_PAV20	gO1a	gO1a	gO1a	gN1
10	KJ872542.1_PAV21	gO1a	gO1a	gO1a	gN1
11	KP745640.1_BE/22/2010	gO1a	gO1a	gO1a	gN1
12	KP745641.1_BE/31/2011	gO1a	gO1a	gO1a	gN1
13	KP745653.1_BE/22/2011	gO1a	gO1a	gO1a	gN1
14	KP745657.1_BE/13/2011	gO1a	gO1a	gO1a	gN1
15	KP745674.1_BE/33/2011	gO1a	gO1a	gO1a	gN1
16	KP745697.1_BE/23/2010	gO1c	gO1a	gO1a	gN1
17	KP745699.1_BE/1/2012	gO1a	gO1a	gO1a	gN1
18	KP745709.1_BE/48/2011	gO1a	gO1a	gO1a	gN1
19	KP745722.1_BE/40/2011	gO1a	gO1a	gO1a	gN1
20	KT634296.1_UKNEQAS2	gO1a	gO1a	gO1a	gN1
21	HQ380895.1_JHC	gO1b	gO1b	gO1b	gN3a
22	JX512207.1_HAN28	gO1b	gO1b	gO1b	gN3a
23	KC519321.1_BE/11/2010	gO1b	gO1b	gO1b	gN3a
24	KF021605.1_TR	gO1b	gO1b	gO1b	gN3a
25	KJ361947.1_2CEN5	gO1b	gO1b	gO1b	gN3a
26	KJ361948.1_2CEN15	gO1b	gO1b	gO1b	gN3a
27	KJ361954.1_HAN32	gO1b	gO1b	gO1b	gN3a
28	KJ361962.1_PAV6	gO1b	gO1b	gO1b	gN3a
29	KJ361967.1_PAV23	gO1b	gO1b	gO1b	gN3a
30	KP745634.1_BE/32/2010	gO1b	gO1b	gO1b	gN3a
31	KP745635.1_BE/5/2012	gO1b	gO1b	gO1b	gN3a
32	KP745639.1_BE/10/2011	gO1b	gO1b	gO1b	gN3a
33	KP745649.1_BE/10/2012	gO1b	gO1b	gO1b	gN3a
34	KP745652.1_BE/2/2011	gO1b	gO1b	gO1b	gN3a
35	KP745658.1_BE/14/2012	gO1b	gO1b	gO1b	gN3a
36	KP745660.1_BE/6/2011	gO1b	gO1b	gO1b	gN3a
37	KP745663.1_BE/5/2010	gO1b	gO1b	gO1b	gN3a
38	KP745668.1_BE/18/2011	gO1b	gO1b	gO1b	gN3a
39	KP745673.1_BE/42/2011	gO1b	gO1b	gO1b	gN3a
40	KP745675.1_BE/23/2011	gO1b	gO1b	gO1b	gN3a
41	KP745676.1_BE/28/2010	gO1b	gO1b	gO1b	gN3a
42	KP745687.1_BE/36/2011	gO1b	gO1b	gO1b	gN3a
43	KP745689.1_BE/17/2011	gO1b	gO1b	gO1b	gN3a
44	KP745691.1_CZ/1/2013	gO1b	gO1b	gO1b	gN3a
45	KP745696.1_BE/27/2011	gO1b	gO1b	gO1b	gN3a
46	KP745698.1_BE/20/2011	gO1b	gO1b	gO1b	gN3a
47	KP745701.1_BE/6/2010	gO1b	gO1b	gO1b	gN3a
48	KP745703.1_BE/26/2011	gO1b	gO1b	gO1b	gN3a
49	KP745704.1_BE/32/2011	gO1b	gO1b	gO1b	gN3a

	GenBank Accession No. & Strain ID	gO Genotype			gN Genotype
		N-terminus	mid-N- terminus	C- terminus	
50	KP745713.1_BE/35/2011	gO1b	gO1b	gO1b	gN3a
51	KP745715.1_BE/44/2011	gO1b	gO1b	gO1b	gN3a
52	KP745725.1_BE/49/2011	gO1b	gO1b	gO1b	gN3a
53	KP745726.1_BE/30/2010	gO1b	gO1b	gO1b	gN3a
54	GQ396663.1_HAN20	gO1c	gO1c	gO1c	gN4c
55	GU179289.1_VR1814	gO1c	gO1c	gO1c	gN4c
56	GU937742.1_Toledo	gO1c	gO1c	gO1c	gN4c
57	JX512201.1_HAN3	gO1c	gO1c	gO1c	gN4c
58	KF297339.1_TB40/E	gO1c	gO1c	gO1c	gN4c
59	KJ361955.1_HAN33	gO1c	gO1c	gO1c	gN4c
60	KJ872540.1_PAV18	gO1c	gO1c	gO1c	gN4c
61	KP745685.1_CZ/3/2012	gO1c	gO1c	gO1c	gN4c
62	KP745710.1_BE/2/2012	gO1c	gO1c	gO1c	gN4c
63	KP745721.1_BE/14/2010	gO1c	gO1c	gO1c	gN4c
64	GQ221973.1_HAN13	gO2a	gO2a	gO2a	gN3b
65	GQ396662.1_HAN38	gO2a	gO2a	gO2a	gN3b
66	JX512198.1_Davis	gO2a	gO2a	gO2a	gN3b
67	JX512199.1_HAN1	gO2a	gO2a	gO2a	gN3b
68	JX512202.1_HAN8	gO2a	gO2a	gO2a	gN3b
69	KC519322.1_BE/21/2010	gO2a	gO2a	gO2a	gN3b
70	KJ361968.1_PAV24	gO2a	gO2a	gO2a	gN3b
71	KP745638.1_BE/15/2010	gO2a	gO2a	gO2a	gN3b
72	KP745643.1_CZ/2/2012	gO2a	gO2a	gO2a	gN3b
73	KP745646.1_BE/8/2012	gO2a	gO2a	gO2a	gN3b
74	KP745677.1_BE/1/2010	gO2a	gO2a	gO2a	gN3b
75	KP745683.1_BE/12/2011	gO2a	gO2a	gO2a	gN3b
76	KP745712.1_BE/19/2010	gO2a	gO2a	gO2a	gN3b
77	KP745724.1_BE/4/2012	gO2a	gO2a	gO2a	gN3b
78	GU179288.1_U8	gO2b	gO2b	gO2b	gN2
79	GU179291.1_AF1	gO2b	gO2b	gO2b	gN2
80	JX512200.1_HAN2	gO2b	gO2b	gO2b	gN2
81	JX512203.1_HAN12	gO2b	gO2b	gO2b	gN2
82	JX512208.1_HAN31	gO2b	gO2b	gO2b	gN2
83	KJ361946.1_2CEN2	gO2b	gO2b	gO2b	gN2
84	KJ361951.1_HAN21	gO2b	gO2b	gO2b	gN2
85	KJ361956.1_HAN36	gO2b	gO2b	gO2b	gN2
86	KJ361965.1_PAV11	gO2b	gO2b	gO2b	gN2
87	KP745654.1_BE/19/2011	gO2b	gO2b	gO2b	gN2
88	KP745656.1_BE/2/2013	gO2b	gO2b	gO2b	gN2
89	KP745667.1_BE/5/2011	gO2b	gO2b	gO2b	gN2
90	KP745672.1_BE/29/2011	gO2b	gO2b	gO2b	gN2
91	KP745678.1_BE/25/2010	gO2b	gO2b	gO2b	gN2
92	KP745679.1_BE/24/2010	gO2b	gO2b	gO2b	gN2
93	KP745686.1_BE/39/2011	gO2b	gO2b	gO2b	gN2
94	KP745705.1_BE/38/2011	gO2b	gO2b	gO2b	gN2
95	KP745716.1_BE/16/2010	gO2b	gO2b	gO2b	gN2
96	JX512197.1_6397	gO3a	gO3	gO3	gN4a
97	JX512204.1_HAN16	gO3a	gO3	gO3	gN4a
98	KJ361949.1_2CEN30	gO3a	gO3	gO3	gN4a
99	KJ361952.1_HAN27	gO3a	gO3	gO3	gN4a

	GenBank Accession No. & Strain ID	gO Genotype			gN Genotype
		N-terminus	mid-N- terminus	C- terminus	
100	KJ361957.1_HAN39	gO3a	gO3	gO3	gN4a
101	KJ361960.1_PAV4	gO3a	gO3	gO3	gN4a
102	KJ361961.1_PAV5	gO3a	gO3	gO3	gN4a
103	KJ361964.1_PAV8	gO3a	gO3	gO3	gN4a
104	KP745633.1_BE/45/2011	gO3a	gO3	gO3	gN4d
105	KP745636.1_BE/7/2011	gO3a	gO3	gO3	gN4a
106	KP745642.1_CZ/1/2012	gO3a	gO3	gO3	gN4a
107	KP745645.1_BE/13/2010	gO3a	gO3	gO3	gN4a
108	KP745650.1_BE/1/2011	gO3a	gO3	gO3	gN4a
109	KP745651.1_BE/9/2012	gO3a	gO3	gO3	gN4a
110	KP745655.1_BE/3/2010	gO3a	gO3	gO3	gN4a
111	KP745659.1_BE/3/2011	gO3a	gO3	gO3	gN4a
112	KP745661.1_BE/33/2010	gO3a	gO3	gO3	gN4a
113	KP745664.1_CZ/2/2013	gO3a	gO3	gO3	gN4a
114	KP745665.1_BE/16/2012	gO3a	gO3	gO3	gN4a
115	KP745666.1_BE/7/2012	gO3a	gO3	gO3	gN4a
116	KP745670.1_BE/30/2011	gO3a	gO3	gO3	gN4a
117	KP745680.1_BE/11/2012	gO3a	gO3	gO3	gN4a
118	KP745681.1_BE/43/2011	gO3a	gO3	gO3	gN4a
119	KP745684.1_BE/11/2011	gO3a	gO3	gO3	gN4a
120	KP745688.1_BE/12/2012	gO3a	gO3	gO3	gN4a
121	KP745692.1_BE/3/2012	gO3a	gO3	gO3	gN4a
122	KP745693.1_BE/15/2012	gO3a	gO3	gO3	gN4a
123	KP745694.1_BE/12/2010	gO3a	gO3	gO3	gN4a
124	KP745700.1_BE/4/2011	gO3a	gO3	gO3	gN4a
125	KP745702.1_BE/21/2011	gO3a	gO3	gO3	gN4a
126	KP745706.1_BE/41/2011	gO3a	gO3	gO3	gN4a
127	KP745707.1_BE/13/2012	gO3a	gO3	gO3	gN4a
128	KP745708.1_BE/8/2010	gO3a	gO3	gO3	gN4a
129	KP745711.1_BE/24/2011	gO3a	gO3	gO3	gN4a
130	KP745714.1_BE/29/2010	gO3a	gO3	gO3	gN4a
131	KP745718.1_CZ/1/2011	gO3a	gO3	gO3	gN4a
132	KP745719.1_BE/26/2010	gO3a	gO3	gO3	gN4a
133	KP745723.1_BE/37/2011	gO3a	gO3	gO3	gN4a
134	FJ616285.1_Towne	gO4	gO4	gO4	gN4b
135	GQ466044.1_3301	gO4	gO4	gO4	gN4b
136	JX512206.1_HAN22	gO4	gO4	gO4	gN4b
137	KC519320.1_BE/10/2010	gO4	gO4	gO4	gN4b
138	KJ361969.1_PAV25	gO4	gO4	gO4	gN4b
139	KJ426589.1_HAN	gO4	gO4	gO4	gN4b
140	KP745637.1_BE/9/2011	gO4	gO4	gO4	gN4b
141	KP745648.1_BE/8/2011	gO4	gO4	gO4	gN4b
142	KP745690.1_BE/34/2011	gO4	gO4	gO4	gN4b
143	KP745717.1_BE/2/2010	gO4	gO4	gO4	gN4b
144	KP745720.1_BE/15/2011	gO4	gO4	gO4	gN4b
145	AY446894.2_Merlin	gO5	gO5	gO5	gN4d
146	GQ221974.1_3157	gO5	gO5	gO5	gN4d
147	GQ221975.1_JP	gO5	gO5	gO5	gN4d
148	JX512205.1_HAN19	gO5	gO5	gO5	gN4d
149	KJ361953.1_HAN30	gO5	gO5	gO5	gN4d

	GenBank Accession No. & Strain ID	gO Genotype			gN Genotype
		N-terminus	mid-N- terminus	C- terminus	
150	KJ361963.1_PAV7	gO5	gO5	gO5	gN4d
151	KJ361970.1_PAV26	gO5	gO5	gO5	gN4d
152	KJ361971.1_UKNEQAS1	gO5	gO5	gO5	gN4d
153	KJ872539.1_PAV16	gO5	gO5	gO5	gN4d
154	KP745644.1_BE/31/2010	gO5	gO5	gO5	gN4d
155	KP745647.1_BE/18/2010	gO5	gO5	gO5	gN4d
156	KP745662.1_BE/20/2010	gO5	gO5	gO5	gN4d
157	KP745669.1_BE/28/2011	gO5	gO5	gO5	gN4d
158	KP745671.1_BE/14/2011	gO5	gO5	gO5	gN4d
159	KP745682.1_BE/46/2011	gO5	gO5	gO5	gN4d
160	KP745695.1_BE/6/2012	gO5	gO5	gO5	gN4d
161	KP745727.1_BE/17/2010	gO5	gO5	gO5	gN4d
162	KP745728.1_BE/4/2010	gO5	gO5	gO5	gN4d
163	KT959235.1_DB	gO5	gO5	gO5	gN4d

Table 5.2. HCMV gO and gN genotype identification using molecular tags developed during this thesis. HCMV gO genotypes were correctly assigned to all 163 complete HCMV genomes available in GenBank release 211. The sequences were screened with our genotyping tags at three loci (2 in the N-terminus and one in the C-terminus) within the gO gene, and one locus at the N-terminus of gN. KJ361950.1 (HAN11) and KP745697.1 (BE/23/2010) showed peculiarities and are discussed under 'Unique Cases' in section 5.2.3.

	GenBank Accession	gO genotype		
		N-terminus tag	mid-N-terminus tag	N-terminus tag
1	AF531320.1	gO1a	gO1a	gO1a
2	AF531322.1	gO1a	gO1a	gO1a
3	AF531325.1	gO1a	gO1a	gO1a
4	AF531330.1	gO1a	gO1a	gO1a
5	AF531337.1	gO1a	gO1a	gO1a
6	AF531344.1	gO1a	gO1a	gO1a
7	AF531346.1	gO1a	gO1a	gO1a
8	AF531347.1	gO1a	gO1a	gO1a
9	EU348361.1	-	gO1a	gO1a
10	EU348362.1	gO1a	gO1a	gO1a
11	EU348363.1	gO1a	gO1a	gO1a
12	GQ227785.1	gO1a	gO1a	gO1a
13	AF531316.1	gO1b	gO1b	gO1b
14	AF531319.1	gO1b	gO1b	gO1b
15	AF531326.1	gO1b	gO1b	gO1b
16	AF531333.1	gO1b	gO1b	gO1b
17	AF531334.1	gO1b	gO1b	gO1b
18	AF531345.1	gO1b	gO1b	gO1b
19	AF531349.1	gO1b	gO1b	gO1b
20	AF531354.1	gO1b	gO1b	gO1b
21	EU348351.1	gO1b	gO1b	gO1b
22	EU348352.1	gO1b	gO1b	gO1b
23	EU348353.1	gO1b	gO1b	gO1b
24	GQ227786.1	gO1b	gO1b	gO1b
25	AF531340.1	gO1c	gO1c	gO1c
26	GQ227762.1	gO1c	gO1c	gO1c
27	AF531315.1	gO2a	gO2a	gO2a
28	AF531324.1	gO2a	gO2a	gO2a
29	AF531336.1	gO2a	gO2a	gO2a
30	AF531343.1	gO2a	gO2a	gO2a
31	EU348354.1	gO2a	gO2a	gO2a
32	AF531317.1	gO2b	gO2b	gO2b
33	AF531328.1	gO2b	gO2b	gO2b
34	AF531329.1	gO2b	gO2b	gO2b
35	AF531335.1	gO2b	gO2b	gO2b
36	AF531339.1	gO2b	gO2b	gO2b
37	AF531341.1	gO2b	gO2b	gO2b
38	AF531342.1	gO2b	gO2b	gO2b
39	AF531353.1	gO2b	gO2b	gO2b
40	EU348355.1	gO2b	gO2b	gO2b
41	GQ227753.1	gO2b	gO2b	gO2b
42	GQ227761.1	gO2b	gO2b	gO2b
43	AF531318.1	gO3a	gO3	gO3
44	AF531327.1	gO3a	gO3	gO3
45	AF531338.1	gO3a	gO3	gO3

	GenBank Accession	gO genotype		
		N-terminus tag	mid-N-terminus tag	N-terminus tag
46	AF531348.1	gO3a	gO3	gO3
47	AF531350.1	gO3a	gO3	gO3
48	AF531351.1	gO3a	gO3	gO3
49	AF531352.1	gO3a	gO3	gO3
50	EU348364.1	gO3a	gO3	gO3
51	GQ227765.1	gO3a	gO3	gO3
52	GQ227770.1	gO3a	gO3	gO3
53	AF531323.1	gO4	gO4	gO4
54	AF531332.1	gO4	gO4	gO4
55	GQ227755.1	gO4	gO4	gO4
56	GQ227760.1	gO4	gO4	gO4
57	AF531321.1	gO5	gO5	gO5
58	EU348356.1	gO5	gO5	gO5
59	EU348357.1	gO5	gO5	gO5
60	EU348358.1	gO5	gO5	gO5
61	EU348359.1	gO5	gO5	gO5
62	EU348360.1	gO5	gO5	gO5
63	GQ227780.1	gO5	gO5	gO5
64	AY326963.1	gO1a	gO1a	-
65	AY326964.1	gO1a	gO1a	-
66	AY326966.1	gO1a	gO1a	-
67	AY326973.1	gO1a	gO1a	-
68	AY326974.1	gO1a	gO1a	-
69	AY326975.1	gO1a	gO1a	-
70	AY326982.1	gO1a	gO1a	-
71	AY326983.1	gO1a	gO1a	-
72	AY326984.1	gO1a	gO1a	-
73	AY326985.1	gO1a	gO1a	-
74	AY326986.1	gO1a	gO1a	-
75	AY582469.1	gO1a	gO1a	-
76	AY582470.1	gO1a	gO1a	-
77	AY582471.1	gO1a	gO1a	-
78	AY582472.1	gO1a	gO1a	-
79	EU686463.1	gO1a	gO1a	-
80	EU686464.1	gO1a	gO1a	-
81	EU686465.1	gO1a	gO1a	-
82	EU686466.1	gO1a	gO1a	-
83	EU686467.1	gO1a	gO1a	-
84	EU686508.1	gO1a	gO1a	-
85	EU686511.1	gO1a	gO1a	-
86	AY326968.1	gO1b	gO1b	-
87	AY326980.1	gO1b	gO1b	-
88	AY582439.1	gO1b	gO1b	-
89	AY582440.1	gO1b	gO1b	-
90	AY582441.1	gO1b	gO1b	-

	GenBank Accession	gO genotype		
		N-terminus tag	mid-N-terminus tag	N-terminus tag
91	AY582442.1	gO1b	gO1b	-
92	AY582443.1	gO1b	gO1b	-
93	AY582444.1	gO1b	gO1b	-
94	AY582445.1	gO1b	gO1b	-
95	AY582473.1	gO1b	gO1b	-
96	EU686468.1	gO1b	gO1b	-
97	EU686469.1	gO1b	gO1b	-
98	EU686470.1	gO1b	gO1b	-
99	EU686471.1	gO1b	gO1b	-
100	EU686472.1	gO1b	gO1b	-
101	EU686473.1	gO1b	gO1b	-
102	EU686474.1	gO1b	gO1b	-
103	EU686475.1	gO1b	gO1b	-
104	EU686476.1	gO1b	gO1b	-
105	EU686477.1	gO1b	gO1b	-
106	EU686478.1	gO1b	gO1b	-
107	EU686507.1	gO1b	gO1b	-
108	EU686509.1	gO1b	gO1b	-
109	EU686512.1	gO1b	gO1b	-
110	EU686513.1	gO1b	gO1b	-
111	EU686514.1	gO1b	gO1b	-
112	EU686515.1	gO1b	gO1b	-
113	EU686516.1	gO1b	gO1b	-
114	AY326979.1	gO1c	gO1c	-
115	AY582436.1	gO1c	gO1c	-
116	AY582448.1	gO1c	gO1c	-
117	AY582449.1	gO1c	gO1c	-
118	AY582450.1	gO1c	gO1c	-
119	AY582451.1	gO1c	gO1c	-
120	EU686479.1	gO1c	gO1c	-
121	EU686480.1	gO1c	gO1c	-
122	AY326981.1	gO2a	gO2a	-
123	EU686481.1	gO2a	gO2a	-
124	EU686482.1	gO2a	gO2a	-
125	EU686510.1	gO2a	gO2a	-
126	AY326962.1	gO2b	gO2b	-
127	AY326971.1	gO2b	gO2b	-
128	AY326972.1	gO2b	gO2b	-
129	AY326978.1	gO2b	gO2b	-
130	AY582447.1	gO2b	gO2b	-
131	AY582480.1	gO2b	gO2b	-
132	AY582481.1	gO2b	gO2b	-
133	EU686483.1	gO2b	gO2b	-
134	EU686484.1	gO2b	gO2b	-
135	EU686485.1	gO2b	gO2b	-

	GenBank Accession	gO genotype		
		N-terminus tag	mid-N-terminus tag	N-terminus tag
136	EU686486.1	gO2b	gO2b	-
137	EU686517.1	gO2b	gO2b	-
138	AY582437.1	gO3	gO3	-
139	AY582438.1	gO3	gO3	-
140	AY582461.1	gO3	gO3	-
141	AY582464.1	gO3	gO3	-
142	AY582465.1	gO3	gO3	-
143	AY582466.1	gO3	gO3	-
144	AY582467.1	gO3	gO3	-
145	AY582468.1	gO3	gO3	-
146	AY582482.1	gO3	gO3	-
147	EU686487.1	gO3	gO3	-
148	EU686488.1	gO3	gO3	-
149	EU686489.1	gO3	gO3	-
150	EU686490.1	gO3	gO3	-
151	EU686491.1	gO3	gO3	-
152	EU686492.1	gO3	gO3	-
153	EU686493.1	gO3	gO3	-
154	EU686494.1	gO3	gO3	-
155	EU686495.1	gO3	gO3	-
156	EU686496.1	gO3	gO3	-
157	AY326967.1	gO4	gO4	-
158	AY326969.1	gO4	gO4	-
159	AY326970.1	gO4	gO4	-
160	EU686497.1	gO4	gO4	-
161	EU686498.1	gO4	gO4	-
162	EU686499.1	gO4	gO4	-
163	EU686500.1	gO4	gO4	-
164	AY582453.1	gO5	gO5	-
165	AY582454.1	gO5	gO5	-
166	AY582456.1	gO5	gO5	-
167	AY582457.1	gO5	gO5	-
168	AY582458.1	gO5	gO5	-
169	AY582459.1	gO5	gO5	-
170	AY582460.1	gO5	gO5	-
171	AY582463.1	gO5	gO5	-
172	AY582475.1	gO5	gO5	-
173	AY582477.1	gO5	gO5	-
174	AY582479.1	gO5	gO5	-
175	EU686501.1	gO5	gO5	-
176	EU686502.1	gO5	gO5	-
177	EU686503.1	gO5	gO5	-
178	EU686504.1	gO5	gO5	-
179	EU686505.1	gO5	gO5	-
180	EU686506.1	gO5	gO5	-

	GenBank Accession	gO genotype		
		N-terminus tag	mid-N-terminus tag	N-terminus tag
181	EU686518.1	gO5	gO5	-
182	AY326977.1	gO1a	-	-
183	AY326965.1	gO1b	-	-
184	AY326976.1	gO2a	-	-
185	AY326961.1	gO4	-	-
186	KT987996.1	-	gO1a	-
187	EU377488.1	-	gO1c	-
188	EU377490.1	-	gO1c	-
189	KT987997.1	-	gO1c	-
190	KT987998.1	-	gO2b	-
191	KT987999.1	-	gO2b	-
192	KT988002.1	-	gO4	-
193	KT988003.1	-	gO4	-
194	KT988000.1*	-	-	-
195	KT988001.1*	-	-	-

Table 5.3. HCMV gO genotype identification in HCMV sequences available in GenBank using molecular tags. HCMV gO genotypes were correctly assigned to 193 partial CDS and complete CDS sequences available in GenBank release 211. The sequences were screened with tags at three loci (2 in the N-terminus and one in the C-terminus) within the gO gene. *KT988000.1 and KT988001.1 are discussed in section 5.1.2 ‘Unique Cases’.

	GenBank Accession No.	Genotype
1	AF224680.1	gN1
2	AF309969.1	gN1
3	AF309970.1	gN1
4	AF309971.1	gN1
5	AF309972.1	gN1
6	AF309973.1	gN1
7	AF309974.1	gN1
8	AF390755.1	gN1
9	AF390757.1	gN1
10	AF390758.1	gN1
11	AF390760.1	gN1
12	AF390766.1	gN1
13	AF390774.1	gN1
14	AF390775.1	gN1
15	AF390776.1	gN1
16	AF390778.1	gN1
17	AF390781.1	gN1
18	AF390783.1	gN1
19	AF390784.1	gN1
20	AF390786.1	gN1
21	AF390787.1	gN1
22	AF390788.1	gN1
23	AF390789.1	gN1
24	AF390796.1	gN1
25	AF390799.1	gN1
26	AF390804.1	gN1
27	AF390806.1	gN1
28	AF390817.1	gN1
29	AF390818.1	gN1
30	AF390826.1	gN1
31	AF390827.1	gN1
32	AF390829.1	gN1
33	AF390830.1	gN1
34	AF390831.1	gN1
35	AF390843.2	gN1
36	AF390844.2	gN1
37	AF390845.2	gN1
38	AF390846.2	gN1
39	AF390847.2	gN1
40	AF390851.1	gN1
41	AF390852.1	gN1
42	AF390854.1	gN1
43	AF395121.1	gN1
44	AF396713.1	gN1
45	AF396714.1	gN1
46	AF396715.1	gN1
47	AF396716.1	gN1
48	AF396723.1	gN1
49	AF396725.1	gN1

	GenBank Accession No.	Genotype
50	AF396726.1	gN1
51	AF396735.1	gN1
52	AF396738.1	gN1
53	AY326987.1	gN1
54	AY326988.1	gN1
55	AY326989.1	gN1
56	AY326990.1	gN1
57	AY326991.1	gN1
58	AY326992.1	gN1
59	AY326993.1	gN1
60	AY326998.1	gN1
61	AY327000.1	gN1
62	AY327001.1	gN1
63	AY327002.1	gN1
64	AY327003.1	gN1
65	AY327004.1	gN1
66	AY327005.1	gN1
67	AY327006.1	gN1
68	AY327007.1	gN1
69	AY327010.1	gN1
70	AY327014.1	gN1
71	AY327016.1	gN1
72	AY327017.1	gN1
73	AY327018.1	gN1
74	AY327019.1	gN1
75	AY327020.1	gN1
76	AY327024.1	gN1
77	AY327025.1	gN1
78	AY327026.1	gN1
79	AY327029.1	gN1
80	AY327030.1	gN1
81	AY327031.1	gN1
82	EU377503.1	gN1
83	EU686416.1	gN1
84	EU686417.1	gN1
85	EU686418.1	gN1
86	EU686419.1	gN1
87	EU686451.1	gN1
88	EU686454.1	gN1
89	EU686458.1	gN1
90	EU686459.1	gN1
91	EU686460.1	gN1
92	EU686461.1	gN1
93	EU686462.1	gN1
94	GQ227785.1	gN1
95	GU441773.1	gN1
96	GU583628.1	gN1
97	GU583629.1	gN1
98	GU583630.1	gN1

	GenBank Accession No.	Genotype
99	GU583631.1	gN1
100	GU583642.1	gN1
101	GU583646.1	gN1
102	GU583647.1	gN1
103	KF875976.1	gN1
104	KF875977.1	gN1
105	KR992948.1	gN1
106	KR992949.1	gN1
107	KR992950.1	gN1
108	KR992952.1	gN1
109	KR992954.1	gN1
110	KR992956.1	gN1
111	KR992967.1	gN1
112	KR992968.1	gN1
113	KR992970.1	gN1
114	KR992983.1	gN1
115	KR992990.1	gN1
116	KR992993.1	gN1
117	KR992994.1	gN1
118	KR992996.1	gN1
119	KR993000.1	gN1
120	KR993007.1	gN1
121	KR993008.1	gN1
122	KR993019.1	gN1
123	KR993025.1	gN1
124	KR993027.1	gN1
125	KR993035.1	gN1
126	KR993041.1	gN1
127	KR993059.1	gN1
128	KT987979.1	gN1
129	KT987980.1	gN1
130	AF309975.1	gN2
131	AF309976.1	gN2
132	AF309977.1	gN2
133	AF309978.1	gN2
134	AF309979.1	gN2
135	EU686420.1	gN2
136	EU686421.1	gN2
137	EU686422.1	gN2
138	GQ227753.1	gN2
139	GQ227761.1	gN2
140	GU376725.1	gN2
141	KF875978.1	gN2
142	KF875979.1	gN2
143	KR992958.1	gN2
144	KR992972.1	gN2
145	KR992976.1	gN2
146	KR992995.1	gN2
147	KR992998.1	gN2
148	KR993002.1	gN2
149	KR993006.1	gN2
150	KR993012.1	gN2
151	KR993036.1	gN2
152	KR993042.1	gN2
153	KR993043.1	gN2
154	KR993046.1	gN2
155	KR993050.1	gN2
156	KR993052.1	gN2
157	KR993056.1	gN2
158	KR993057.1	gN2
159	AF309980.1	gN3a
160	AF309981.1	gN3a
161	AF309982.1	gN3a
162	AF309983.1	gN3a
163	AF309984.1	gN3a
164	AF309985.1	gN3a
165	AF309986.1	gN3a
166	AF390794.1	gN3a
167	AF390825.1	gN3a
168	AF390841.2	gN3a
169	AF390842.2	gN3a
170	AF390856.1	gN3a
171	AF396717.1	gN3a
172	AF396718.1	gN3a
173	AF396719.1	gN3a
174	AF396720.1	gN3a
175	AF396732.1	gN3a
176	AF396734.1	gN3a
177	AF396745.1	gN3a
178	AY326994.1	gN3a
179	AY326995.1	gN3a
180	AY326996.1	gN3a
181	AY326997.1	gN3a
182	AY326999.1	gN3a
183	AY327008.1	gN3a
184	AY327009.1	gN3a
185	AY327015.1	gN3a
186	AY327027.1	gN3a
187	AY327028.1	gN3a
188	AY327032.1	gN3a
189	EU686423.1	gN3a
190	EU686424.1	gN3a
191	EU686425.1	gN3a
192	EU686426.1	gN3a
193	EU686427.1	gN3a
194	EU686428.1	gN3a
195	EU686429.1	gN3a
196	EU686446.1	gN3a

	GenBank Accession No.	Genotype
197	EU686447.1	gN3a
198	EU686448.1	gN3a
199	EU686449.1	gN3a
200	EU686450.1	gN3a
201	EU686452.1	gN3a
202	EU686455.1	gN3a
203	EU686457.1	gN3a
204	GQ227786.1	gN3a
205	GU376726.1	gN3a
206	GU583636.1	gN3a
207	KR992951.1	gN3a
208	KR992960.1	gN3a
209	KR992966.1	gN3a
210	KR992975.1	gN3a
211	KR992977.1	gN3a
212	KR992979.1	gN3a
213	KR992980.1	gN3a
214	KR992987.1	gN3a
215	KR992991.1	gN3a
216	KR993005.1	gN3a
217	KR993021.1	gN3a
218	KR993031.1	gN3a
219	KR993033.1	gN3a
220	KR993038.1	gN3a
221	KR993045.1	gN3a
222	KR993053.1	gN3a
223	KR993058.1	gN3a
224	KR993060.1	gN3a
225	KT987981.1	gN3a
226	KT987982.1	gN3a
227	AF390770.1	gN3b
228	AF390773.1	gN3b
229	AF390802.1	gN3b
230	AF390812.1	gN3b
231	AF390823.1	gN3b
232	AF390835.2	gN3b
233	AY327013.1	gN3b
234	AY327021.1	gN3b
235	EU377506.1	gN3b
236	EU686430.1	gN3b
237	GU376720.1	gN3b
238	KF875980.1	gN3b
239	KF875981.1	gN3b
240	KR992971.1	gN3b
241	KR992974.1	gN3b
242	KR992984.1	gN3b
243	KR992997.1	gN3b
244	KR992999.1	gN3b
245	KR993010.1	gN3b
246	KR993011.1	gN3b
247	KR993020.1	gN3b
248	KR993023.1	gN3b
249	KR993029.1	gN3b
250	KR993030.1	gN3b
251	KR993040.1	gN3b
252	KR993047.1	gN3b
253	KR993048.1	gN3b
254	KR993049.1	gN3b
255	KT987983.1	gN3b
256	AF224679.1	gN4a
257	AF309987.1	gN4a
258	AF309988.1	gN4a
259	AF309989.1	gN4a
260	AF309990.1	gN4a
261	AF309991.1	gN4a
262	AF309992.1	gN4a
263	AF309993.1	gN4a
264	AF309994.1	gN4a
265	AF390752.1	gN4a
266	AF390753.1	gN4a
267	AF390762.1	gN4a
268	AF390767.1	gN4a
269	AF390768.1	gN4a
270	AF390769.1	gN4a
271	AF390779.1	gN4a
272	AF390782.1	gN4a
273	AF390785.1	gN4a
274	AF390791.1	gN4a
275	AF390792.1	gN4a
276	AF390793.1	gN4a
277	AF390795.1	gN4a
278	AF390801.1	gN4a
279	AF390813.1	gN4a
280	AF390815.1	gN4a
281	AF390819.1	gN4a
282	AF390821.1	gN4a
283	AF390836.2	gN4a
284	AF390837.2	gN4a
285	AF390848.1	gN4a
286	AF390849.1	gN4a
287	AF390855.1	gN4a
288	AF396721.1	gN4a
289	AF396722.1	gN4a
290	AF396724.1	gN4a
291	AF396727.1	gN4a
292	AF396730.1	gN4a
293	AF396731.1	gN4a
294	AF396739.1	gN4a

	GenBank Accession No.	Genotype
295	AF396741.1	gN4a
296	AF396743.1	gN4a
297	EU377505.1	gN4a
298	EU377509.1	gN4a
299	EU686431.1	gN4a
300	EU686432.1	gN4a
301	EU686453.1	gN4a
302	GQ227765.1	gN4a
303	GQ227770.1	gN4a
304	GU376721.1	gN4a
305	GU376722.1	gN4a
306	GU583637.1	gN4a
307	GU583644.1	gN4a
308	KR992957.1	gN4a
309	KR992959.1	gN4a
310	KR992962.1	gN4a
311	KR992981.1	gN4a
312	KR992982.1	gN4a
313	KR992985.1	gN4a
314	KR992988.1	gN4a
315	KR993003.1	gN4a
316	KR993013.1	gN4a
317	KR993015.1	gN4a
318	KR993016.1	gN4a
319	KR993018.1	gN4a
320	KR993024.1	gN4a
321	KR993032.1	gN4a
322	KR993051.1	gN4a
323	KR993054.1	gN4a
324	KR993055.1	gN4a
325	KT987984.1	gN4a
326	KT987985.1	gN4a
327	AF309995.1	gN4b
328	AF309996.1	gN4b
329	AF309997.1	gN4b
330	AF309998.1	gN4b
331	AF309999.1	gN4b
332	AF390749.1	gN4b
333	AF390750.1	gN4b
334	AF390751.1	gN4b
335	AF390754.1	gN4b
336	AF390756.1	gN4b
337	AF390765.1	gN4b
338	AF390780.1	gN4b
339	AF390797.1	gN4b
340	AF390810.1	gN4b
341	AF390816.1	gN4b
342	AF390820.1	gN4b
343	AF390832.1	gN4b
344	AF390834.1	gN4b
345	AF390850.1	gN4b
346	AF390857.1	gN4b
347	AF396733.1	gN4b
348	AF396736.1	gN4b
349	EU377510.1	gN4b
350	EU686433.1	gN4b
351	EU686434.1	gN4b
352	EU686435.1	gN4b
353	GQ227755.1	gN4b
354	GQ227760.1	gN4b
355	GU376723.1	gN4b
356	GU583632.1	gN4b
357	GU583634.1	gN4b
358	GU583638.1	gN4b
359	GU583640.1	gN4b
360	GU583641.1	gN4b
361	GU583645.1	gN4b
362	KR261654.1	gN4b
363	KR992955.1	gN4b
364	KR992963.1	gN4b
365	KR992964.1	gN4b
366	KR992989.1	gN4b
367	KR993014.1	gN4b
368	KR993022.1	gN4b
369	KR993026.1	gN4b
370	KR993028.1	gN4b
371	KR993039.1	gN4b
372	KT987986.1	gN4b
373	KT987987.1	gN4b
374	AF390748.1	gN4c
375	AF390761.1	gN4c
376	AF390790.1	gN4c
377	AF390805.1	gN4c
378	AF390814.1	gN4c
379	AF390838.2	gN4c
380	AF395119.1	gN4c
381	AF395120.1	gN4c
382	AF396728.1	gN4c
383	AF396737.1	gN4c
384	EU686436.1	gN4c
385	EU686437.1	gN4c
386	GQ227762.1	gN4c
387	GU647095.1	gN4c
388	KR992965.1	gN4c
389	KR992973.1	gN4c
390	KR992992.1	gN4c
391	KR993001.1	gN4c
392	KR993004.1	gN4c

	GenBank Accession No.	Genotype
393	KR993009.1	gN4c
394	KR993044.1	gN4c
395	KT987988.1	gN4c
396	KT987989.1	gN4c
397	AF224681.1	gN4d
398	AF224682.1	gN4d
399	AF224683.1	gN4d
400	AF224684.1	gN4d
401	AF224685.1	gN4d
402	AF224686.1	gN4d
403	AF224687.1	gN4d
404	AF310000.1	gN4d
405	AF310001.1	gN4d
406	AF310002.1	gN4d
407	AF310003.1	gN4d
408	AF310004.1	gN4d
409	AF310005.1	gN4d
410	AF310006.1	gN4d
411	AF390759.1	gN4d
412	AF390763.1	gN4d
413	AF390764.1	gN4d
414	AF390771.1	gN4d
415	AF390772.1	gN4d
416	AF390798.1	gN4d
417	AF390800.1	gN4d
418	AF390803.1	gN4d
419	AF390807.1	gN4d
420	AF390808.1	gN4d
421	AF390809.1	gN4d
422	AF390811.1	gN4d
423	AF390822.1	gN4d
424	AF390824.1	gN4d
425	AF390828.1	gN4d
426	AF390833.1	gN4d
427	AF390840.2	gN4d
428	AF395118.1	gN4d
429	AF395122.1	gN4d
430	AF396729.1	gN4d
431	AF396740.1	gN4d
432	AF396742.1	gN4d
433	AF396744.1	gN4d
434	EU686438.1	gN4d
435	EU686439.1	gN4d
436	EU686440.1	gN4d
437	EU686441.1	gN4d
438	EU686442.1	gN4d
439	EU686443.1	gN4d
440	EU686444.1	gN4d
441	EU686456.1	gN4d

	GenBank Accession No.	Genotype
442	GQ227780.1	gN4d
443	GU376724.1	gN4d
444	GU376727.1	gN4d
445	GU583633.1	gN4d
446	GU583635.1	gN4d
447	GU583643.1	gN4d
448	KF875982.1	gN4d
449	KF875983.1	gN4d
450	KR992953.1	gN4d
451	KR992969.1	gN4d
452	KR992978.1	gN4d
453	KR992986.1	gN4d
454	KR993017.1	gN4d
455	KR993034.1	gN4d
456	KR993037.1	gN4d
457	AF224688.1*	-
458	AF390777.1*	-
459	AF390853.1*	-
460	AY327011.1*	-
461	AY327022.1*	-
462	AY327023.1*	-
463	EU686445.1*	-
464	GU583639.1*	-
465	KR992961.1*	-

Table 5.4. HCMV gN genotype identification using molecular tags. HCMV gN genotypes were correctly assigned to 456 out of 465 partial or complete CDSs available in GenBank release 211. * Nine sequences could not be assigned a genotype by our tags for various reasons discussed in section 5.2.3 ‘Unique Cases’.

5.2.3 Unique Cases

1. KP745697.1 (HCMV strain BE/23/2010) – potential recombinant

This sequence is from urine-derived virus passaged four times in human fibroblasts. UL1 and UL9 are mutated (https://www.ncbi.nlm.nih.gov/nuccore/822901463). The BE/23/2010 gO sequence shares similarities with both gO1c and gO1a. Amino acid sequence alignment (figure 5.9) shows that the first 140 amino acid residues are identical to gO1c reference strain Toledo (GU937742.1) while the remaining 324 residues are identical to gO1a reference strain AD169 (FJ527563.1). However, at nucleotide level the similarity with gO1a is earlier at base 294 [codon 98] (figure 5.10). These observations suggest that strain BE/23/2010 is a gO1a/gO1c recombinant or a compilation error from mixed infection.

```
FJ527563.1_AD169 [gO1a]   MGRKEMMVRDVPKMVFLISISFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL  60
GU937742.1_Toledo [gO1c]  MGRKGD-MRSISKLFFIISLTVLLFSIINCKVVRP--PGRYWLGTVLSTIGKQKLDKFKL  57
KP745697.1_BE/23/2010     MGRKGD-MRSISKLFFIISLTVLLFSIINCKVVRP--PGRYWLGTVLSTIGKQKLDKFKL  57
                          ****   :*.: *:.*:**::.**.*:*****:      * * * ***.*** ***::**

FJ527563.1_AD169 [gO1a]   EILRQLETTISTKY-NVSKQPVKNLTMNMTEFPQYYILAGPIQNYSITYLWFDFYSTQLR  119
GU937742.1_Toledo [gO1c]  EILKQLEREPYTKYFNMTRQHVKNLTMNMTQFPQYYILAGPIRNDSITYLWFDFYSTQLR  117
KP745697.1_BE/23/2010     EILKQLEREPYTKYFNMTRQHVKNLTMNMTQFPQYYILAGPIRNDSITYLWFDFYSTQLR  117
                          ***:***    *** *:::* *********:***********:* ***************

FJ527563.1_AD169 [gO1a]   KPAKYVYSQYNHTAKTITFRPPPCGTVPSMTCLSEMLNVSKRNDTGEQGCGNFTTFNPMF  179
GU937742.1_Toledo [gO1c]  KPAKYVYSQYNHTAKTITFRPPSCGTVPSMTCLSEMLNVSKRNDTGEQGCGNFTTFNPMF  177
KP745697.1_BE/23/2010     KPAKYVYSQYNHTAKTITFRPPPCGTVPSMTCLSEMLNVSKRNDTGEQGCGNFTTFNPMF  177
                          ********************** *************************************

FJ527563.1_AD169 [gO1a]   FNVPRWNTKLYVGPTKVNVDSQTIYFLGLTALLLRYAQRNCTHSFYLVNAMSRNLFRVPK  239
GU937742.1_Toledo [gO1c]  FNVPRWNTKLYVGPTKVNVDSQTIYFLGLTALLLRYAQRNCTHSFYLVNAMSRNLFRVPK  237
KP745697.1_BE/23/2010     FNVPRWNTKLYVGPTKVNVDSQTIYFLGLTALLLRYAQRNCTHSFYLVNAMSRNLFRVPK  237
                          ************************************************************

FJ527563.1_AD169 [gO1a]   YINGTKLKNTMRKLKRKQAPVKEQFEKKAKKTQSTTTPYFSYTTSAALNVTTNVTYSITT  299
GU937742.1_Toledo [gO1c]  YINGTKLKNTMRKLKRKQAPVKEQLEKKTKKSQSTTTPYFSYTTSTALNVTTNATYKVTT  297
KP745697.1_BE/23/2010     YINGTKLKNTMRKLKRKQAPVKEQFEKKAKKTQSTTTPYFSYTTSAALNVTTNVTYSITT  297
                          ************************:***:**:*************:*******.**.:**

FJ527563.1_AD169 [gO1a]   AARRVSTSTIAYRPDSSFMKSIMATQLRDLATWVYTTLRYRQNPFCEPSRNRTAVSEFMK  359
GU937742.1_Toledo [gO1c]  SAKRIPTSTIAYRPDSSFMKSIMATQLRDLATWVYTTLRYRNEPFCKPDRNRTAVSEFMK  357
KP745697.1_BE/23/2010     AARRVSTSTIAYRPDSSFMKSIMATQLRDLATWVYTTLRYRQNPFCEPSRNRTAVSEFMK  357
                          :*:*: ***********************************::***:*.***********

FJ527563.1_AD169 [gO1a]   NTHVLIRNETPYTIYGTLDMSSLYYNETMFVENKTASDSNKTTPTSPSMGFQRTFIDPLW  419
GU937742.1_Toledo [gO1c]  NTHVLIRNETPYTIYGTLDMSSLYYNETMSVENETASDNNETTPTSPSTRFQKTFIDPLW  417
KP745697.1_BE/23/2010     NTHVLIRNETPYTIYGTLDMSSLYYNETMFVENKTASDSNKTTPTSPSMGFQRTFIDPLW  417
                          ***************************** ***:****.*:*******  **:*******

FJ527563.1_AD169 [gO1a]   DYLDSLLFLDEIRNFSLRSPTYVNLTPPEHRRAVNLSTLNSLWWWLQ    466
GU937742.1_Toledo [gO1c]  DYLDSLLFLDKIRNFSLQLPAYGNLTPPEHRRAVNLSTLNSLWWWSQ    464
KP745697.1_BE/23/2010     DYLDSLLFLDEIRNFSLRSPTYVNLTPPEHRRAVNLSTLNSLWWWLQ    464
                          **********:******: *:* ********************** *
```

Figure 5.9. Full-length HCMV gO amino acid sequence alignment of BE/23/2010, AD169 (gO1a reference strain) and Toledo (gO1c reference strain). The red box indicates the position where sequence similarity of BE/23/2010 switches from gO1c to gO1a (residue 140). Alignment was done using CLUSTAL Omega version 1.2.2.

```
FJ527563.1_AD169.[gO1a]  ATGGGGAGAAAAGAGATGATGGTGAGAGACGTCCCTAAGATGGTGTTTCTAATATCTATA  60
GU937742.1_Toledo.[gO1c] ATGGGGAGAAAA---GGAGACATGAGAAGCATTTCTAAATTATTCTTTATTATATCACTG  57
KP745697.1_BE/23/2010.   ATGGGGAGAAAA---GGAGACATGAGAAGCATTTCTAAATTATTCTTTATTATATCACTA  57
                         ************          *****  * *  ****  *  * *** * *****  *

FJ527563.1_AD169.[gO1a]  TCTTTCTTGCTTGTTTCTTTCATAAACTGTAAAGTTATGTCAAAAGCGCTTTATAATCGT  120
GU937742.1_Toledo.[gO1c] ACTGTCCTGTTATTTTCTATAATAAACTGTAAGGTCGTCAGACCACC------GGGACGT  111
KP745697.1_BE/23/2010.   ACTGTCCTGTTATTTTCTATAATAAACTGTAAGGTCGTCAGACCACC------GGGACGT  111
                          ** ** ** *  ***** * *********** **  *   *  * *          ***

FJ527563.1_AD169.[gO1a]  CCTTGGAGGGGCTTGGTACTGTCTAAGATAGGCAAATATAAATTAGATCAGCTTAAGTTA  180
GU937742.1_Toledo.[gO1c] TACTGGTTAGGTACAGTACTTTCTACGATAGGCAAGCAAAAACTAGATAAATTCAAGTTA  171
KP745697.1_BE/23/2010.   TACTGGTTAGGTACAGTACTTTCTACGATAGGCAAGCAAAAACTAGATAAATTCAAGTTA  171
                            ***   **    ***** **** *********  * *** ***** *  * ******

FJ527563.1_AD169.[gO1a]  GAAATTTTGAGACAACTAGAAACGACTATTTCTA---CAAAATACAATGTAAGTAAACAA  237
GU937742.1_Toledo.[gO1c] GAGATTTTAAAACAATTAGAAAGGGAACCTTATACAAAATACTTCAATATGACTAGACAA  231
KP745697.1_BE/23/2010.   GAGATTTTAAAACAATTAGAAAGGGAACCTTATACAAAATACTTCAATATGACTAGGCAA  231
                         ** ***** * **** ****** *     ** **    * * * **** * * **  ***

FJ527563.1_AD169.[gO1a]  CCGGTTAAAAATCTCACTATGAACATGACAGAGTTTCCACAATACTACATTTTAGCGGGC  297
GU937742.1_Toledo.[gO1c] CACGTTAAAAATCTTACTATGAATATGACCCAGTTTCCACAATACTACATTCTAGCAGGT  291
KP745697.1_BE/23/2010.   CACGTTAAAAATCTTACTATGAATATGACCCAGTTTCCACAATACTACATTCTAGCGGGC  291
                         *  *********** ******** *****  ******************** **** **

FJ527563.1_AD169.[gO1a]  CCCATTCAGAATTATAGTATAACCTATCTGTGGTTTGATTTTTATAGTACCCAGCTTAGA  357
GU937742.1_Toledo.[gO1c] CCCATTCGAAACGATAGTATAACCTATCTGTGGTTCGATTTTTATAGTACCCAGCTTAGA  351
KP745697.1_BE/23/2010.   CCCATTCGAAATGATAGTATAACCTATCTGTGGTTCGATTTTTATAGTACCCAGCTTAGA  351
                         *******  **  ********************** ************************

FJ527563.1_AD169.[gO1a]  AAACCCGCAAAATACGTTTACTCACAGTACAATCATACGGCTAAAACGATAACATTCAGA  417
GU937742.1_Toledo.[gO1c] AAACCCGCCAAATACGTCTACTCACAGTACAATCATACGGCTAAAACGATAACATTCAGA  411
KP745697.1_BE/23/2010.   AAACCCGCAAAATACGTCTACTCACAGTACAATCATACGGCTAAAACGATAACATTCAGA  411
                         ******** ******** ******************************************

FJ527563.1_AD169.[gO1a]  CCCCCACCTTGTGGTACTGTGCCTTCCATGACTTGTCTTTCCGAAATGCTAAACGTTTCC  477
GU937742.1_Toledo.[gO1c] CCCCCATCTTGTGGTACAGTGCCTTCAATGACTTGTCTTTCCGAAATGTTAAACGTTTCC  471
KP745697.1_BE/23/2010.   CCCCCACCTTGTGGTACTGTGCCTTCCATGACTTGTCTTTCCGAAATGCTAAACGTTTCC  471
                         ****** ********** ******** ********************* ***********

FJ527563.1_AD169.[gO1a]  AAACGTAATGATACTGGCGAACAAGGTTGCGGTAATTTCACCACGTTCAACCCCATGTTT  537
GU937742.1_Toledo.[gO1c] AAACGTAATGATACTGGCGAACAAGGCTGCGGTAATTTTACCACGTTCAACCCCATGTTT  531
KP745697.1_BE/23/2010.   AAACGTAATGATACTGGCGAACAAGGTTGCGGTAATTTCACCACGTTCAACCCCATGTTT  531
                         ************************** *********** *********************

FJ527563.1_AD169.[gO1a]  TTCAATGTACCGCGTTGGAACACCAAATTGTACGTGGGTCCGACTAAGGTTAACGTAGAT  597
GU937742.1_Toledo.[gO1c] TTCAATGTACCACGTTGGAATACCAAATTGTACGTGGGTCCGACTAAAGTTAACGTAGAT  591
KP745697.1_BE/23/2010.   TTCAATGTACCGCGTTGGAACACCAAATTGTACGTGGGTCCGACTAAGGTTAACGTAGAT  591
                         *********** ******** ************************** ************

FJ527563.1_AD169.[gO1a]  AGTCAAACGATTTATTTTCTAGGTTTAACCGCCCTGCTTTTACGTTACGCACAACGCAAC  657
GU937742.1_Toledo.[gO1c] AGTCAAACGATTTATTTTTTGGGTTTAACCGCCCTACTTCTACGTTACGCACAACGCAAC  651
KP745697.1_BE/23/2010.   AGTCAAACGATTTATTTTCTAGGTTTAACCGCCCTGCTTTTACGTTACGCACAACGCAAC  651
                         ****************** * ************** *** ********************

FJ527563.1_AD169.[gO1a]  TGTACACACAGTTTCTACCTGGTTAACGCCATGAGCCGGAATCTATTTCGCGTCCCCAAG  717
GU937742.1_Toledo.[gO1c] TGCACACACAGTTTCTACCTGGTTAACGCCATGAGCCGGAATCTATTCCGCGTCCCCAAG  711
KP745697.1_BE/23/2010.   TGTACACACAGTTTCTACCTGGTTAACGCCATGAGCCGGAATCTATTTCGCGTCCCCAAG  711
                         ** ******************************************** ************

FJ527563.1_AD169.[gO1a]  TATATTAACGGCACCAAGTTAAAAAACACTATGCGAAAACTAAAACGTAAACAAGCGCCC  777
GU937742.1_Toledo.[gO1c] TATATTAACGGCACCAAGTTGAAAAACACTATGCGAAAACTAAAACGTAAACAAGCGCCA  771
KP745697.1_BE/23/2010.   TATATTAACGGCACCAAGTTAAAAAACACTATGCGAAAACTAAAACGTAAACAAGCGCCC  771
                         ******************** **************************************

FJ527563.1_AD169.[gO1a]  GTTAAGGAACAATTCGAAAAAAAAGCTAAGAAAACTCAGAGTACTACTACGCCATACTTT  837
GU937742.1_Toledo.[gO1c] GTCAAAGAACAATTAGAAAAAAAGACTAAAAAATCTCAGAGTACTACTACGCCATATTTT  831
KP745697.1_BE/23/2010.   GTTAAGGAACAATTCGAAAAAAAGGCTAAGAAAACTCAGAGTACTACTACGCCATACTTT  831
                         ** ** ******** ********  **** *** ********************** ***

FJ527563.1_AD169.[gO1a]  TCCTATACAACGTCTGCCGCTCTCAACGTCACTACTAACGTGACTTATAGTATTACTACC  897
GU937742.1_Toledo.[gO1c] TCCTATACAACGTCTACCGCTCTCAACGTCACTACTAACGCGACTTATAAGGTTACCACC  891
KP745697.1_BE/23/2010.   TCCTATACAACGTCTGCCGCTCTCAACGTCACTACTAACGTGACTTATAGTATTACTACC  891
                         *************** ************************ ********   **** ***

FJ527563.1_AD169.[gO1a]  GCCGCAAGGCGGGTTTCCACGTCTACAATTGCTTATCGTCCTGATAGCAGCTTTATGAAG  957
GU937742.1_Toledo.[gO1c] AGCGCAAAGCGGATTCCCACATCTACGATTGCTTATCGTCCTGATAGCAGCTTTATGAAG  951
KP745697.1_BE/23/2010.   GCCGCAAGGCGGGTTTCCACGTCTACAATTGCTTATCGTCCTGATAGCAGCTTTATGAAG  951
                           ***** **** ** **** ***** *********************************

FJ527563.1_AD169.[gO1a]  TCCATTATGGCCACACAGTTAAGGGACCTAGCAACGTGGGTGTATACCACTCTACGTTAC  1017
GU937742.1_Toledo.[gO1c] TCCATTATGGCCACGCAGTTAAGGGATCTAGCGACATGGGTGTATACTACTCTACGGTAT  1011
KP745697.1_BE/23/2010.   TCCATTATGGCCACACAGTTAAGGGACCTAGCAACGTGGGTGTATACCACTCTACGTTAC  1011
                         ************** *********** ***** ** *********** ******** **
```

```
FJ527563.1_AD169.[gO1a]  CGGCAAAATCCTTTTTGTGAACCAAGCCGCAACCGAACCGCCGTGTCAGAATTTATGAAA  1077
GU937742.1_Toledo.[gO1c] CGGAATGAACCCTTTTGTAAACCAGACCGTAACCGTACCGCCGTGTCAGAATTTATGAAA  1071
KP745697.1_BE/23/2010.   CGGCAAAATCCTTTTTGTGAACCAAGCCGCAACCGAACCGCCGTGTCAGAATTTATGAAA  1071
                         *** *  * ** ****** *****  *** ***** ************************

FJ527563.1_AD169.[gO1a]  AACACGCACGTACTAATCCGTAACGAAACGCCGTACACTATTTACGGTACTCTCGACATG  1137
GU937742.1_Toledo.[gO1c] AACACGCACGTATTGATTCGTAACGAAACACCGTACACTATTTACGGCACTCTTGACATG  1131
KP745697.1_BE/23/2010.   AACACGCACGTACTAATCCGTAACGAAACGCCGTACACTATTTACGGTACTCTTGACATG  1131
                         ************ * ** *********** ***************** ***** ******

FJ527563.1_AD169.[gO1a]  AGCTCCTTATATTACAACGAAACCATGTTCGTGGAAAACAAAACAGCTTCCGATAGTAAC  1197
GU937742.1_Toledo.[gO1c] AGCTCCTTATACTACAACGAAACCATGTCCGTGGAAAACGAAACGGCTTCCGATAATAAC  1191
KP745697.1_BE/23/2010.   AGCTCCTTATATTACAACGAAACCATGTTCGTGGAAAACAAAACAGCTTCCGATAGTAAC  1191
                         *********** **************** ********** **** ********** ****

FJ527563.1_AD169.[gO1a]  AAAACTACACCTACGTCACCATCAATGGGGTTTCAGAGAACATTTATAGATCCCCTGTGG  1257
GU937742.1_Toledo.[gO1c] GAAACTACACCTACGTCACCATCGACGAGGTTTCAGAAAACGTTTATAGATCCCTTATGG  1251
KP745697.1_BE/23/2010.   AAAACTACACCTACGTCACCATCAATGGGGTTTCAGAGAACATTTATAGATCCCCTGTGG  1251
                          ********************** * * ********* *** ************ * ***

FJ527563.1_AD169.[gO1a]  GACTATCTAGACTCGCTGCTGTTTCTAGATGAGATTCGTAACTTTAGCCTCCGGTCACCC  1317
GU937742.1_Toledo.[gO1c] GACTATCTAGACTCGCTGCTGTTTCTAGATAAAATCCGTAACTTTAGCCTCCAATTACCC  1311
KP745697.1_BE/23/2010.   GACTATCTAGACTCGCTGCTGTTTCTAGATGAGATTCGTAACTTTAGCCTCCGGTCACCC  1311
                         ****************************** * ** ****************  * ****

FJ527563.1_AD169.[gO1a]  ACGTATGTAAACCTTACCCCGCCGGAACACCGCCGGGCTGTAAATCTGTCCACCCTCAAT  1377
GU937742.1_Toledo.[gO1c] GCGTATGGAAATCTTACCCCGCCGGAACACCGCCGGGCTGTAAATCTATCCACCCTCAAT  1371
KP745697.1_BE/23/2010.   ACGTATGTAAACCTTACCCCGCCGGAACACCGCCGGGCTGTAAATCTGTCCACCCTCAAT  1371
                          ****** *** *********************************** ************

FJ527563.1_AD169.[gO1a]  AGCCTTTGGTGGTGGTTGCAGTAA    1401
GU937742.1_Toledo.[gO1c] AGCCTTTGGTGGTGGTCGCAGTAA    1395
KP745697.1_BE/23/2010.   AGCCTTTGGTGGTGGTTGCAGTAA    1395
                         **************** *******
```

Figure 5.10. Full-length HCMV gO nucleotide sequence alignment of BE/23/2010, AD169 (gO1a reference strain) and Toledo (gO1c reference strain). The red box indicates the position where sequence similarity of BE/23/2010 switches from gO1c to gO1a (base position 294). The alignment was performed using CLUSTAL Omega version 1.2.2.

2. KJ361950.1 (HCMV strain HAN 11) – potential recombinant

HCMV strain HAN 11 was sequenced from virus isolated from a bronchoalveolar lavage sample and passaged 3 times in human fibroblasts. It contains three mutated genes (RL5A, UL150 and US9) (https://www.ncbi.nlm.nih.gov/nuccore/586833288). HAN 11 has gN of genotype gN3a (figure 5.5) and gO of genotype gO1 (figure 5.2). As previously shown, genotype gN3a forms a linkage group with gO3, while gO1 links with gN1a (Bates et al., 2008). The unusual combination of gN3a and gN1, as seen here for HAN 11, is to our knowledge the only case where the established gN/gO linkage relationship is abrogated, and indicates a recombination point in between these genes, or again the possibility of sequence compilation from a mixed infection.

3. KT988000.1 and KT988001.1 – incomplete sequences, do not include tags

Sequences KT988000.1 (isolate human-wt/ITA/PR206/1995) and KT988001.1 (isolate human-wt/ITA/PR2197/2012) were derived from urine in a cohort of congenitally and postnatally infected children (Arcangeletti et al., 2015). Both sequences are relatively short, each being 267bp in length, spanning base positions 439 to 705 relative to the gO sequence of strain HAN16 (gO3). This region falls outside the three loci targeted by our tags. However phylogenetic tree construction confirms both KT988000.1 and KT988001.1 are genotype gO3 (figure 5.9), as described in the sequence feature notes (http://www.ncbi.nlm.nih.gov/nuccore/KT988000.1 and http://www.ncbi.nlm.nih.gov/nuccore/KT988001.1 respectively).

4. Partial gN CDSs with Unassigned Genotype – possible sequencing errors

Nine partial gN CDSs could not be assigned a genotype by our script and tags (table 5.4). Of these, seven were found to contain substitutions that gave rise to mismatches with our tags; one has unresolved ambiguous bases; and one has a 3bp deletion, as summarised in table 5.5. These could be sequencing errors or variants.

Accession No.	Isolate ID	Reason gN genotype unassigned by Molecular tag
AF224688.1	E16	A to G substitution at position 104 (1st base of gN4d tag) w.r.t reference strain Merlin
AF390777.1	RV	3nt (ACC) deletion at position 115-117 (last 3 bases of gN4d tag) w.r.t reference strain Merlin
AF390853.1	C9	Contains 4 unresolved ambiguous ('N') bases , including at position 110, G to A-substitution at position 106, and A to G-substitution at position 114 w.r.t gN4d reference strain Merlin
AY327011.1	Riu	GCT to ACT and A to C substitutions at positions 120-122 and 126 respectively w.r.t gN3a reference strain TR
AY327022.1	U1r	C to T substitution at position 122 w.r.t gN3a reference strain TR
AY327023.1	U3r	C to T substitution at position 122 w.r.t gN3a reference strain TR
EU686445.1	35M6	C to A substitution at position 124 w.r.t gN3a reference strain TR
GU583639.1	323F	C to T substitution at position 117 w.r.t gN4d reference strain Merlin
KR992961.1	9C	GC to AT substitution at position 104-105 w.r.t gN4c reference strain Toledo

Table 5.5. Partial HCMV gN CDSs that could not be assigned genotypes by our molecular tags. The nine partial HCMV gN CDSs detailed here could not be assigned to a gN genotype for reasons related to the nature of their sequences. w.r.t, with respect to.

5.2.4 Genotype Detection Rate

Overall the tags for gO combined with the script was able to genotype gO 356/358 (99.4%) and gN 619/628 (98.6%) correctly, demonstrating that the tags were robust to genotype and the *perl* Script could be employed to analyse NGS data. Therefore, this method could now be applied to clinical samples and NGS data for quantification of mixed infections and evaluation of infection burden.

5.3 ANALYSIS OF HCMV NGS SEQUENCES FROM BREAST MILK

5.3.1 NGS Data Pre-processing

We sequenced twenty-one DNA samples (17 HIV-positive and 4 HIV-negative) which had sufficient remaining volume for NGS. The DNA was extracted from breast milk collected at weeks (W), W4, W12, and W16 postpartum from 12 women – eight HIV-positive and four HIV-negative. HCMV target amplification using the SureSelect system (Agilent Technologies Inc.), library preparation and Illumina® MiSeq™ paired-end sequencing were performed by our collaborating partners, the Andrew Davison group at the MRC Center for Virus Research, University of Glasgow. Upon obtaining the FASTQ read files (see Methods Section 2.6), we used FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to analyse read quality, and Trimmomatic (Bolger et al., 2014) to trim adaptors and filter the reads. Trimmomatic was used in standard format (phred+33) with the following settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:100. Table 5.7 summarises key outputs from this pre-processing step.

5.3.2 Application of *perl* script to analyse genotype variants

Having validated the molecular tags and *perl* script as described in section 5.2, the next step was to apply these tools to directly analyse FASTQ reads from NGS sequences from our 21 clinical (breast milk) samples collected from 12 individuals – four HIV-negative and eight HIV-positive. Milk samples included four from the HIV-negative and 17 from HIV-positive mothers. Overall, four samples were collected at week (W) 4, five at W12, and ten at W16 postpartum, as detailed in table 5.6. In figures presented in this section, the point of sample collection is designated as a subscript to the sample ID, thus $240R_{12}$.

The primary focus of this analysis was on gO genotypes; however, each sample was also checked for gN genotypes (Table 5.8). Verification was performed by mapping the FASTQ reads to the appropriate reference for each gO genotype and then calculating the read depth using the SAMtools mpileup tool or the GATK DepthofCoverage tool. In all cases, this was consistent with the number of variants detected by our perl Script approach. Although it was possible to detect gO genotypes present to as low as 0.2%, we set a cut-off of 1% and minimum read depth of 10, consistent with previously established thresholds in our laboratory for analyses of Illumina based sequence reads for a related human betaherpesvirus to detect minor variants (Tweedy et al., 2015a, Tweedy et al., 2015b, Tweedy et al., 2016), and as established elsewhere for other genomic systems (Depledge et al., 2011).

	ID	HIV Status	Breast from which milk collected	Sample Collection time-point (postpartum)		
				W4	W12	W16
1	141	Positive	Right			✓
2	158	Negative	Left			✓
3	174	Positive	Left			✓
			Right			✓
4	232	Negative	Right		✓	
5	240	Negative	Right		✓	
6	243	Positive	Left			✓
			Right	✓	✓	✓
7	248	Positive	Right	✓		
8	259	Positive	Left			✓
			Right			✓
9	278	Positive	Left			✓
			Right	✓	✓	✓
10	280	Negative	Right		✓	
11	281	Positive	Right	✓		
12	288	Positive	Right	✓	✓	

Table 5.6. Breast milk samples analysed by NGS. Twenty-one samples collected from 12 individuals were available for NGS analysis.

Figures 5.11 and 5.12 chart the gO genotype proportions in milk from HIV-negative (N=4) and HIV-positive (N=17) mothers respectively, showing vividly that in both groups there were multiple genotypes present. With our cut-offs multiple infections were detected in all but four samples: $234R_{16}$, $243R_{4}$, $243R_{16}$, and $248R_{4}$. For visual clarity and consistency, the eight gO genotypes are colour-coded: gO1a (grey); gO1b (black); gO1c (magenta); gO2a (orange); gO2b (dark red); gO3 (yellow); gO4 (green); and gO5 (blue). Notably, gO2b was not detected among HIV-negative samples.

In figure 5.13, comparison was made between the gO proportions in three HIV-positive and three HIV-negative samples collected at the same time point, week 16. In these six samples between 2 and 7 different genotypes were detected per sample, irrespective of HIV status.

In Figure 5. 14, further comparisons were made, in this case genotype variants in 4 individuals who had samples collected from both the left (L) and right (R) breast at the same time point. All four individuals for whom such sample pairs were available were HIV-positive. There was striking concordance between the gO genotype proportions in samples from the two sides (figure 5.14) indicating similarity in the routes of infection or virus reactivation in this tissue.

ID	HIV Status	Input Read Pairs	Paired Reads Retained		Unpaired Forward Reads Retained		Unpaired Reverse Reads Retained		Total Reads Retained		Total Reads Dropped	
		Number	*Number*	*%*	*Number*	*%*	*Number*	*%*	*Number*	*%*	*Number*	*%*
141R16	Positive	1,146,537	890,615	77.68	178,153	15.54	26,676	2.33	1,095,444	95.55	51,093	4.46
174L16	Positive	916,944	737,347	80.41	118,244	12.9	23,979	2.62	879,570	95.93	37,374	4.08
174R16	Positive	1,083,750	827,817	76.38	186,721	17.23	21,513	1.99	1,036,051	95.6	47,699	4.4
243L16	Positive	1,132,222	847,209	74.83	174,948	15.45	25,436	2.25	1,047,593	92.53	84,629	7.47
243R4	Positive	1,473,048	1,006,620	68.34	327,497	22.23	24,086	1.64	1,358,203	92.21	114,845	7.8
243R12	Positive	1,400,640	1,145,618	81.79	133,375	9.52	43,541	3.11	1,322,534	94.42	78,106	5.58
243R16	Positive	886,331	685,668	77.36	136,340	15.38	21,640	2.44	843,648	95.18	42,683	4.82
248R4	Positive	854,330	657,792	77.0	138,720	16.24	17,525	2.05	814,037	95.29	40,293	4.72
259L16	Positive	1,620,030	1,067,874	65.92	401,160	24.76	23,079	1.42	1,492,113	92.1	127,917	7.9
259R16	Positive	1,451,513	909,158	62.64	383,704	26.43	23,133	1.59	1,315,995	90.66	135,518	9.34
278L16	Positive	1,480,282	975,707	65.91	365,136	24.67	20,927	1.41	1,361,770	91.99	118,512	8.01
278R4	Positive	1,752,804	1,208,782	68.96	384,070	21.91	29,648	1.69	1,622,500	92.56	130,304	7.43
278R12	Positive	1,947,062	1,685,104	86.55	126,845	6.51	59,910	3.08	1,871,859	96.14	75,203	3.86
278R16	Positive	1,483,322	949,871	64.04	389,601	26.27	19,995	1.35	1,359,467	91.66	123,855	8.35
281R4	Positive	2,718,523	2,020,988	74.34	449,676	16.54	67,925	2.5	2,538,589	93.38	179,934	6.62
288R4	Positive	1,486,816	992,768	66.77	346,684	23.32	27,013	1.82	1,366,465	91.91	120,351	8.09
288R12	Positive	1,521,583	1,263,945	83.07	145,835	9.58	40,457	2.66	1,450,237	95.31	71,346	4.69
158L16	Negative	885,658	675,776	76.3	132,258	14.93	28,493	3.22	836,527	94.45	49,131	5.55
232R12	Negative	1,732,535	1,409,192	81.34	169,036	9.76	55,011	3.18	1,633,239	94.28	99,296	5.73
240R12	Negative	1,499,202	1,256,577	83.82	127,782	8.52	44,586	2.97	1,428,945	95.31	70,257	4.69
280R12	Negative	1,561,860	1,305,644	83.6	135,680	8.69	45,835	2.93	1,487,159	95.22	74,701	4.78

Table 5.7. Summary of FASTQ reads pre-processing for DNA derived from 21 breast milk samples. Trimmomatic (Bolger et al., 2014) in standard format (phred+33) was used to filter the reads. Only reads with a minimum length of 100 bases were retained for further analysis.

ID	HIV Status	gO genotype proportion (%)								gN genotype proportion (%)							
		gO1a	gO1b	gO1c	gO2a	gO2b	gO3	gO4	gO5	gN1	gN2	gN3a	gN3b	gN4a	gN4b	gN4c	gN4d
141R16	Positive		48.9		1.3	48.8		1.0			25.0	75.0					
158L16	Negative	2.4	12.6	6.9	6.9		60.1	4.8	6.2					100.0			
174L16	Positive					26.4	68.1	5.5		5.1	33.9	3.8		54.1	3.1		
174R16	Positive		2.9			31.8	63.6	1.6		1.7	14.8	1.4		82.1			
232R12	Negative		62.8						37.2			57.9					42.1
240R12	Negative	5.5	3.6	15.9	1.8		17.4	13.4	42.4	4.1				3.1	9.3		83.5
243L16	Positive							100				6.5			93.5		
243R4	Positive							100							100.0		
243R12	Positive		4.3				1.6	94.1				2.7			97.3		
243R16	Positive							100							100.0		
248R4	Positive				100.0								100.0				
259L16	Positive	8.8				38.1		53.1		21.0	36.9				42.1		
259R16	Positive	9.5				23.2		67.3		14.9	22.8				62.3		
278L16	Positive		88.5						11.5			79.2					20.8
278R4	Positive		76.0	0.8	0.7		1.3	0.8	20.3			71.6					28.4
278R12	Positive	9.0	9.7	41.2	1.5		21.8	14.8	2.0								
278R16	Positive		88.1						11.9			84.8					15.2
280R12	Negative	58.5	2.3	16.1	1.2		14.3	7.6		100.0							
281R4	Positive	0.7	11.8	1.5	1.4		4.0	2.0	78.5			4.9					95.1
288R4	Positive	3.4	2.0	1.0	0.8		2.1	9.2	81.5	13.0				4.6		8.2	74.2
288R12	Positive	9.4	10.4	29.3	8.1		26.4	12.9	3.6				100.0				

Table 5.8. Confirmation of HCMV gO and gN genotype linkage in breast milk samples analysed by NGS. The table summarises the proportions of gO and gN variants analysed using our tags and perl script. Highlighted in grey are the dominant variants, showing correlation between the linked gN/gO genotype.

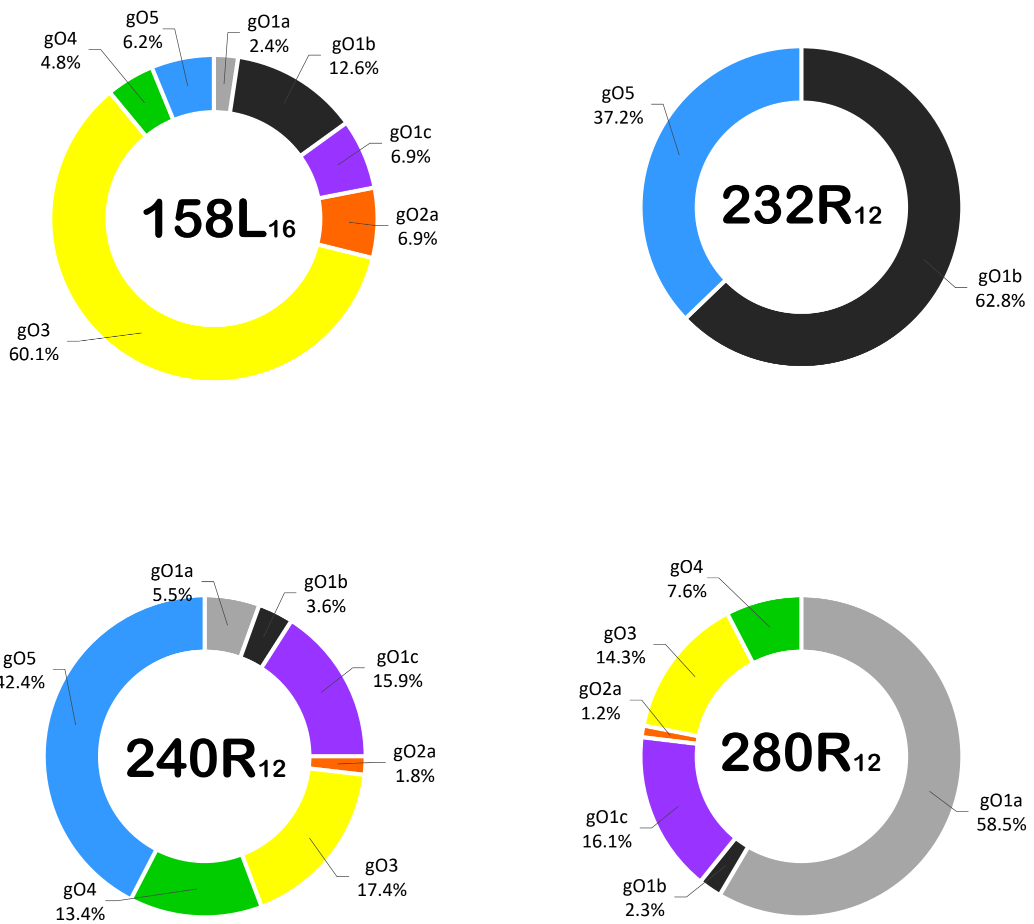
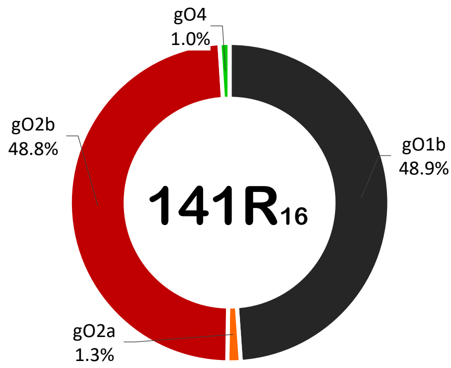
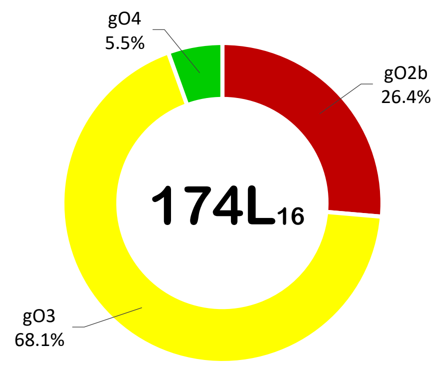
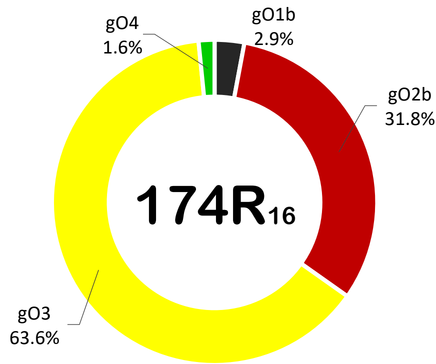
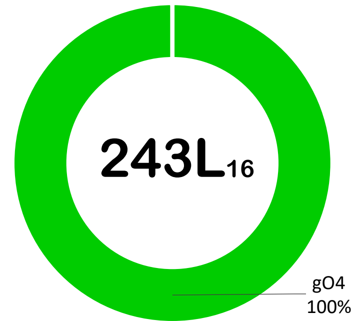
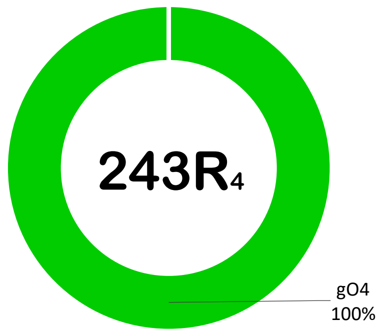
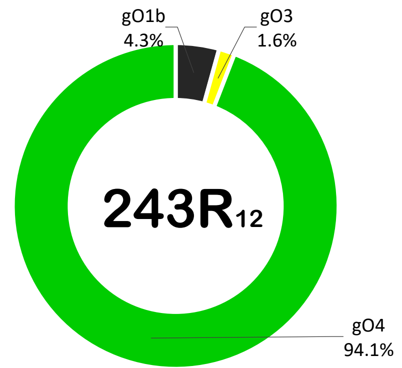
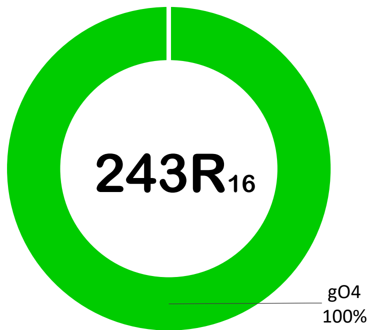
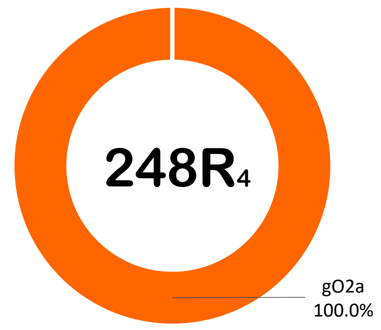
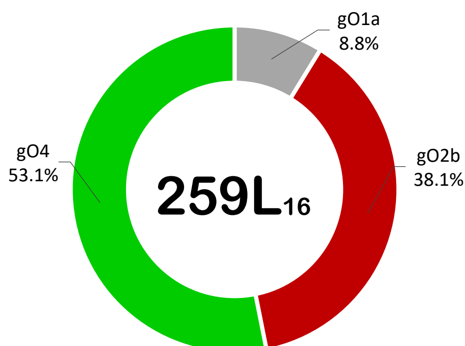
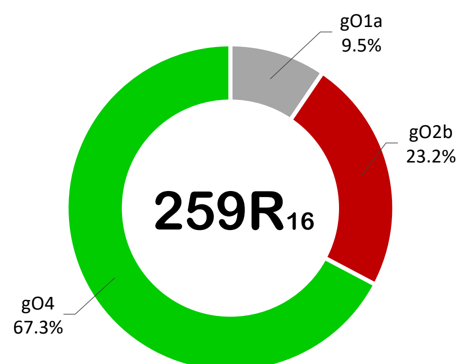


Figure 5.11. HCMV gO genotype proportions in four breast milk samples from HIV-negative mothers. Sample IDs are indicated in the centre of each doughnut, with the subscript indicating the time-point in weeks postpartum when the milk sample was collected. Genotypes are colour-coded as follows: gO1a – grey; gO1b – black; gO1c – magenta; gO2a – orange; gO2b – dark red; gO3 – yellow; gO4 – green; gO5 – blue.


gO4
1.0%
gO2b
48.8%
gO1b
48.9%
141R16
gO2a
1.3%

gO4
5.5%
gO2b
26.4%
174L16
gO3
68.1%

gO4
1.6%
gO1b
2.9%
gO2b
31.8%
174R16
gO3
63.6%

243L16
gO4
100%

243R4
gO4
100%

gO1b
4.3%
gO3
1.6%
243R12
gO4
94.1%

243R16
gO4
100%

248R4
gO2a
100.0%

gO1a
8.8%
gO4
53.1%
gO2b
38.1%
259L16

gO1a
9.5%
gO2b
23.2%
259R16
gO4
67.3%

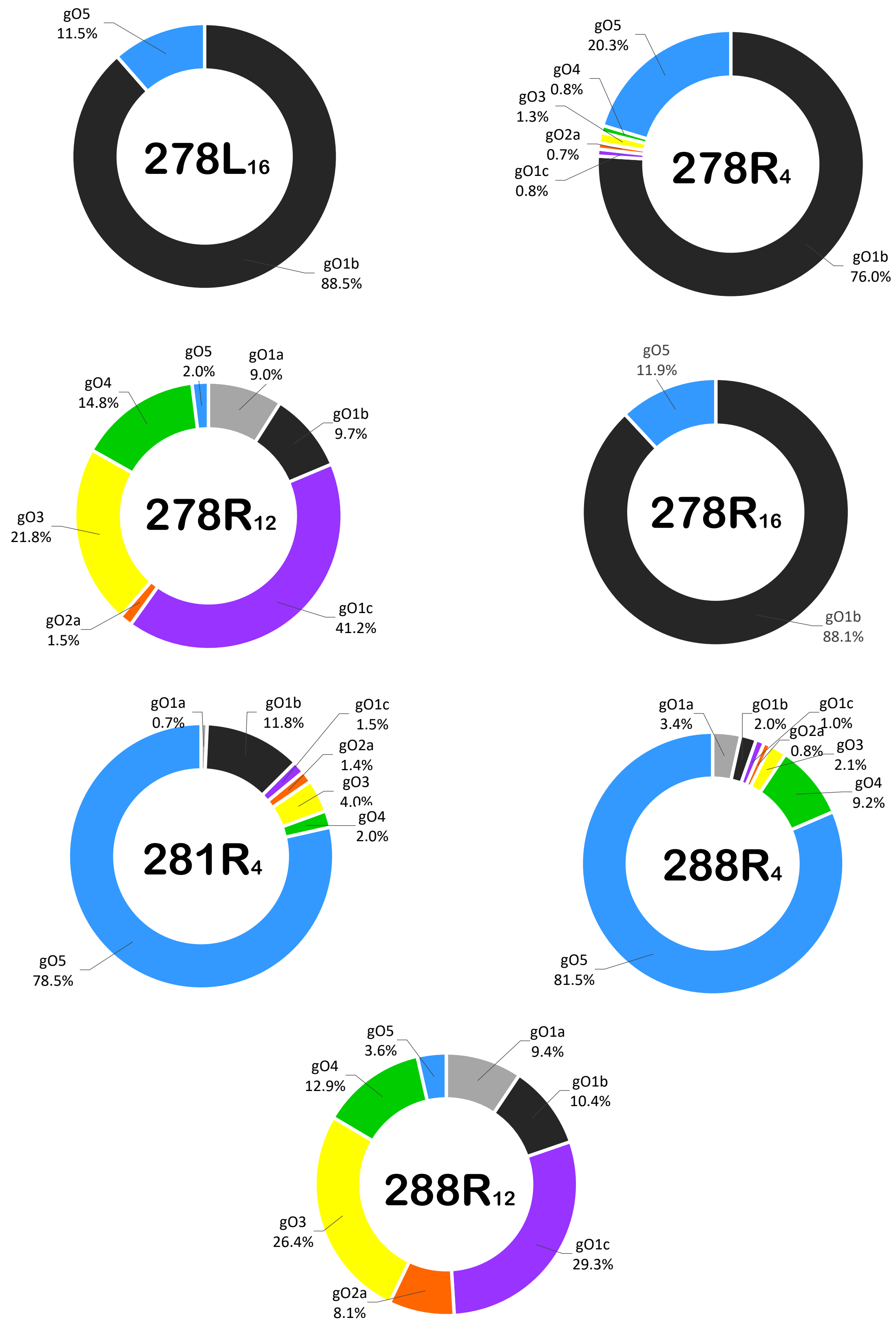


Figure 5.12. HCMV gO genotype proportions in 17 breast milk samples from HIV-positive mothers. Sample IDs are indicated in the centre of each doughnut, with the subscript indicating the time-point in weeks postpartum when the milk sample was collected. Genotypes are colour-coded as follows: gO1a – grey; gO1b – black; gO1c – magenta; gO2a – orange; gO2b – dark red; gO3 – yellow; gO4 – green; gO5 – blue.

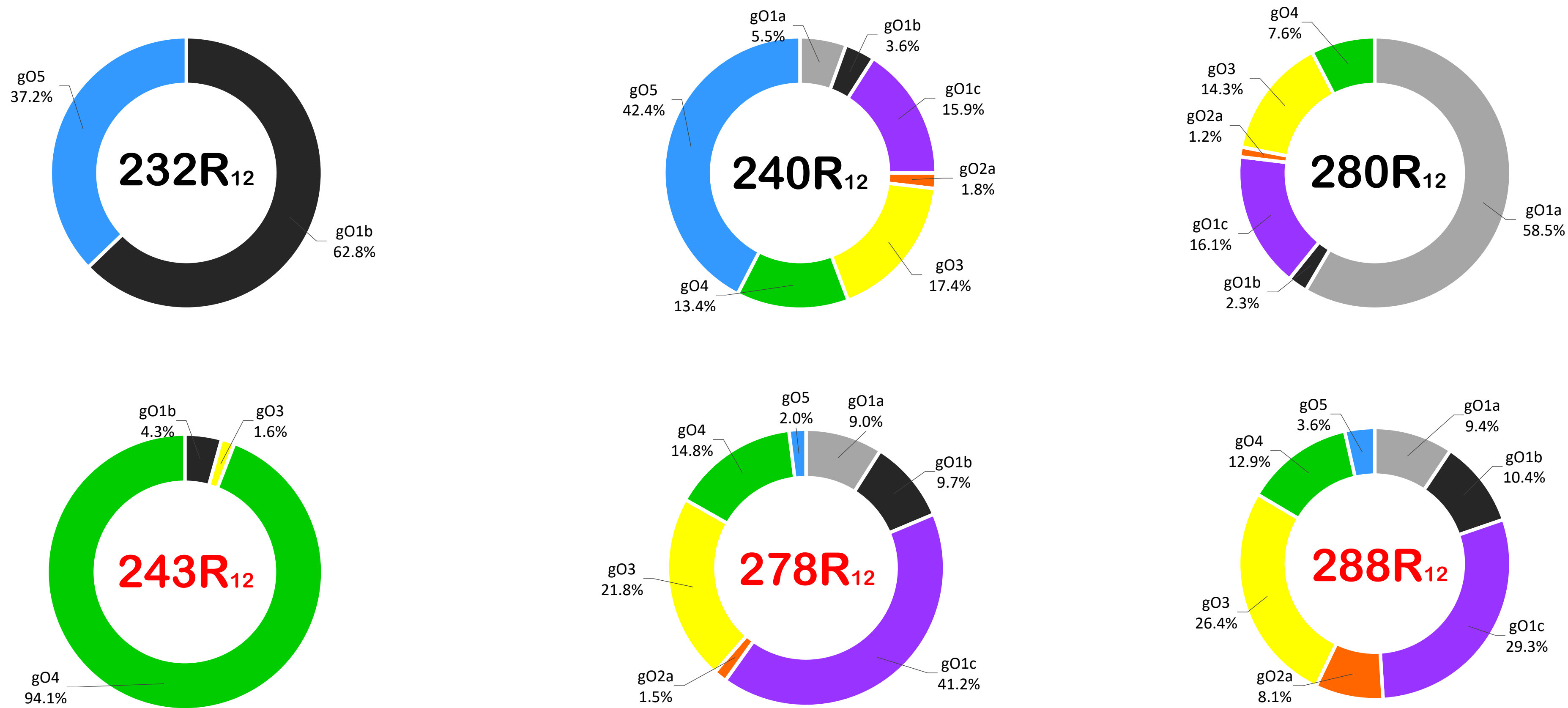


Figure 5.13. HCMV gO genotypes in breast milk from three HIV-positive and three HIV-negative mothers at postpartum week 12. Sample IDs are indicated in the centre of each doughnut, with HIV-positive samples in shown red font and HIV-negative in black. Genotypes are colour-coded as follows: gO1a – grey; gO1b – black; gO1c – magenta; gO2a – orange; gO2b – dark red; gO3 – yellow; gO4 – green; gO5 – blue.

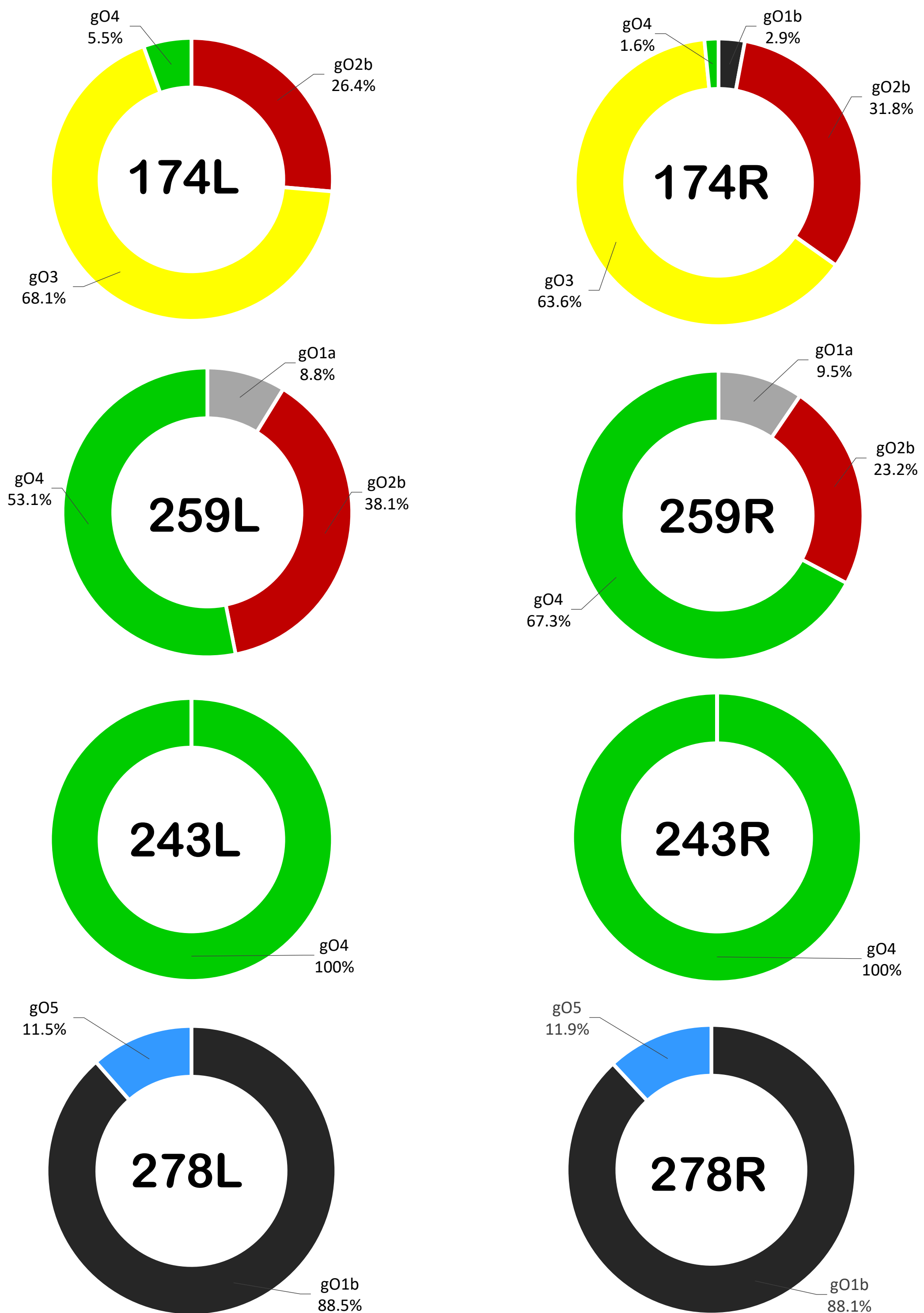


Figure 5.14. HCMV gO genotypes in breast milk from four mothers with milk samples collected from both left (L) and right (R) breasts at week 16 postpartum. Sample IDs are indicated in the centre of each doughnut. All four mothers were HIV-positive. Genotypes are colour-coded as follows: gO1a – grey; gO1b – black; gO1c – magenta; gO2a – orange; gO2b – dark red; gO3 – yellow; gO4 – green; gO5 – blue.

5.3.3 HCMV gO genotype, mixed infections kinetics, and viral load

In order to compare the burden of infection with viral load, the two data sets were combined for samples available from this cohort. In the final dataset, there were three individuals for whom milk samples collected at multiple time-points were sequenced by NGS, and for whom HCMV viral load data were also available: 243, 278, and 288. These were all from HIV-positive women. The HCMV gO proportions were mapped against the viral load in order to assess the effect of viral load on gO genotype (figures 5.15 – 5.17). The data suggest that for these samples, multiple infections were already present from week 4 postpartum. There was an apparent ‘flaring up’ of the minor variants from W4 to W12, and for samples 243 and 278 with available W16 data, this was followed by reversion to the W4 status quo. This may suggest enhanced immune pressure on the virus between postpartum weeks 4 and 12, viral reactivation and import of additional strains into the mammary gland, or superinfection.

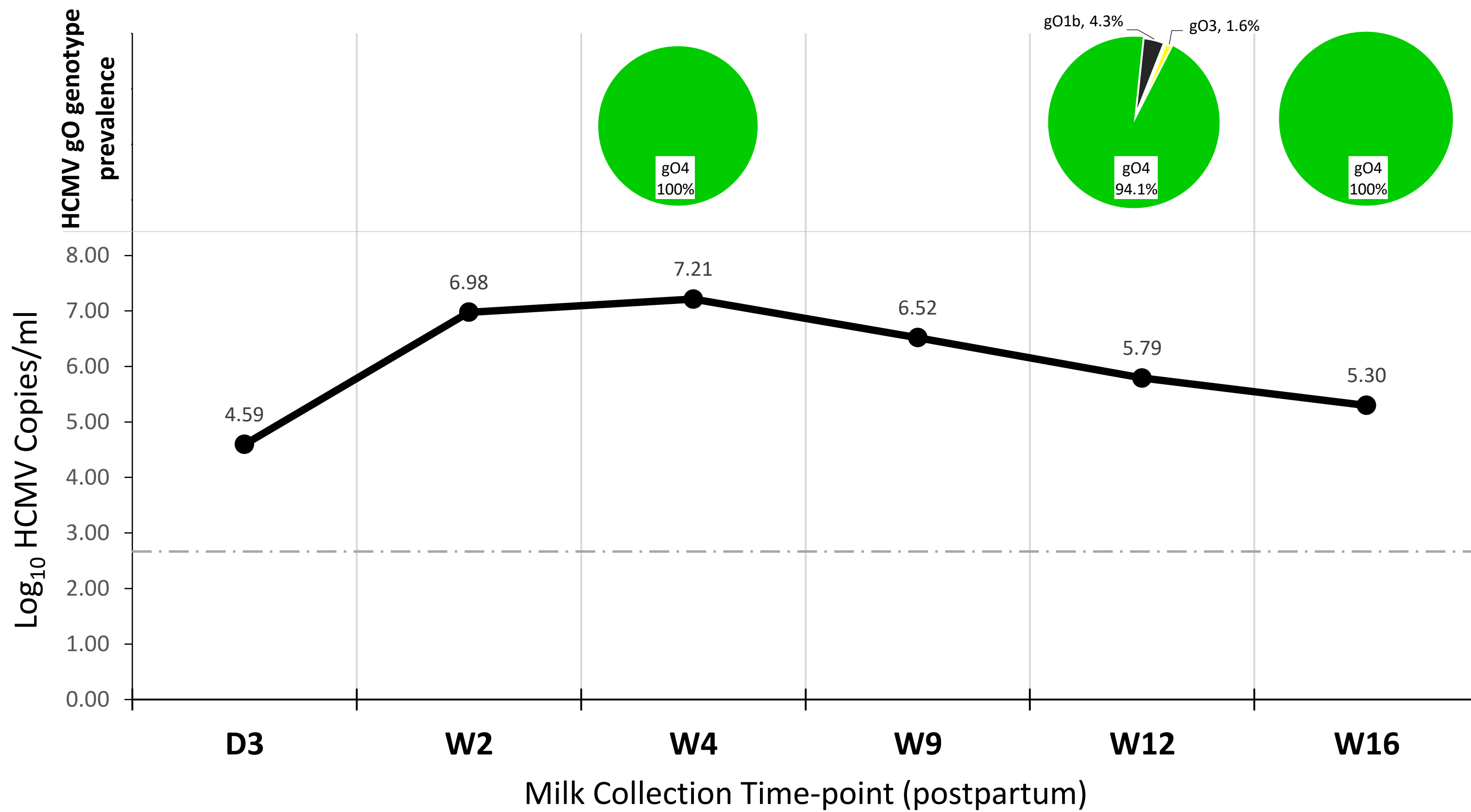


Figure 5.15. HCMV gO genotypes and viral load: sample 243R. The lower panel is a graph of breast milk HCMV viral load kinetics from postpartum day (D) 3 to week (W) 16, and the top panel shows HCMV gO genotypes from NGS analysis at three time-points: W4, W12 and W16.

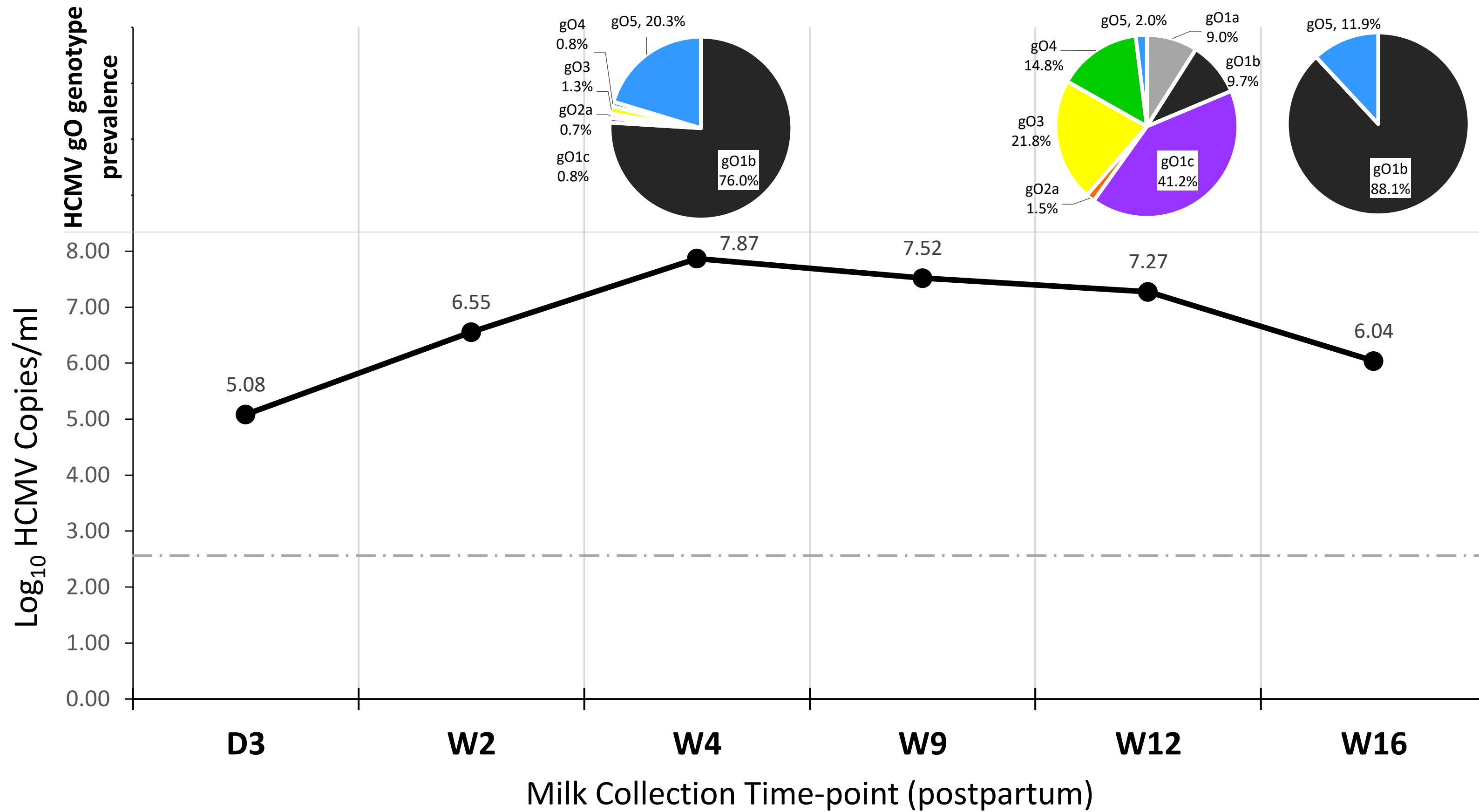


Figure 5.16. HCMV gO genotypes and viral load: sample 278R. The lower panel is a graph of breast milk HCMV viral load kinetics from postpartum day (D) 3 to week (W) 16, and the top panel shows HCMV gO genotypes from NGS analysis at three time-points: W4, W12 and W16.

288R

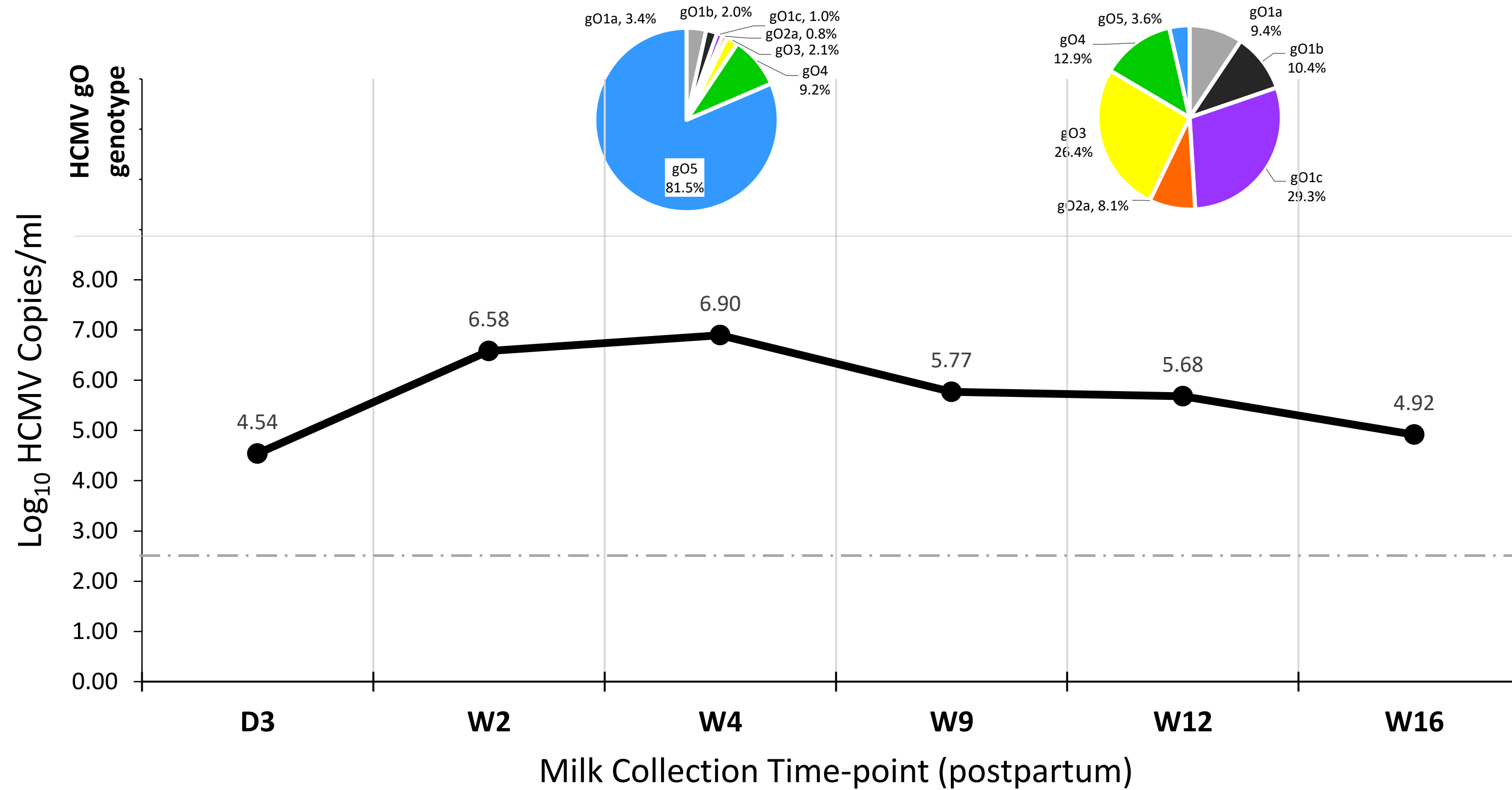


Figure 5.17. HCMV gO genotypes and viral load: sample 288R. The lower panel is a graph of breast milk HCMV viral load kinetics from postpartum day (D) 3 to week (W) 16, and the top panel shows HCMV gO genotypes from NGS analysis at two time-points: W4 and W12.

5.3.4 Summary

The results showed that the molecular tags and scripts could accurately genotype and quantify mixed infections in the breast milk samples. Mixed infections were detected in both HIV-positive and negative women, and the gO and gN genotypes could be used as markers for burden of infection. There was some evidence for wider genotypes indicating more HCMV infections in breast milk samples from HIV-positive women. Where samples were available from both breasts, we found remarkable consistency in genotypes prevalence. In analyses of burden of infection and viral load over time, there appeared in two samples increases in burden of infection after initial virus reactivation and replication, followed by control as the viral load decreased. In the samples analysed, mixed infections were detected from the earliest time point post-partum.

5.4 ASSEMBLY OF COMPLETE HCMV GENOMES FROM BREAST MILK

Despite the apparent global ubiquity of HCMV, scrutiny of GenBank release 211 (National Center for Biotechnology Information (NCBI), 2015) revealed that the 163 publicly available complete HCMV genome sequences were from just 10 countries! Six of the 10 countries are European, and accounted for 155/163 (95%) of the complete HCMV genomes, with Belgium alone contributing 95 sequences (58.3% overall). Figure 5.18, a map constructed to depict the geographical origin of the 163 complete HCMV genomes, vividly shows the paucity of information from the majority of global regions, particularly the Southern Hemisphere. Various factors could account for this picture, including limited access to sequencing technology, and differences in research and health priorities at country / regional levels.

The final part of this thesis presents compilation of complete HCMV genomes assembled from our NGS reads, with detailed analysis of ZMB240 as a clinical representative of the reference strain Merlin. Following analyses of genotype diversity, as described in section 5.3 above, samples with a dominant gO genotype were selected for genome compilation. In other laboratory projects, we also analysed other hypervariable loci in these samples using the molecular tags approach described in Section 5.2, and multiple genotypes were detected for those hypervariable genes (data not shown). After the

initial quality checks and sequence filtering (see section 5.3.1 and table 5.7), two approaches were used to assemble genomes – *de novo* assembly and reference mapping – to ensure there were no gaps in the derived genomes assembly. For *de novo* assembly, FASTQ reads were optimised by VelvetOptimiser (Zerbino and Birney, 2008) and scaffolds were constructed from resultant contigs using SSPACE (Boetzer et al., 2011) and ABACAS (Assefa et al., 2009) with default settings. Contigs were verified by employing the consensus sequence as mapping reference. Reference mapping was accomplished using a variety of genomics tools including the Burrows-Wheeler Aligner, BWA (Li and Durbin, 2009, Li and Durbin, 2010), SAMtools / BCFtools (Li et al., 2009) and GATK (McKenna et al., 2010) for sequence alignment, indexing, mapping, and variant calling. Artemis (Rutherford et al., 2000) was employed for visualisation of mapped reads. Genome annotations were transferred from the genome of HCMV reference strain Merlin (accession number AY446894.2) using the rapid annotation transfer tool, RATT (Otto et al., 2011) as part of the post-assembly genome-improvement toolkit, PAGIT suite (Swain et al., 2012). Each annotated gene was inspected manually and using NCBI BLAST (Altschul et al., 1990), and where necessary, manual adjustments/corrections were made.

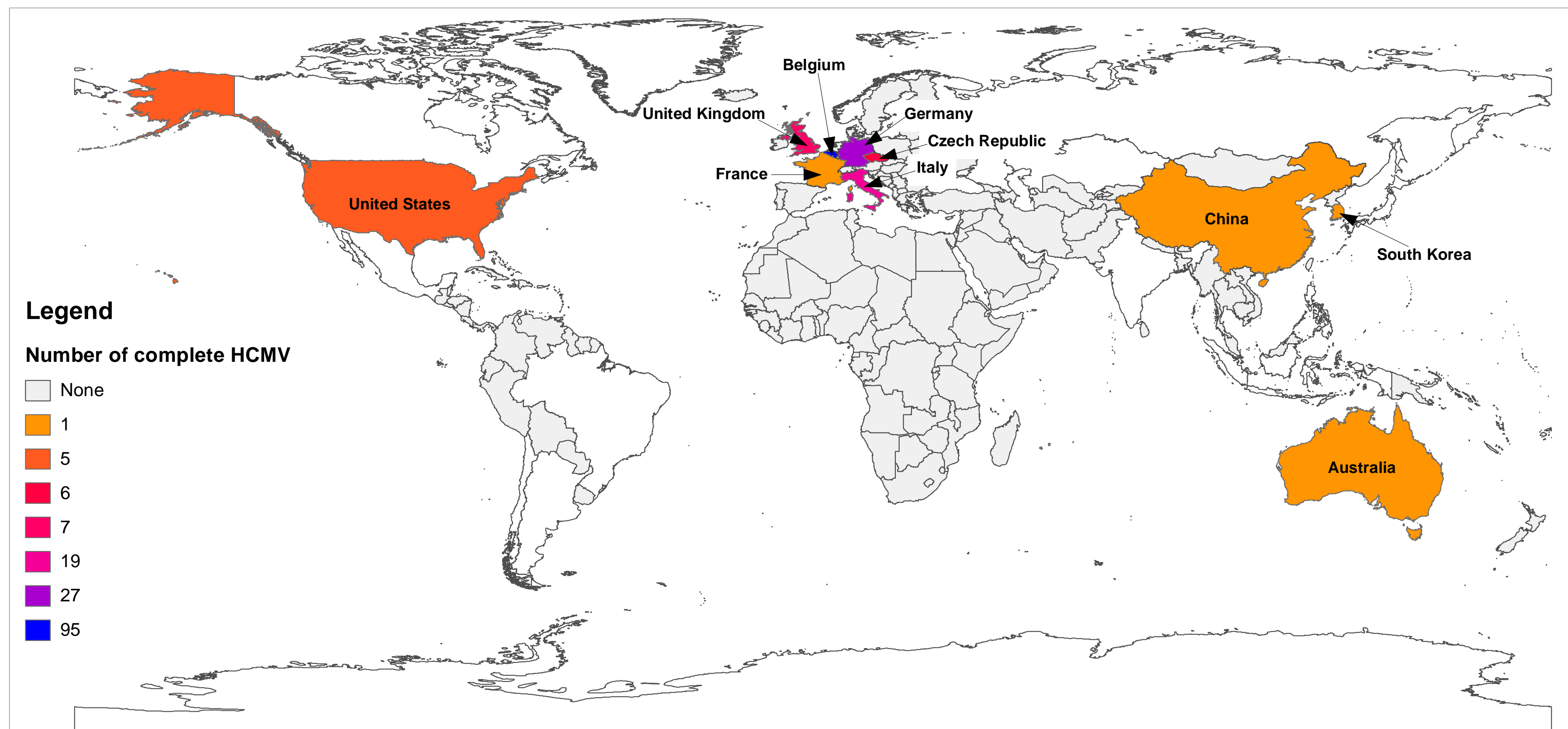


Figure 5.18. World map showing global distribution of complete HCMV genomes by country. The number of genomes published from each country is indicated by a colour gradient, as shown in the legend. Data were compiled from GenBank database release 211, accessed 21st December 2015. The map was created using ArcGIS® 10.3 and ArcMap™ software for Desktop (Esri, www.esri.com).

5.4.1 HCMV strain ZMB240

Genome Architecture

The final compilation consensus genome was performed on ZMB240 (accession number TBA). This sequence was derived from DNA extracted from a breast milk sample from an HIV-negative mother at 12 weeks postpartum. Similar to other HCMV strains (see Appendix 5.1), ZMB240 is of length 236,211bp and 57.6% G+C content, with a typical class E herpesviral genome architecture consisting of two unique regions, long unique (U_L) and short unique (U_S), flanked by inverted repeats, giving the overall arrangement:

$$ab\text{-}U_L\text{-}b'a'c'\text{-}U_S\text{-}ca$$

where *ba*/*b′a′* and *ca*/*c′a′* are the inverted repeats.

The layout and coordinates of the six main regions TR_L, U_L, IR_L, IR_S, U_S and TR_S of the ZMB240 genome are represented in figure 5.19, and the extent of these genome regions is compared to HCMV reference strain Merlin (AY446894.2) in table 5.9.

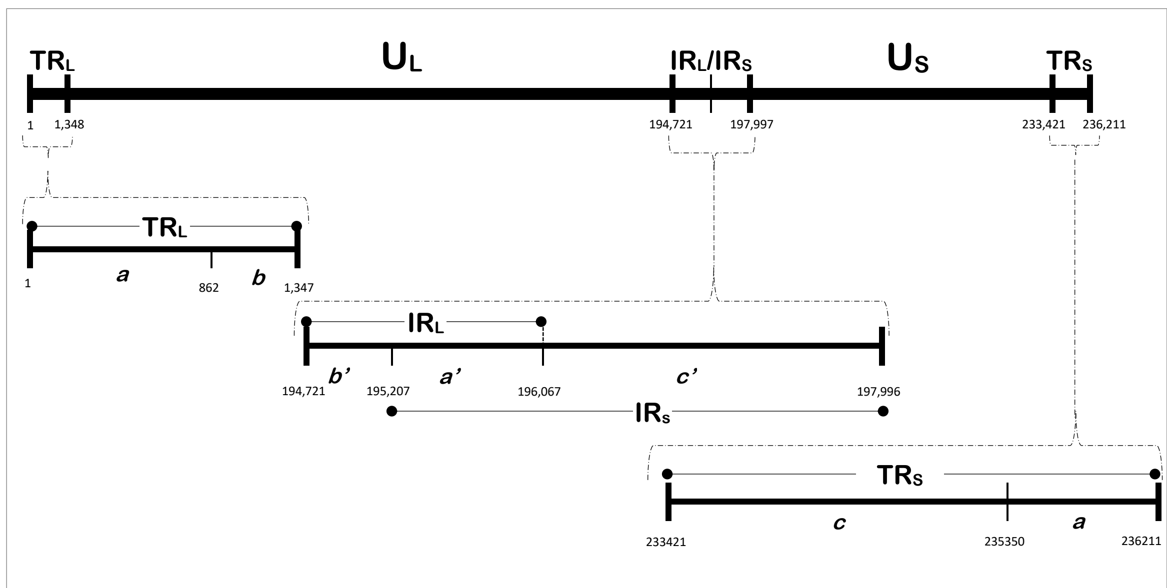


Figure 5.19. HCMV strain ZMB240 genome layout. The genome is layout is in the typical class E herpesviral genome architecture TR_L–U_L–IR_L–IR_S–U_S–TR_S. Numbers indicate the start coordinates for each genome region. TR_L, Long terminal repeat; U_L, long unique region; IR_L, long internal repeat; IR_S, short internal repeat; U_S, unique short region, TR_S, short terminal repeat; *ba*/*b′a′*/*ca*/*c′a′*, inverted repeats.

A: Genome Landmarks	HCMV strain ZMB240			HCMV strain Merlin
	Start	End	Length (bp)	Length (bp)
Full sequence	1	236211	236211	235646
'a' sequence	1	861	861	578
TRL	1	1347	1347	1324
UL	1348	194720	193373	193019
OriLyt-associated repeat	94174	94313	140	140
IRL	194721	196067	1347	1324
'a' sequence	195207	196067	861	578
IRS	195207	197996	2790	2537
US	197997	233420	35424	35482
TRS	233421	236211	2791	2538
'a' sequence	235350	236211	862	579
B: Other Features			Count	Count
CDS			169	169
Spliced genes			16	15
Non-spliced genes			153	154
Repeat regions			8	8
ncRNA regions			30	21
GC percentage			57.6	57.5

Table 5.9. Overview of HCMV strain ZMB240 genome content. Genome features are compared to HCMV reference strain Merlin (accession AY446894).

5.4.2 ZMB240 Genes Compared to HCMV Strain Merlin (AY446894.2)

5.4.2.1 Gene Content

ZMB240 includes a complement of 169 ORFs, of which eight are RL genes, 127 are UL genes and 34 are US genes. ZMB240 possesses an additional putative ORF, which in other HCMV strains including Merlin, encodes UL6. In ZMB240 however, this ORF is initiated by the start codon variant TTG instead of the conventional ATG, and is thus annotated as a “miscellaneous feature”. Bioinformatics analysis including NCBI BLAST searches showed that the TTG start codon in putative UL6 is documented in at least 25 other complete-genome HCMV strains, as shown in figure 5.20.

Among the 169 ORFs in ZMB240, 69 are in the forward direction and 100 in the reverse. Further, among the resultant gene products, 130 have the same overall length as in HCMV strain Merlin, including products of 35/40 core genes, 20/27 beta genes, all 7

betagamma genes, and the immediate early genes UL150A and US2. NCBI BLASTp searches showed that the variation in amino acid length due to deletions and/or insertions in various ZMB240 gene products including the 5 core genes (UL48, UL50, UL95, UL100, UL102), 7 beta genes (UL22A, UL32, UL37, UL44, UL74, UL77, UL117) and the spliced gene UL119 is in fact also present in other HCMV strains. For these key genes the number of complete HCMV genomes in GenBank release 211 with the same product lengths as ZMB240 is 61 (UL48), 4 (UL50), 5 (UL95), 10 (UL100), 5 (UL102), 3 (UL22A), 10 (UL37), 14 (UL119), and one each for UL32, UL42, UL74, UL77, and UL117 as per table 5.10. This was confirmed by protein alignments (figures 5.21 to 5.36) performed using NCBI's COBALT tool (Papadopoulos and Agarwala, 2007) with default settings at https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web.

5.4.2.2 UL128 and the ULb' Region

The UL128 locus encodes a membrane protein which is a component of the pentameric complex (gH-gL-pUL128-pUL130-pUL131) critical for virus entry into non-fibroblast cells such as endothelial, epithelial, and myeloid cells (Hahn et al., 2004, Wang and Shenk, 2005b, Wang and Shenk, 2005a, Ryckman et al., 2008b). HCMV strain Merlin lacks gene UL128 owing to the effects of selection through propagation in cell culture giving rise to an in-frame nucleotide substitution that introduces a stop codon leading to premature sequence termination (Akter et al., 2003, Dolan et al., 2004). Notably, ZMB240 – sequenced directly from a clinical sample – has an intact UL128, with the classical 'C' variation at nucleotide position 634 in the mature transcript. Further, as demonstrated in other clinical strains, the ULb' region is intact in ZMB240 and located at the 3' end of UL. This 13-15kbp region, which encodes at least 20 ORFs (UL133–UL150A), has been termed "the virulence region" and is deleted or modified in fibroblast-cultured laboratory-adapted HCMV strains (Cha et al., 1996, Murphy et al., 2003, Dolan et al., 2004).

The full list of ZMB240 genes is given in table 5.12, which also highlights relative differences from HCMV strain Merlin.

```
                                   10        20        30        40        50        60        70
                                    |         |         |         |         |         |         |
ZMB240                  1 TTGACGTGGAACCACGGTATTTATGGATTTACGCAGGCTAAAATTCCTGTTAATGGAAATGGAACTGTAACG
KP745726.1_BE/30/2010   1 ........................................................................
JX512207.1_HAN28        1 ........................................................................
KJ361957.1_HAN39        1 ........................................................................
KP745692.1_BE/3/2012    1 ........................................................................
KP745693.1_BE/15/2012   1 ........................................................................
KP745680.1_BE/11/2012   1 ........................................................................
KP745690.1_BE/34/2011   1 ........................................................................
KP745682.1_BE/46/2011   1 ........................................................................
KP745689.1_BE/17/2011   1 ........................................................................
GU179288.1_U8           1 ........................................................................
KP745702.1_BE/21/2011   1 ........................................................................
KP745719.1_BE/26/2010   1 ........................................................................
KP745704.1_BE/32/2011   1 ........................................................................
KP745641.1_BE/31/2011   1 ........................................................................
KP745644.1_BE/31/2010   1 ........................................................................
JX512206.1_HAN22        1 ........................................................................
KP745636.1_BE/7/2011    1 ........................................................................
JX512208.1_HAN31        1 ........................................................................
KP745666.1_BE/7/2012    1 ........................................................................
KP745670.1_BE/30/2011   1 ........................................................................
KP745652.1_BE/2/2011    1 ........................................................................
KP745665.1_BE/16/2012   1 ........................................................................
KP745659.1_BE/3/2011    1 ........................................................................
KJ872540.1_PAV18        1 ........................................................................
KJ361951.1_HAN21        1 .....................................................................G..
AY446894.2_Merlin       1 A......T..TAT.T.TGGA.GCG.ATAA...T.GTAA.G.GC.AAAA.CG.CGACTGGGA.T.A.T.T..C
```

Figure 5.20. Nucleotide sequence alignment of the putative UL6 ORF of strain ZMB240 and 25 other HCMV strains. The alignment shows the first 72 nucleotides of putative UL6 sequences for ZMB240 and 25 other HCMV strains in GenBank release 211. In these strains, the ORF is initiated by TTG, compared with the conventional ATG in strains such as the HCMV reference Merlin, shown in the last row.

Strain	Accession	UL48 є	UL50 є	UL95 є	UL100 є	UL102 є	UL22A βṩ	UL32 β	UL37 βṩ	UL42 β	UL74 β	UL77 β	UL117 β	UL29 ṩ	UL33 ṩ	UL112 ṩ	UL119 ṩ
3157	GQ221974.1	X							X								
3301	GQ466044.1						X									X	
6397	JX512197.1	X					X									X	
2CEN2	KJ361946.1				X												
2CEN5	KJ361947.1	X															
AD169	FJ527563.1		X											X	X	X	
AF1	GU179291.1	X														X	
BE/1/2010	KP745677.1			X													
BE/1/2011	KP745650.1	X															
BE/1/2012	KP745699.1	X															
BE/10/2011	KP745639.1	X															
BE/10/2012	KP745649.1		X													X	
BE/11/2010	KC519321.1							X		X						X	
BE/11/2012	KP745680.1	X															X
BE/13/2010	KP745645.1	X															X
BE/13/2012	KP745707.1	X															
BE/14/2011	KP745671.1				X												
BE/15/2010	KP745638.1	X															
BE/15/2012	KP745693.1															X	X
BE/16/2012	KP745665.1																X
BE/17/2010	KP745727.1	X															X
BE/17/2011	KP745689.1	X															
BE/19/2010	KP745712.1	X															
BE/19/2011	KP745654.1															X	
BE/2/2010	KP745717.1	X															
BE/2/2011	KP745652.1	X															
BE/2/2012	KP745710.1	X														X	
BE/2/2013	KP745656.1	X														X	
BE/20/2010	KP745662.1	X	X														
BE/21/2010	KC519322.1																X
BE/22/2011	KP745653.1	X														X	
BE/23/2010	KP745697.1				X												
BE/24/2011	KP745711.1	X														X	
BE/25/2010	KP745678.1															X	
BE/26/2010	KP745719.1																X
BE/26/2011	KP745703.1	X															
BE/27/2010	KC519323.1	X															
BE/28/2010	KP745676.1	X															
BE/29/2010	KP745714.1	X															
BE/29/2011	KP745672.1															X	
BE/3/2010	KP745655.1	X															
BE/3/2011	KP745659.1	X															
BE/30/2010	KP745726.1	X															
BE/30/2011	KP745670.1																X
BE/31/2011	KP745641.1	X					X										
BE/32/2010	KP745634.1	X															
BE/33/2011	KP745674.1	X															
BE/37/2011	KP745723.1															X	
BE/39/2011	KP745686.1															X	
BE/4/2011	KP745700.1	X			X												X
BE/40/2011	KP745722.1	X															
BE/42/2011	KP745673.1	X															
BE/46/2011	KP745682.1		X	X													
BE/49/2011	KP745725.1	X															
BE/6/2010	KP745701.1	X															
BE/6/2012	KP745695.1			X												X	
BE/7/2011	KP745636.1	X															
BE/7/2012	KP745666.1																X

Strain	Accession	UL48 Є	UL50 Є	UL95 Є	UL100 Є	UL102 Є	UL22A βṨ	UL32 β	UL37 βṨ	UL42 β	UL74 β	UL77 β	UL117 β	UL29 Ṩ	UL33 Ṩ	UL112 Ṩ	UL119 Ṩ
BE/8/2011	KP745648.1			X												X	
BE/9/2010	KC519319.1		X														
BE/9/2011	KP745637.1	X															
CZ/1/2011	KP745718.1	X															
CZ/1/2012	KP745642.1	X														X	
CZ/3/2012	KP745685.1	X															X
HAN1	JX512199.1				X											X	
HAN11	KJ361950.1	X			X												
HAN12	JX512203.1	X															
HAN13	GQ221973.1				X											X	
HAN16	JX512204.1	X															
HAN2	JX512200.1				X											X	
HAN20	GQ396663.1	X														X	
HAN21	KJ361951.1	X														X	
HAN22	JX512206.1	X															
HAN28	JX512207.1	X				X											
HAN3	JX512201.1	X															
HAN30	KJ361953.1															X	
HAN33	KJ361955.1	X														X	
HAN38	GQ396662.1															X	
HAN40	KJ361958.1				X												
HAN8	JX512202.1	X															
JP	GQ221975.1										X						
PAV11	KJ361965.1	X														X	X
PAV23	KJ361967.1	X															
PAV24	KJ361968.1															X	
PAV26	KJ361970.1	X															
PAV4	KJ361960.1	X															
PAV5	KJ361961.1	X															
PAV7	KJ361963.1															X	
PAV8	KJ361964.1	X															
Toledo	GU937742.1															X	X
Towne	FJ616285.1	X		X								X	X				
U11	GU179290.1															X	
U8	GU179288.1	X														X	X
UKNEQAS1	KJ361971.1	X														X	
UKNEQAS2	KT634296.1	X			X												
VR1814	GU179289.1	X															

Table 5.10. ZMB240 Core and Beta genes whose products are of variable length compared to HCMV reference strain Merlin (AY446894.2). HCMV strains listed in the first column have amino acid products of the same length as ZMB240 for the genes indicated by an X. Є, Core gene; β, beta gene; Ṩ, spliced gene.

```
ZMB240      1 MARRLWILSLLAVTLTVALAAPSQKSKRSVTVEQPSTSTNSDGNNTTRNKDVTLSQGGSTTDGDEDYSGGDYDVLITDTD 80
6397        1 MARRLWILSLLAVTLTVALAAPSQKSKRSVTVEQPSTSTNSDGNNTTRNKDVTLSQGGSTTDEDEDYSGGDYDVLITDTD 80
BE/31/2011  1 MARRLWILSLLAVTLTVALAAPSQKSKRSVTVEQPSTSTNSDGNNTTRNKDVTLSQGGSTTDGDEDYSGEEYDVLITDGD 80
3301        1 MARRLWILSLLAVTLTVALAAPSQKSKRSVTVEQPSTSTNSDGNNTTRNKDVTLSQGGSTTDGDEDYSGGDYDVLITDTD 80
Merlin      1 MARRLWILSLLAVTLTVALAAPSQKSKRSVTVEQPSTSTNSDGN-TTPSKNVTLSQGGSTTDGDEDYSGGDYDVLITDTD 79

ZMB240     81 GGNHQQPQKKTNEHKEEHTKENEKTQ 106
6397       81 GGNHQQPQEKTNEHKEEHTKENEKTQ 106
BE/31/2011 81 GSEHQQPQEKTDEHKGEHTKKNEKTQ 106
3301       81 GGNHQQPQEKTDEHKEEHTKENEKTQ 106
Merlin     80 GGNHQQPQEKTDEHKGEHTKENEKTQ 105
```

Figure 5.21. ZMB240 UL22A amino acid sequence alignment. Compared to HCMV strain Merlin (AY446894.2), the ZMB240 UL22A amino acid sequence has an insertion after position 44, highlighted by the red border. This insertion is also present in HCMV strains 6397, BE/31/2011, and 3301, as shown in the alignment.

```
ZMB240 1     MSGRRKGCSAATASSSSSSPPSRLPPLPGHARRPRRKRCLVPEVFCTRDLADLCVRRDYEGLRRYLRRFEGSCVSLGWPS  80
AD169  1     MSGRRKGCSAATASSSSSSPPSRLPPLPGHARRPRRKRCLVPEVFCTRDLADLCVRRDYEGLRRYLRRFEGSCVSLGWPS  80
Merlin 1     MSGRRKGCSAATASSSSSSPPSRLP-LPGHARRPRRKRCLVPEVFCTRDLADLCVRRDYEGLRRYLRRFEGSCVSLGWPS  79

ZMB240 81    QCIYVVGGEHSPHSLTEIDLEHCQNDFFGEFRALHLIGTVSHATCRYQVFVDAYGAVFAYDAQEDCLYELASDLAGFFAK  160
AD169  81    QCIYVVGGEHSPHSLTEIDLEHCQNDFFGEFRALHLIGTVSHATCRYQVFVDAYGAVFAYDAQEDCLYELASDLAGFFAK  160
Merlin 80    QCIYVVGGEHSPHSLTEIDLEHCQNDFFGEFRALHLIGTVSHATCRYQVFVDAYGAVFAYDAQEDCLYELASDLAGFFAK  159

ZMB240 161   GMIRCDPVHESICARLQPNVPLVHPDHRAELCRRSRASARGRYLRSLLAFRELLACEDTAARCAYVEAHREAQLTLIWPE  240
AD169  161   GMIRCDPVHESICARLQPNVPLVHPDHRAELCRRSRASARGRYLRSLLAFRELLACEDTAARCAYVEAHREAQLTLIWPE  240
Merlin 160   GMIRCDPVHESICARLQPNVPLVHPDHRAELCRRSRASARGRYLRSLLAFRELLACEDTAARCAYVEAHREAQLTLIWPE  239

ZMB240 241   KHSLVLRTARDLGLSASMLRRFQRSLYTREPVMPLGEIEGAEDKTFFHRVRILCGDTGTVYAALVGQDKLVRLARDLRGF  320
AD169  241   KHSLVLRTARDLGLSASMLRRFQRSLYTREPVMPLGEIEGAEDKTFFHRVRILCGDTGTVYAALVGQDKLVRLARDLRGF  320
Merlin 240   KHSLVLRTARDLGLSASMLRRFQRSLYTREPVMPLGEIEGAEDKTFFHRVRILCGDTGTVYAALVGQDKLVRLSRDLRGF  319

ZMB240 321   VRVGLALLIDDFRYESIGPVDRSSLYEANPELRLPFKKRRLVVGYFDSLSSLYLRGQPKFSSIWRGLRDAWTHKRPKPRE  400
AD169  321   VRVGLALLIDDFRYESIGPVDRSSLYEANPELRLPFKKRRLVVGYFDSLSSLYLRGQPKFSSIWRGLRDAWTHKRPKPRE  400
Merlin 320   VRVGLALLIDDFRYESIGPVDRSSLYEANPELRLPFKKRRLVVGYFDSLSSLYLRGQPKFSSIWRGLRDAWTHKRPKPRE  399

ZMB240 401   RASGVHLQRYVRATAGRWLPLCWPPLHGIMLGDTQYFGVVRDHKTYRRFSCLRQAGRLYFIGLVSVYECVPDANTAPEIW  480
AD169  401   RASGVHLQRYVRATAGRWLPLCWPPLHGIMLGDTQYFGVVRDHKTYRRFSCLRQAGRLYFIGLVSVYECVPDANTAPEIW  480
Merlin 400   RASGVHLQRYVRATAGRWLPLCWPPLHGIMLGDTQYFGVVRDHKTYRRFSCLRQAGRLYFIGLVSVYECVPDANTAPEIW  479

ZMB240 481   VSGHGHAFAYLPGEDKVYVLGLSFGEFFENGLFAVYSFFERDYVDEIVEGAWFKHTFAGMYELSQILHDRANLLRVCQLH  560
AD169  481   VSGHGHAFAYLPGEDKVYVLGLSFGEFFENGLFAVYSFFERDYVDEIVEGAWFKHTFAGMYELSQILHDRANLLRVCQLH  560
Merlin 480   VSGHGHAFAYLPGEDKVYVLGLSFGEFFENGLFAVYSFFERDYVDEIVEGAWFKHTFAGMYELSQILHDRANLLRVCQLH  559

ZMB240 561   AGSKIRLGGSPACTFTFGSWNVAEADEANNFVIGVLEQAHFVVIGWMEPVNKAVFMDAHGGIHVLLYGTMLVKLAETLRG  640
AD169  561   AGSKIRLGGSPACTFTFGSWNVAEADEANNFVIGVLEQAHFVVIGWMEPVNKAVFMDAHGGIHVLLYGTMLVKLAETLRG  640
Merlin 560   AGSKIRLGGSPACTFTFGSWNVAEADEANNFVIGVLEQAHFVVIGWMEPVNKAVFMDAHGGIHVLLYGTMLVKLAETLRG  639

ZMB240 641   FIRQGSFWFRCPRRFCFSPLDSSATVAAKPVSSHTSPAYDVSEYVFSGRSVLDSVSGTGAS  701
AD169  641   FIRQGSFWFRCPRRFCFSPLDSSATVAAKPVSSHTSPAYDVSEYVFSGRSVLDSVSGTGAS  701
Merlin 640   FIRQGSFWFRCPRRFCFSPLDSSATVAAKPVSSHTSPAYDVSEYVFSGRSVLDSVSGTGAS  700
```

Figure 5.22. ZMB240 UL29 amino acid sequence alignment. Similar to HCMV strain AD169, the ZMB240 UL29 amino acid sequence has an insertion after position 25, which introduces a glycosylation site, as highlighted by the red border.

```
ZMB240      1 MSLQFIGLQRRDVVALVNFLRHLTQKPDVDLEAHPKILKKCGEKRLHRRTVLFNELMLWLGYYRELRFHNPDLSSVLEEF  80
BE/11/2010 1 MSLQFIGLQRRDVVALVNFLRHLTQKPDVDLEAHPKILKKCGEKRLHRRTVLFNELMLWLGYYRELRFHNPDLSSVLEEF  80
Merlin      1 MSLQFIGLQRRDVVALVNFLRHLTQKPDVDLEAHPKILKKCGEKRLHRRTVLFNELMLWLGYYRELRFHNPDLSSVLEEF  80

ZMB240      81 EVRCAAVARRGYTYPFGDRGKARDHLAVLDHTEFDTDVRHDAEIVERALVSAVILAKMSVRETLVTAIGQTEPIAFVHLK  160
BE/11/2010 81 EVRCAAVARRGYTYPFGDRGKARDHLAVLDRTEFDTDVRHDAEIVERALVSAVILAKMSVRETLVTAIGQTEPIAFVHLK  160
Merlin      81 EVRCAAVARRGYTYPFGDRGKARDHLAVLDRTEFDTDVRHDAEIVERALVSAVILAKMSVRETLVTAIGQTEPIAFVHLK  160

ZMB240      161 DTEVQRIEENLEGVRRNMFCVKPLDLNLDRHANTALVNAVNKLVYTGRLIMNVRRSWEELERKCLARIQERCKLLVKELR  240
BE/11/2010 161 DTEVQRIEENLEGVRRNMFCVKPLDLNLDRHANTALVNAVNKLVYTGRLIMNVRRSWEELERKCLARIQERCKLLVKELR  240
Merlin      161 DTEVQRIEENLEGVRRNMFCAKPLDLNLDRHANTALVNAVNKLVYTGRLIMNVRRSWEELERKCLARIQERCKLLVKELR  240

ZMB240      241 MCLSFDSNYCRNILKHAVENGDSADTLLELLIEDFDIYVDSFPQSAHTFLGARSPSLEFDDDANLLSLGGGSAFSSVPKK  320
BE/11/2010 241 MCLSFDSNYCRNILKHAVENGDSADTLLELLIEDFDIYVDSFPQSAHTFLGARSPSLEFDDDANLLSLGGGSAFSSVPKK  320
Merlin      241 MCLSFDSNYCRNILKHAVENGDSADTLLELLIEDFDIYVDSFPQSAHTFLGARSPSLEFDDDANLLSLGGGSAFSSVPKK  320

ZMB240      321 HVPTQPLDGWSWIASPWKGHKPFRFEAHGSLAPAAEAHAARSAAVGYYDEEEKRRERQKRVDDEVVQREKQQLKAWEERQ  400
BE/11/2010 321 HVPTQPLDGWSWIASPWKGHKPFRFEAHGSLAPAAEAHAARSAAVGYYDEEEKRRERQKRVDDEVVQREKQQLKAWEERQ  400
Merlin      321 HVPTQPLDGWSWIASPWKGHKPFRFEAHGSLAPAAEAHAARSAAVGYYDEEEKRRERQKRVDDEVVQREKQQLKAWEERQ  400

ZMB240      401 QNLQQRQQQPPPSTRKPGASRRLFGSSADE-DDDDDDDEKNIFTPIKKPGTSGKGAASGGGVSNIFSGLLSSGSQKPTSG  479
BE/11/2010 401 QNLQQRQQQPPPPARKPSASRRLFGSSADE-DDDDDDDEKNIFTPIKKPGTSGKGAASGGGVSSIFSGLLSSGSQKPTSG  479
Merlin      401 QNLQQRQQQPPPPTRKPGASRRLFGSSADEDDDDDDDDEKNIFTPIKKPGTSGKGAASGGGVSSIFSGLLSSGSQKPTSG  480

ZMB240      480 PLNIPQQQQRHAAFSLVSPQVTKASPGRVRRDSAWDVRPLTETRGDLFSGDEDSDSSDGYPPNRQDPRFTDTLVDITDTE  559
BE/11/2010 480 PLNIPQQQQRHAAFSLVSPQVTKASPGRVRRDSAWDVRPLTETRGDLFSGDEDSDSSDGYPPNRQDPRFTDTLVDITDTE  559
Merlin      481 PLNIPQQQQRHAAFSLVSPQVTKASPGRVRRDSAWDVRPLTETRGDLFSGDEDSDSSDGYPPNRQDPRFTDTLVDITDTE  560

ZMB240      560 TNVKPPVTTAYKFEQPTLTFGAGVNVPAGAGAAILTPTPVNPSTAPAPAPTPTFAGTQTPVNGNSPWAPTAPLPGDMNPA  639
BE/11/2010 560 TSAKPPVTTAYKFEQPTLTFGAGVNVPAGAGAAILTPTPVNPSTAPAPAPTPTFAGTQTPVNGNSPWAPTAPLPGDMNPA  639
Merlin      561 TSAKPPVTTAYKFEQPTLTFGAGVNVPAGAGAAILTPTPVNPSTAPAPAPTPTFAGTQTPVNGNSPWAPTAPLPGDMNPA  640

ZMB240      640 NWPRERAWALKNPHLAYNPFRMPTTSTASQNTVSTTPRRPSTPRAAVTQTASQNAADEVWALRDQTAESPVEDSEEEDDD  719
BE/11/2010 640 NWPRERAWALKNPHLAYNPFRMPTTSTASQNNVSTTPRRPSTPRAAVTQTASQNAADEVWALRDQTAESPVEDSEEEDDD  719
Merlin      641 NWPRERAWALKNPHLAYNPFRMPTTSTASQNTVSTTPRRPSTPRAAVTQTASRDAADEVWALRDQTAESPVEDSEEEDDD  720

ZMB240      720 SSDTGSVVSLGHTTPSSDY-NDVISPPSQTPEQSTPSRIRKAKLSSPMTTTSTSQKPVLGKRVATPHASARAQTVTSTPV  798
BE/11/2010 720 SSDTGSVVSLGHTTPSSDY-NDVISPPSQTPEQSTPSRIRKAKLSSPMTTTSTSQKPVLGKRVATPHASARAQTVTSTPV  798
Merlin      721 SSDTGSVVSLGHTTPSSDYNNDVISPPSQTPEQSTPSRIRKAKLSSPMTTTSTSQKPVLGKRVATPHASARAQTVTSTPV  800

ZMB240      799 QGRLEKQVSGTPSTVPATLLQHQPASSKTTSSRNVTSGAGTSSASSARQPSASASVLSPTEDDVVSPATSPLSMLSSASP  878
BE/11/2010 799 QGRLEKQVSGTPSTVPATLLQPQPASSKTTSSRNVTSGAGTSSASSARQPSASASVLSPTEDDVVSPATSPLSMLSSASP  878
Merlin      801 QGRLEKQVSGTPSTVPATLLQPQPASSKTTSSRNVTSGAGTSSASSARQPSASASVLSPTEDDVVSPATSPLSMLSSASP  880

ZMB240      879 SPAKSAPPSPVRGRGSRVGVVPSLKPTLGGKAVVGRPPSVPVSGSAPGRLSGTSRAASTTPTYPAVTTVYPPSSTAKSSV  958
BE/11/2010 879 SPAKSAPPSPVKGRGSRVGVVPSLKPTLGGKAVVGRPPSVPVSGSAPGRLSGTSRAASTTPTYPAVTTVYPPSSTAKSSV  958
Merlin      881 SPAKSAPPSPVKGRGSRVG-VPSLKPTLGGKAVVGRPPSVPVSGSAPGRLSGSSRAASTTPTYPAVTTVYPPSSTAKSSV  959

ZMB240      959 SNAPPVASPSILKPGASAALQSRRSTGTAAVGSPVKSTTGMKTVAFDLSSPQKSGTGPQPGSAGMGGAKTPSDAVQNILQ  1038
BE/11/2010 959 SNAPPVASPSILKPGASAALQSRRSTGTAAVGSLVKSTTGMKTVAFDLSSPQKSGTGPQPGSAGMGGAKTPSDAVQNILQ  1038
Merlin      960 SNAPPVASPSILKPGASAALQSRRSTGIAAVGSPVKSTTGMKTVAFDLSSPQKSGTGPQPGSAGMGGAKTPSDAVQNILQ  1039

ZMB240      1039 KIEKIKKTEE  1048
BE/11/2010 1039 KIEKIKNTEE  1048
Merlin      1040 KIEKIKNTEE  1049
```

Figure 5.23. ZMB240 UL32 amino acid alignment. Relative to HCMV strain Merlin (AY446894.2), the ZMB240 UL32 amino acid sequence has two deletions at positions 431 and 720 and an insertion after position 899, highlighted by the red borders. These changes are also present in HCMV strain BE/11/2010 as shown here.

```
ZMB240 1    MDTIIHNTTNRSTSTPHVNSTCNMTEPLSAIRTTEAVINTFIIFVGGPLNAIVLVTQLLTNRVLGYSTPTIYMTNLYSTN  80
AD169  1    MDTIIHNTTNRSTSTPHVNSTCNMTEPLSAIRTTEAVINTFIIFVGGPLNAIVLVTQLLTNRVLGYSTPTIYMTNLYSTN  80
Merlin 1    MDTIIHNTTNRSTDTPHVNITCNITEPLSAIRTTEAVINTFIIFVGGPLNAIVLITQLLTNRVLGYSTPTIYMTNLYSTN  80

ZMB240 81   FLTLTVLPFIVLSNQWLLPASVTSCKFLSVIYYSSCTVGFATVALIAADRYRVLHKRTYARQSYRSTYIILLLTWFAGLI  160
AD169  81   FLTLTVLPFIVLSNQWLLPASVTSCKFLSVIYYSSCTVGFATVALIAADRYRVLHKRTYARQSYRSTYIILLLTWFAGLI  160
Merlin 81   FLTLTVLPFIVLSNQWLLPASVASCKFLSVIYYSSCTVGFATVALIAADRYRVLHKRTYARQSYRSTYIILLLTWFAGLI  160

ZMB240 161  FSMPAAVYTTVVIHNGTD-ENTNGHATCVLYFIADEVYTVLLSWKVLLTLVWGAAPVIMMTWFYAFFYSTVQRASQKQRS  239
AD169  161  FSMPAAVYTTVVIHNGTD-ENTNGHATCVLYFIADEVYTVLLSWKVLLTLVWGAAPVIMMTWFYAFFYSTVQRASQKQRS  239
Merlin 161  FSMPAAVYTTVVIHNGTNGQSSNGHATCVLYFIADEVYTVLLSWKVLLTLVWGAAPVIMMTWFYAFFYSTVQRASQKQRS  240

ZMB240 240  RTLTFVSVLLISFVALQTPYVSIMIFNSYATAAWPMDCEHLTLRRTIGTLSRLVPHLHCLINPILYALLGHDFLQRMRQC  319
AD169  240  RTLTFVSVLLISFVALQTPYVSIMIFNSYATAAWPMDCEHLTLRRTIGTLSRLVPHLHCLINPILYALLGHDFLQRMRQC  319
Merlin 241  RTLTFVSVLLISFVALQTPYVSIMIFNSYATAAWPMDCEHLTLRRTIGTLSRLVPHLHCLINPILYALLGHDFLQRMRQC  320

ZMB240 320  FRGQLLDRRAFLRSQQNQRATAETNLAAGNNSQSVATSLDTNSKNCNQHAKRSVSFNFPSGTWKGGQKTASNDTSTKIPH  399
AD169  320  FRGQLLDRRAFLRSQQNQRATAETNLAAGNNSQSVATSLDTNSKNCNQHAKRSVSFNFPSGTWKGGQKTASNDTSTKIPH  399
Merlin 321  FRGQLLDRRAFLRSQQNQRATAETNLAAGNNSQSVATSLDTSSKNCNQHAKRSVSFNFPSGTWKGGQKTASNDTSTKIPH  400

ZMB240 400  RLSQSHHNLSGV  411
AD169  400  RLSQSHHNLSGV  411
Merlin 401  RLSQSHHNLSGV  412
```

Figure 5.24. ZMB240 UL33 amino acid alignment. The ZMB240 UL33 amino acid sequence has a deletion at position 179, also present in HCMV strain AD169, relative to HCMV strain Merlin AY446894.2. This deletion is highlighted by the red border.

```
ZMB240      1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
HAN40       1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
HAN21       1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
BE/29/2010  1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRRKKACALTRCSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
BE/1/2010   1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
BE/10/2011  1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
PAV20       1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
PAV18       1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
VR1814      1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
3157        1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
JP          1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80
Merlin      1   MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLA 80

ZMB240      81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRARLLQHNGPQQ 160
HAN40       81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRARLLQHNGPQQ 160
HAN21       81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRAQLLQHNGPQQ 160
BE/29/2010  81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRAQLLQHNGPQQ 160
BE/1/2010   81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRARLLQHNGPQQ 160
BE/10/2011  81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRARLLQHNGPQQ 160
PAV20       81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRARLLQHNGPQQ 160
PAV18       81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRAQLLQHNGPQQ 160
VR1814      81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRAQLLQHNGPQQ 160
3157        81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRQRARMLQHNGPQQ 160
JP          81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRQARLLQHNGPQQ 160
Merlin      81  EAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPMDAEGLSGLAREACEALKKALRRHRFLWQRRRRARLLQHNGPQQ 160

ZMB240      161 SHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
HAN40       161 SHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
HAN21       161 SHHAAGFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCHLNCTCNNTQIYNGPCVGTEARLP 240
BE/29/2010  161 SHHAAVFCRVHGLRGFQVSVWLLLTLFLSTGHGVSVRCTYHGTDINVTSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
BE/1/2010   161 SHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
BE/10/2011  161 FHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
PAV20       161 SHHAAVFCRVHGLRGFQVSVWLLLTLLWNTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
PAV18       161 SHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
VR1814      161 SHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
3157        161 FHHAAGFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
JP          161 SHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLP 240
Merlin      161 SHHAAVFCRVHGLRGFQVSVWLLLTLFWSTGYGVSVRCTYHGTDINVTSNATSMNCRLNCTCNHTQIYNGPCAGAESKLP 240
```

```
ZMB240     241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
HAN40      241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
HAN21      241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVRGEIACFTNDTNVTVGQLMLNFTGHSYVLQAVAHT 320
BE/29/2010 241 LNVTFQQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESRETNVTEVRGEIACFTNDTNVTVGQLMLNFTGHSYVLQAVAHT 320
BE/1/2010  241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
BE/10/2011 241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
PAV20      241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
PAV18      241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
VR1814     241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
3157       241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
JP         241 LNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHT 320
Merlin     241 LNVTFRQSRRQWHSVMLTFGFQYHLEGWFPLRILNESRDINVTEVYGEVACFTNDTNITMGQLTLNLTGRSYVLRALART 320

ZMB240     321 SPFESYVHWEETNATSNMTSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKITKRKTINATFTGTNVTSLPM 400
HAN40      321 SPFESYVHWEETNATSNMTSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKITKRKTINATFTGTNVTSLPM 400
HAN21      321 SPFESYVHWEETNVTSDVTSNVTSLENTTTVMLILTKYAESDYIFLQDMCPRFLRKTLKLTRNKKRNSTFTGTNVTSLPM 400
BE/29/2010 321 SPFESYVHWEETNVTSNVTSNVTSVENTTTVMLILTKYAESDYIFLQDMCPRFLRRTLKLTRNKKRNSTFTGTNVTSLPM 400
BE/1/2010  321 SPFESYVHWEETNATSNMTSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKITKRKTINATFTGTNVTSLPM 400
BE/10/2011 321 SPFESYVHWEETNATSNMTSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKITKRKTINATFTGTNVTSLPM 400
PAV20      321 SPFESHVHWEETNATSNITSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKLTKNKTRNATFTGTNVTSLPM 400
PAV18      321 SPFESYVHWEETNATSNMTSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKITKRKTINATFTGTNVTSLPM 400
VR1814     321 SPFESHVHWEETNATSNITSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKLTKNKTRNATFTGTNVTSLPM 400
3157       321 SPFESHVHWEETNATSNITSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKLTKKKTRNATFTGTNVTSLPM 400
JP         321 SPFESHVHWEETNATSNITSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKLTKNKTRNATFTGTNVTSLPM 400
Merlin     321 SPFESSVNWEETNVTDNATS---SENNTVTVMSVLTVYAESDYIFLQDMCPRFLKRSVKLAKNNRRNTTFTGTNVTSLPE 397

ZMB240     401 WT-PECEGWKYWTTLSIMWKNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
HAN40      401 WT-PECEGWKYWTTLSIMWKNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
HAN21      401 WT-PECEGWKYWTTLSVMWRNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
BE/29/2010 401 WT-PECEGWKYWITLSTMWKNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
BE/1/2010  401 WT-PECEGWKYWTTLSIMWKNRRSDLLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
BE/10/2011 401 WT-PECEGWKYWTTLSIMWKNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
PAV20      401 WT-PKCDGWKYWITLSMMWRNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
PAV18      401 WT-PECEGWKYWTTLSIMWKNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
VR1814     401 WT-PKCDGWKYWITLSMMWRNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGG 479
3157       401 WT-PQCEGWKYWITLSMMWRNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 479
JP         401 WT-PKCDGWKYWITLSMMWRNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGG 479
Merlin     398 WTLQQCQGWKYWTTLSIMWKNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGS 477

ZMB240     480  SSTKNDVHRIV  490
HAN40      480  SSTKNDVHRIV  490
HAN21      480  SSTKNDVHRIV  490
BE/29/2010 480  SSTKNDVHRIV  490
BE/1/2010  480  SSTKNDVHRIV  490
BE/10/2011 480  SSTKNDVHRIV  490
PAV20      480  SSTKNDVHRIV  490
PAV18      480  SSTKNDVHRIV  490
VR1814     480  SSTKNDVHRIV  490
3157       480  SSTKNDVHRIV  490
JP         480  SSTKNDVHRIV  490
Merlin     478  SSTKNDVHWIV  488
```

Figure 5.25. ZMB240 UL37 amino acid alignment. The sequence has three amino acids inserted after position 340, giving an additional glycosylation site (NVT). Further, one residue is deleted at position 400. These sequence differences – relative to HCMV strain Merlin AY446894.2 – are highlighted by the red borders, and also present in the 10 HCMV strains included in the alignment.

```
ZMB240     1 MEPTPMLRDRDHDDAPPTYEQAMGLCPTTVSTPPPPPPPDCSPPPYRPPYCLVSSPSPRHTFDMDMMEMPATMHPTTGAYF 80
BE/11/2010 1 MEPTPMLRDRDHDDAPPTYEQAMGLCPTTVSTPPPPPPPDCSPPPYRPPYCLVSSPSPRHTFDMDMMEMPATMHPTTGAYF 80
Merlin     1 MEPTPMLRDRDHDDAPPTYEQAMGLCPTTVSTPPPPPPPDCSPPPYRPPYCLVSSPSPRHTFDMDMMEMPATMHPTTGAYF 80

ZMB240     81 DNGWKWTFALLVVAILGIIFLAVVFTVVINRDNST-TTGT--TSG 122
BE/11/2010 81 DNGWKWTFALLVVAILGIIFLAVVFTVVINRDNST-TTGT--TSG 122
Merlin     81 DNGWKWTFALLVVAILGIIFLAVVFTVVINRDSANITTGTQASSG 125
```

Figure 5.26. ZMB242 UL42 amino acid alignment. There are three deletions in the ZMB240 sequence at positions 116, 121, and 122, highlighted by the red borders, relative to HCMV strain Merlin AY446894.2. The deletions are also present in HCMV strain BE/11/2010, as shown here.

```
ZMB240 1    MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKKLPAGGRLPVY  80
Towne  1    MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKKLPAGGRLPVY  80
3157   1    MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKKLPAGGRLPVY  80
HAN20  1    MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKKLPAGGRLPVY  80
VR1814 1    MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKKLPAGGRLPVY  80
U8     1    MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKKLPAGGRLPVY  80
Merlin 1    MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKKLPAGGRLPVY  80

ZMB240 81   RLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFD  160
Towne  81   RLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFD  160
3157   81   RLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFD  160
HAN20  81   RLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFD  160
VR1814 81   RLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFD  160
U8     81   RLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFD  160
Merlin 81   RLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFD  160

ZMB240 161  PHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD  240
Towne  161  PHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD  240
3157   161  PHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD  240
HAN20  161  PHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD  240
VR1814 161  PHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD  240
U8     161  PHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD  240
Merlin 161  PHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD  240

ZMB240 241  TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPDKRRKDLSGSKHGGKKKPPSTTSKTLATASSS-SA  319
Towne  241  TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPEKRRKDLSGSKHGGKKKPPSTTSKTLATASSS-SA  319
3157   241  TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPEKRRKDLSGSKHGGKKKPPSTTSKTLATASSS-SA  319
HAN20  241  TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPEKRRKDLSGSKHGGKKKPPSTTSKTLATASSS-SA  319
VR1814 241  TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPEKRRKDLSGSKHGGKKKPPSTTSKTLATASSS-SA  319
U8     241  TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPEKRRKDLSGSKHGGKKKPPSTTSKTLATASSS-SA  319
Merlin 241  TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPEKRRKDLSGSKHGGKKKPPSTTSKTLATASSSPSA  320

ZMB240 320  IAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLT  399
Towne  320  IAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLT  399
3157   320  IAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLT  399
HAN20  320  IAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLT  399
VR1814 320  IAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLT  399
U8     320  IAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLT  399
Merlin 321  IAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLT  400
```

Figure 5.27. ZMB240 UL48 amino acid alignment. There is an amino acid deletion at position 318 (red border) relative to HCMV strain Merlin AY446894.2. This deletion is also present in 61 full genome HCMV strains (table 5.10), five of which are included in this alignment. Only the first 400 of 2241 amino acid residues are shown here.

```
ZMB240      1   MEINKVLHQDLVQATRRILKLGPSELRVTDAGLICKNPNYSVCDAMLKTDTVYCVEYLLSYWESRTDHVPCFIFKNTGCA 80
AD169       1   MEMNKVLHQDLVQATRRILKLGPSELRVTDAGLICKNPNYSVCDAMLKTDTVYCVEYLLSYWESRTDHVPCFIFKNTGCA 80
BE/46/2011 1   MEMNKVLHQDLVQATRRILKLGPSELRVTDAGLICKNPNYSVCDAMLKTDTVYCVEYLLSYWESRTDHVPCFIFKNTGCA 80
BE/10/2010 1   MEMNKVLHQDLVQATRRILKLGPSELRVTDAGLICKNPNYSVCDAMLKTDTVYCVEYLLSYWESRTDHVPCFIFKNTGCA 80
BE/9/2010   1   MEMNKVLHQDLVQATRRILKLGPSELRVTDAGLICKNPNYSVCDAMLKTDTVYCVEYLLSYWESRTDHVPCFIFKNTGCA 80
Merlin      1   MEMNKVLHQDLVQATRRILKLGPSELRVTDAGLICKNPNYSVCDAMLKTDTVYCVEYLLSYWESRTDHVPCFIFKNTGCA 80

ZMB240      81  VSLCCFVRAPVKLVSPARHVGEFNVLKVNESLIVTLKDIEEIKPSAYGVLTKCVVRKSNSASVFNIELIAFGPENEGEYE 160
AD169       81  VSLCCFVRAPVKLVSPARHVGEFNVLKVNESLIVTLKDIEEIKPSAYGVLTKCVVRKSNSASVFNIELIAFGPENEGEYE 160
BE/46/2011 81  VSLCCFVRAPVKLVSPARHVGEFNVLKVNESLIVTLKDIEEIKPSAYGVLTKCVVRKSNSASVFNIELIAFGPENEGEYE 160
BE/10/2010 81  VSLCCFVRAPVKLVSPARHVGEFNVLKVNESLIVTLKDIEEIKPSAYGVLTKCVVRKSNSASVFNIELIAFGPENEGEYE 160
BE/9/2010   81  VSLCCFVRAPVKLVSPARHVGEFNVLKVNESLIVTLKDIEEIKPSAYGVLTKCVVRKSNSASVFNIELIAFGPENEGEYE 160
Merlin      81  VSLCCFVRAPVKLVSPARHVGEFNVLKVNESLIVTLKDIEEIKPSAYGVLTKCVVRKSNSASVFNIELIAFGPENEGEYE 160

ZMB240      161 NLLRELYAKKAASTSLAVRNHVTVSSHSGSGPSLWRARMSAALTRTAGKRSSRTAS-PPPPPRHPSCSPTMVAAGGAAAG 239
AD169       161 NLLRELYAKKAASTSLAVRNHVTVSSHSGSGPSLWRARMSAALTRTAGKRSSRTAS-PPPPPRHPSCSPTMVAAGGAAAG 239
BE/46/2011 161 NLLRELYAKKAASTSLAVRNHVTVSSHSGSGPSLWRARMSAALTRTAGKRSSRTAS-PPPPPRHPSCSPTMVAAGGAAAG 239
BE/10/2010 161 NLLRELYAKKAASTSLAVRNHVTVSSHSGSGPSLWRARMSAALTRTAGKRSSRTAS-PPPPPRHPSCSPTMVAAGGAAAG 239
BE/9/2010   161 NLLRELYAKKAASTSFAVRNHVTVSSHSGSGPSLWRARMSAALTRTAGKRSSRTAS-PPPPPRHPSCSPTMVAAGGAAAG 239
Merlin      161 NLLRELYAKKAASTSLAVRNHVTVSSHSGSGPSLWRARMSAALTRTAGKRSPRTASPPPPPPRHPSCSPTMVAAGGAAAG 240

ZMB240      240 PRPPPPPMAAGSWRLCRCEACMGRCGCASEGDADEEEEELLALAGEGKAAAAAAGQDVGGSARRPLEEHVSRRRGVSTHH 319
AD169       240 PRPPPPPMAAGSWRLCRCEACMGRCGCASEGDADEEEEELLALAGEGKAAAAAAGQDVGGSARRPLEEHVSRRRGVSTHH 319
BE/46/2011 240 PRPPPPPMAAGSWRLCRCEACMGRCGCASEGDADEEEEELLALAGEGKAAAAAAGQDVGGSARRPLEEHVSRRRGVSTHH 319
BE/10/2010 240 PRPPPPPMAAGSWRLCRCEACMGRCGCASEGDADEEEEELLALAGEGKAATAAAGQDVGGSARRPLEEHVSRRRGVSTHH 319
BE/9/2010   240 PRPPPPPMAAGSWRLCRCEACMGRCGCASEGDADEEEEELLALAGEGKAAAAAAGQDVGGSARRPLEEHVSRRRGVSTHH 319
Merlin      241 PRPPPPPMAAGSWRLCRCEACMGRCGCASEGDADEEEEELLALAGEGKAAAAAAGQDIGGSARRPLEEHVSRRRGVSTHH 320

ZMB240      320 RHPPSPPCAPSLERTGYRWAPSSWWRARSGPSRPQSGPWLPARFAILGPLVLALLLVLALLWRGHGQSSSPTRSAHRD 397
AD169       320 RHPPSPPCAPSLERTGYRWAPSSWWRARSGPSRPQSGPWLPARFATLGPLVLALLLVLALLWRGHGQSSSPTRSAHRD 397
BE/46/2011 320 RHPPSPPCAPSLERTGYRWAPSSWWRARSGSSRPQSGPWLPARFATLGPLVLALLLVLALLWRGHGQSSSPTRSAHRD 397
BE/10/2010 320 RHPPSPPCAPSLERTGYRWAPSSWWRARSGPSRPQSGPWLPARFATLGPLVLALLLVLALLWRGHGQSSSPTRSAHRD 397
BE/9/2010   320 RHPPSPPCAPSLERTGYRWAPSSWWRARSGPSRPQSGPWLPARFATLGPLVLALLLVLALLWRGHGQSSSPTRSAHRD 397
Merlin      321 RHPPSPPCTPSLERTGYRWAPSSWWRARSGPSRPQSGPWLPARFATLGPLVLALLLVLALLWRGHGQSSSPTRSAHRD 398
```

Figure 5.28. ZMB240 UL50 amino acid alignment. Similar to the four HCMV strains included in this alignment, the ZMB240 UL50 product has a single residue deleted at position 217 (red border) relative to Merlin (AY446894.2).

```
ZMB240 1     -----MIMVKGIPKIMLLISITFLLLSLINCNVLVNSRGTRRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYRFLM  75
JP     1     -----MIMVKGIPKIMLLISITFLLLSLINCNVLVNSRGTRRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYRFLM  75
Merlin 1     MGKKEMIMVKGIPKIMLLISITFLLLSLINCNVLVNSRGTRRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYRFLM  80

ZMB240 76    YPSQQKFHAIVISMDKFPQDYILAGPIRNDSITHMWFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVPSMNCLSE  155
JP     76    YPSQQKFHAIVISMDKFPQDYILAGPIRNDSITHMWFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVPSMNCLSE  155
Merlin 81    YPSQQKFHAIVISMDKFPQDYILAGPIRNDSITHMWFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVPSMNCLSE  160

ZMB240 156   MLNVSKRNDTGEKGCGNFTTFNPMFFNVPRWNTKLYIGSNKVNVDSQTIYFLGLTALLLRYAQRNCTRSFYLVNAMSRNL  235
JP     156   MLNVSKRNDTGEKGCGNFTTFNPMFFNVPRWNTKLYIGSNKVNVDSQTIYFLGLTALLLRYAQRNCTRSFYLVNAMSRNL  235
Merlin 161   MLNVSKRNDTGEKGCGNFTTFNPMFFNVPRWNTKLYIGSNKVNVDSQTIYFLGLTALLLRYAQRNCTRSFYLVNAMSRNL  240

ZMB240 236   FRVPKYINGTKLKNTMRKLKRKQALVKEQPQKKNKKSQSTTTPYLSYTTSTAFNVTTNVTYSATAAVTRVATSTTGYRPD  315
JP     236   FRVPKYINGTKLKNTMRKLKRKQALVKEQPQKKNKKSQSTTTPYLSYTTSTAFNVTTNVTYSATAAVTRVATSTTGYRPD  315
Merlin 241   FRVPKYINGTKLKNTMRKLKRKQALVKEQPQKKNKKSQSTTTPYLSYTTSTAFNVTTNVTYSATAAVTRVATSTTGYRPD  320

ZMB240 316   SNFMKSIMATQLRDLATWVYTTLRYRNEPFCKPDRNRTAVSEFMKNTHVLIRNETPYTIYGTLDMSSLYYNETMSVENET  395
JP     316   SNFMKSIMATQLRDLATWVYTTLRYRNEPFCKPDRNRTAVSEFMKNTHVLIRNETPYTIYGTLDMSSLYYNETMSVENET  395
Merlin 321   SNFMKSIMATQLRDLATWVYTTLRYRNEPFCKPDRNRTAVSEFMKNTHVLIRNETPYTIYGTLDMSSLYYNETMSVENET  400

ZMB240 396   ASDNNETTPTSPSTRFQRTFIDPLWDYLDSLLFLDKIRNFSLQLPAYGNLTPPEHRRAANLSTLNSLWWWSQ  467
JP     396   ASDNNETTPTSPSTRFQRTFIDPLWDYLDSLLFLDKIRNFSLQLPAYGNLTPPEHRRAANLSTLNSLWWWSQ  467
Merlin 401   ASDNNETTPTSPSTRFQRTFIDPLWDYLDSLLFLDKIRNFSLQLPAYGNLTPPEHRRAANLSTLNSLWWWSQ  472
```

Figure 5.29. ZMB240 UL74 amino acid alignment. Compared to HCMV strain Merlin (AY446894.2), the first five amino acids are deleted in HCMV genotype gO5 strains ZMB240 and JP as highlighted by the red border.

```
ZMB240 1     MSLLHTFWRLPVAVFFEPHEENVLRCPERVLRRLLEDAAVAMRGGGWREDVLMDRVRKRYLRQELRDLGHRVQTYCEDLE  80
Towne  1     MSLLHTFWRLPVAVFFEPHEENVLRCPERVLRRLLEDAAVAMRGGGWREDVLMDRVRKRYLRQELRDLGHRVQTYCEDLE  80
Merlin 1     MSLLHTFWRLPVAVFFEPHEENVLRCPERVLRRLLEDAAVAMRGGGWREDVLMDRVRKRYLRQELRDLGHRVQTYCEDLE  80

ZMB240 81    GRVSEAEALLNQQCELDEGPSPRTLLQPPCRPRSSSPGTGVAGASAVPHGLYSRHDAITGP-ATPSDAATASAAAGASST  159
Towne  81    GRVSEAEALLNQQCELDEGPSPRTLLQPPCRPRSSSPGTGVAGASAVPHGLYSRHDAITGP-ATPSDAATASAAAGASST  159
Merlin 81    GRVSEAEALLNQQCELDEGPSPRTLLQPPCRPRSSSPGTGVAGASAVPHGLYSRHDAITGPVAAPSDAVAASAAAGASST  160

ZMB240 160   WLAQCAERPLPGNVPSYFGITQNDPFIRFHTDFRGEVVNTMFENASTWTFSFGIWYYRLKRGLYTQPRWKRVYHLAQMDN  239
Towne  160   WLAQCAERPLPGNVPSYFGITQNDPFIRFHTDFRGEVVNTMFENASTWTFSFGIWYYRLKRGLYTQPRWKRVYHLAQMDN  239
Merlin 161   WLAQCAEQPLPGNVPNYFGITQNDPFIRFHTDFRGEVVNTMFENASTWTFSFGIWYYRLKRGLYTQPRWKRVYHLAQMDN  240

ZMB240 240   FSISQELLLGVVNALENVTVYPTYDCVLSDLEAAACLLAAYGHALWEGRDPPDSVATVLGELPQLLPRLADDVSREIAAW  319
Towne  240   FSISQELLLGVVNALENVTVYPTYDCVLSDLEAAACLLAAYGHALWEGRDPPDSVATVLGELPQLLPRLADDVSREIAAW  319
Merlin 241   FSISQELLLGVVNALENVTVYPTYDCVLSDLEAAACLLVAYGHALWEGRDPPDSVTAVLSELPQLLPRLADDVSREIAAW  320

ZMB240 320   EGPVAAGNNYYAYRDSPDLRYYMPLSGGRHYHPGTFDRHVLVRLFHKRGVIQHLPGYGTITEELVQERLSGQVRDDVLSL  399
Towne  320   EGPVAAGNNYYAYRDSPDLRYYMPLSGGRHYHPGTFDRHVLVRLFHKRGVIQHLPGYGTITEELVQERLSGQVRDDVLSL  399
Merlin 321   EGPVAAGNNYYAYRDSPDLRYYMPLSGGRHYHPGTFDRHVLVRLFHKRGVLQHLPGYGTITEELVQERLSGQVRDDVLSL  400

ZMB240 400   WSRRLLVGKLGRDVPVFVHEQQYLRSGLTCLAGLLLLWKVTNADSVFAPRTGKFTLADLLGSDAVAGGGLPGGRAGGEEE  479
Towne  400   WSRRLLVGKLGRDVPVFVHEQQYLRSGLTCLAGLLLLWKVTNADSVFAPRTGKFTLADLLGSDAVAGGGLPGGRAGGEEE  479
Merlin 401   WSRRLLVGKLGRDVPVFVHEQQYLRSGLTCLAGLLLLWKVTNADSVFAPRTGKFTLADLLGSDAVAGGGLPGGRAGGEEK  480

ZMB240 480   GYGGRHGRVRNFEFLVQYYIGPWYARDPAVTLSQLFPGLALLAVTESVRSGWDPSRREDSAGGGDGGGAVLMQLSKSNPV  559
Towne  480   GYGGRHGRVRNFEFLVQYYIGPWYARDPAVTLSQLFPGLALLAVTESVRSGWDPSRREDSAGGGDGGGAVLMQLSKSNPV  559
Merlin 481   GYGGRHGRVRNFEFLVQYYIGPWYARDPAVTLSQLFPGLALLAVTESVRSGWDPSRREDSAGGGDGGGAVLMQLSKSNPV  560

ZMB240 560   ADYMFAQSSKQYGDLRRLEVHDALLFHYEHGLGRLLSVTLPRHRVSTLGSSLFNVNDIYELLYFLVLGFLPSVAVL  635
Towne  560   ADYMFAQSSKQYGDLRRLEVHDALLFHYEHGLGRLLSVTLPRHRVSTLGSSLFNVNDIYELLYFLVLGFLPSVAVL  635
Merlin 561   ADYMFAQSSKQYGDLRRLEVHDALLFHYEHGLGRLLSVTLPRHRVSTLGSSLFNVNDIYELLYFLVLGFLPSVAVL  636
```

Figure 5.30. ZMB240 UL77 amino acid alignment. Like HCMV strain Towne, ZMB240 UL77 has an amino acid deletion at position 142, highlighted by the red border, relative to Merlin AY446894.2.

```
ZMB240      1    MMAAAVVRAEVRRQRREERKKMAAARTTEDPPENDVVADVACGTGAVTRSSSSSLVVSSSSASGSDESSSASPLSFPVSS  80
BE/6/2012   1    MMAAAVVRAEVRRQRREERKKMAAARTTEDPPENHVVADVACGTGAVTRSSSSSLVVSSSSASGSDESSSASPLSFPVCS  80
BE/46/2011  1    MMAAAVVRAEVRRQRREERKKMAVARTTEDPPENHVVADVACGTGAVTRSSSSSLVVSSSSASGSDEPSSASPLSFPVCS  80
BE/1/2010   1    MMAAAVVRAEVRRQRREERKKMAAARTTEDPPENHVVADVVCGTGAVTRSSSSSLVVSSSSASGSDEPSSASPLSFPVCS  80
BE/8/2011   1    MMAAAVVRAEVRRQRREERKKMAAARTTEDPPENHVVADVACGTGAVTRSSSSSLVVSSSSASGSDESSSASPLSFPVSS  80
Towne       1    MMAAAVVRAEVRRQRREERKKMAAARTTEDPPENHVVADVACGTGAVTRSSSSSLVVSSSSASGSDEPSSASPLSFPVCS  80
Merlin      1    MMAAAVVRAEVRRQRREERKKMASARTTEDPPENHVVADVACGTGAVTRSSSSSLVVSSSSASGSDESSSASPLSFPVSS  80

ZMB240      81   PSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDRFRLIETPDENFLLVTNVI  160
BE/6/2012   81   PSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDRFRLIETPDENFLLVTNVI  160
BE/46/2011  81   PSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDRFRLIETPDENFLLVTNVI  160
BE/1/2010   81   PSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDRFRLIETPDENFLLVTNVI  160
BE/8/2011   81   PSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDRFRLIETPDENFLLVTNVI  160
Towne       81   PSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDRFRLIETPDENFLLVTNVI  160
Merlin      81   PSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDRFRLIETPDENFLLVTNVI  160

ZMB240      161  PRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKNDAATPSFLRRHDVLERFAAAAEPLPS  240
BE/6/2012   161  PRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKNDAATPSFLRRHDVLERFAAAAEPLPS  240
BE/46/2011  161  PRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKNDAATPSFLRRHDVLERFAAAAEPLPS  240
BE/1/2010   161  PRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKNDAATPSFLRRHDVLERFAAAAEPLPS  240
BE/8/2011   161  PRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKNDAATPSFLRRHDVLERFAAAAEPLPS  240
Towne       161  PRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKNDAATPSFLRRHDVLERFAAAAEPLPS  240
Merlin      161  PRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKNDAATPSFLRRHDVLERFAAAAEPLPS  240

ZMB240      241  LCVRDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEAVLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSF  320
BE/6/2012   241  LCVRDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEAVLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSF  320
BE/46/2011  241  LCVRDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEAVLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSF  320
BE/1/2010   241  LCVRDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEAVLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSF  320
BE/8/2011   241  LCVRDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEAVLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSF  320
Towne       241  LCVRDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEAVLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSF  320
Merlin      241  LCVHDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEAVLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSF  320

ZMB240      321  CARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLAELYVFVWYRGYEFPAPTPQATTA------GSGGGGGAGACAV  394
BE/6/2012   321  CARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLAELYVFVWYRGYEFPAPTPQATTA------GSGGGGGAGACAV  394
BE/46/2011  321  CARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLAELYVFVWYRGYEFPAPTPQATTA------GSGGGGGAGACAV  394
BE/1/2010   321  CARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLAELYVFVWYRGYEFPAPTPQATTA------GSGGGGGAGACAV  394
BE/8/2011   321  CARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLAELYVFVWYRGYEFPAPTPQATTA------GSGGGGGAGACAV  394
Towne       321  CARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLAELYVFVWYRGYEFPAPTPQATTA------GSGGGGGAGACAV  394
Merlin      321  CARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLAELYVFVWYRGYEFPAPTPQATTAGGGGGGGSGGGGGAGACAV  400
```

```
ZMB240      395  ETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQAVRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYS  474
BE/6/2012   395  ETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQAVRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYS  474
BE/46/2011  395  ETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQAVRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYS  474
BE/1/2010   395  ETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQAVRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYS  474
BE/8/2011   395  ETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQAVRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYS  474
Towne       395  ETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQAVRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYS  474
Merlin      401  ETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQAVRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYS  480

ZMB240      475  PIVIQDCAAAVTDVTLSHILPGQSTVSLFPVYHVGKLLDALSLNDAGLITLNL  527
BE/6/2012   475  PIVIQDCAAAVTDVTLSHILPGQSTVSLFPVYHVGKLLDALSLNDAGLITLNL  527
BE/46/2011  475  PIVIQDCAAAVTDVTLSHILPGQSTVSLFPVYHVGKLLDALSLNDAGLITLNL  527
BE/1/2010   475  PIVIQDCAAAVTDVTLSHILPGQSTVSLFPVYHVGKLLDALSLNDAGLITLNL  527
BE/8/2011   475  PIVIQDCAAAVTDVTLSHILPGQSTVSLFPVYHVGKLLDALSLNDAGLITLNL  527
Towne       475  PIVIQDCAAAVTDVTLSHILPGQSTVSLFPVYHVGKLLDALSLNDAGLITLNL  527
Merlin      481  PIVIQDCAAAVTDVTLSHILPGQSTVSLFPVYHVGKLLDALSLNDAGLITLNL  533
```

Figure 5.31. ZMB242 UL95 amino acid alignment. Relative to HCMV strain Merlin AY446894.2, ZMB240 UL95 has 6 deletions spanning positions 382-387, highlighted by the red border. Five other HCMV strains have this deletion as shown in this alignment.

```
ZMB240      1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
HAN40       1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
HAN11       1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
HAN2        1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
HAN1        1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
2CEN2       1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
UKNEQAS2    1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
BE/4/2011   1    MTPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
BE/23/2010  1    MAPSHVDKVNTRTWSASILFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
BE/14/2011  1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
HAN13       1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80
Merlin      1    MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCY  80

ZMB240      81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
HAN40       81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
HAN11       81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
HAN2        81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
HAN1        81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
2CEN2       81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
UKNEQAS2    81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
BE/4/2011   81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
BE/23/2010  81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
BE/14/2011  81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
HAN13       81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160
Merlin      81   AVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVH  160

ZMB240      161  FFCLTIFNVSMVTQYRSYKRSLFFFSRLHPKLKGTVQFRTLIVNLVEVALGFNTTVLAMALCYGFGNNFFVRTGHMVLAV  240
HAN40       161  FFCLTIFNVSMVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
HAN11       161  FFCLTIFNVSMVTQYRSYKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
HAN2        161  FFCLTIFNVSIVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
HAN1        161  FFCLTIFNVSMVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
2CEN2       161  FFCLTIFNVSIVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
UKNEQAS2    161  FFCLTIFNVSMVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
BE/4/2011   161  FFCLTIFNVSMVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
BE/23/2010  161  FFCLTIFNVSMVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVIAMALCYGFGNNFFVRTGHMVLAV  240
BE/14/2011  161  FFCLTIFNVSMVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
HAN13       161  FFCLTIFNVSMVTQYRSFKRSLFFFSRLHPKLKGTVQFRTLIVNLVEMALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240
Merlin      161  FFCLTIFNVSMVTQYRSYKRSLFFFSRLHPKLKGTVQFRTLIVNLVEVALGFNTTVVAMALCYGFGNNFFVRTGHMVLAV  240

ZMB240      241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
HAN40       241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
HAN11       241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
HAN2        241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
HAN1        241  FVVYAIISIIYFLLIEAIFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
2CEN2       241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
UKNEQAS2    241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
BE/4/2011   241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
BE/23/2010  241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
BE/14/2011  241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
HAN13       241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAV  320
Merlin      241  FVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYDTF-LSNEYRTGISWSFGMLFFIWAMFTTCRAV  319
```

```
ZMB240      321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
HAN40       321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDDDFEDA  373
HAN11       321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
HAN2        321 RYFRGRGSSSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
HAN1        321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
2CEN2       321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
UKNEQAS2    321 RYFRGRGSGSVKYQALATASSEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
BE/4/2011   321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDDDFEDA  373
BE/23/2010  321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
BE/14/2011  321 RYFRGRGSGSVKYQALATASGEEVAALSHHNSLESRRLREEEDDDDDEDFEDA  373
HAN13       321 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA  373
Merlin      320 RYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEE-DDDDEDFEDA  371
```

Figure 5.32. ZMB242 UL100 amino acid alignment. Relative to HCMV strain Merlin AY446894.2, ZMB240 UL100 has two insertions after positions 290 and 361, highlighted by red borders. These insertions are also present in 10 other HCMV strains included in the alignment.

```
ZMB240 1     MTAQPPLHHRHHPYTLFGTSCHLSWYGLLEASVPIVQCLFLDLGGGRAEPRLHTFVVRGDRLPPAEVRAVHRASYAALAS  80
HAN28  1     MTAQPPLHHRHHPYTLFGTSCHLSWYGLLEASVPIVQCLFLDLGGGRAEPRLHTFVVRGDRLPPAEVRVVHRASYAALAS  80
Merlin 1     MTAQPPLHHRHHPYTLFGTSCHLSWYGLLEASVPIVQCLFLDLGGGRAEPRLHTFVVRGDRLPPAEVRAVHRASYAALAS  80

ZMB240 81    AVTTDADERRRGLEQRSAVLARVLLEGSALIRVLARTFTPVQIQTDASGVEILEAAPALGVETAALSNALSLFHVAKLVV  160
HAN28  81    AVTTDADERRRGLEQRSAVLARVLLEGSALIRVLARTFTPVQIQTDASGVEILEAAPALGVETAALSNALSLFHVAKLVV  160
Merlin 81    AVTTDADERRRGLEQRSAVLARVLLEGSALIRVLARTFTPVQIQTDASGVEILEAAPALGVETTALSNALSLFHVAKLVV  160

ZMB240 161   IGSYPEVHEPRVVTHAAERVSEEYGTHAHKKLRRGYYAYDLAMSFRVGTHKYVLERDDEAVLARLFEVREVCFLRTCLRL  240
HAN28  161   IGSYPEVHEPRVVTHAAERVSEEYGTHAHKKLRRGYYAYDLAMSFRVGTHKYVLERDDEAVLARLFEVREVCFLRTCLRL  240
Merlin 161   IGSYPEVHEPRVVTHAAERVSEEYGTHAHKKLRRGYYAYDLAMSFRVGTHKYVLERDDEAVLARLFEVREVCFLRTCLRL  240

ZMB240 241   VTPVGFVAVAVTDEQCCLLLQSAWTHLYDVLFRGFAGQPPLRDYLGPDLFETGAARSFFFPGFPPVPVYAVHGLHTLMRE  320
HAN28  241   VTPVGFVAVAVTDEQCCLLLQSAWTHLYDVLFRGFAGQPPLRDYLGPDLFETGAARSFFFPGFPPVPVYAVHGLHTLMRE  320
Merlin 241   VTPVGFVAVAVTDEQCCLLLQSAWTHLYDVLFRGFAGQPPLRDYLGPDLFETGAARSFFFPGFPPVPVYAVHGLHTLMRE  320

ZMB240 321   TALDAAAEVLSWCGLPDIVGSAGKLEVEPCALSLGVPEDEWQVFGTEAGGGAVRLNATAFRERPAGGDRRWLLPPLPRDD  400
HAN28  321   TALDAAAEVLSWCGLPDIVGSAGKLEVEPCALSLGVPEDEWQVFGTEAGGGAVRLNATAFRERPAGGDRRWLLPPLPRDD  400
Merlin 321   TALDAAAEVLSWCGLPDIVGSAGKLEVEPCALSLGVPEDEWQVFGTEAGGGAVRLNATAFRERPAGGDRRWLLPPLPRDD  400

ZMB240 401   GDGENNVVEVSSSTGGAHPPSDDATFTVHVRDATLHRVLIVDLVERVLAKCVRARDFNPYVRYSHRLHTYAVCEKFIENL  480
HAN28  401   GDGENNVVEVSSSTGGAHPPSDDATFTVHVRDATLHRVLIVDLVERVLAKCVRARDFNPYVRYSHRLHTYAVCEKFIENL  480
Merlin 401   GDGENNVVEVSSSTGGAHPPSDDATFTVHVRDATLHRVLIVDLVERVLAKCVRARDFNPYVRYSHRLHTYAVCEKFIENL  480

ZMB240 481   RFRSRRAFWQIQSLLGYISEHVTSACASAGLLWVLSRGHREFYVYDGYSGHGPVSAEVCVRTVVDCYWRKLFGGDDPGPT  560
HAN28  481   RFRSRRAFWQIQSLLGYISEHVTSACASAGLLWVLSRGHREFYVYDGYSGHGPVSAEVCVRTVVDCYWRKLFGGDDPGPT  560
Merlin 481   RFRSRRAFWQIQSLLGYISEHVTSACASAGLLWVLSRGHREFYVYDGYSGHGPVSAEVCVRTVVDCYWRKLFGGDDPGPT  560

ZMB240 561   CRVQESAPGVLLVWGDERLVGPFNFFYGNGGAGGSPLHGVVGGFAAGHCGGACCAGCVVTHRHSSGGGGSGVGDADHASG  640
HAN28  561   CRVQESAPGVLLVWGDERLVGPFNFFYGNGGAGGSPLHGVVGGFAAGHCGGACCAGCVVTHRHSSGGGGSGVGDADHASG  640
Merlin 561   CRVQESAPGVLLVWGDERLVGPFNFFYGNGGAGGSPLHGVVGGFAAGHCGGACCAGCVVTHRHSSGGGGSGVGDADHASG  640

ZMB240 641   GGLDAAAGSGHNGGSDRVSPSTPPAALGGCCCAAGGDWLSAVGHVLGRLPALLRERVSVSELEAVYREILFRFVARRNDV  720
HAN28  641   GGLDAAAGSGHNGGSDRVSPSTPPAALGGCCCAAGGDWLSAVGHVLGRLPALLRERVSVSELEAVYREILFRFVARRNDV  720
Merlin 641   GGLDAAAGSGHNGGSDRVSPSTPPAALGGCCCAAGGDWLSAVGHVLGRLPALLRERVSVSELEAVYREILFRFVARRNDV  720

ZMB240 721   DFWLLRFQPGENEVRPHAGVIDCAPFHGVWAEQGQIIVQSRDTALAADIGYGVYVDKAFAMLTACVEVWARELLSSSTAS  800
HAN28  721   DFWLLRFQPGENEVRPHAGVIDCAPFHGVWAEQGQIIVQSRDTALAADIGYGVYVDKAFAMLTACVEVWARELLSSSTAS  800
Merlin 721   DFWLLRFQPGENEVRPHAGVIDCAPFHGVWAEQGQIIVQSRDTALAADIGYGVYVDKAFAMLTACVEVWARELLSSSTAS  800

ZMB240 801   TTAC-SSSVLSSALPSVTSSSSGTATVSPPSCSSSSATWLEERDEWVRSLAVDAQHAAKRVASEGLRFFRLNA  872
HAN28  801   TTTC-SSSVLSSALPSVTSSSSGTATVSPPSCSSSSATWLEERDEWVRSLAVDAQHAAKRVASEGLRFFRLNA  872
Merlin 801   TTTCSSSSVLSSALPSVTSSSSGTATVSPPSCSSSSATWLEERDEWVRSLAVDAQHAAKRVASEGLRFFRLNA  873
```

Figure 5.33. ZMB242 UL102 amino acid alignment. Compared to HCMV strain Merlin AY446894.2, ZMB240 and HCMV strain HAN28 contain a deletion at position 805, highlighted here by the red border.

```
ZMB240 1     MDLPTTVVRKYWTFANPNRILHQSVNQTFDVRQFVFDTARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFM  80
AD169  1     MDLPTTVVRKYWTFANPNRILHQSVNQTFDVRQFVFDTARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFM  80
6397   1     MDLPTTVVRKYWTFANPNRILHQSVNQTFDVRQFVFDTARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFM  80
AF1    1     MDLPTTVVRKYWTFTNPNRILHQSVNQTFDVRQFVFDNARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFM  80
U11    1     MDLPTTVVRKYWTFTNPNRILHQSVNQTFDVRQFVFDNARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFM  80
3301   1     MDLPTTVVRKYWTFTNPNRILHQSVNQTFDVRQFVFDNARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFM  80
Merlin 1     MDLPTTVVRKYWTFTNPNRILHQSVNQTFDVRQFVFDNARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFM  80

ZMB240 81    SIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSGGGGGCGAASTASA  160
AD169  81    SIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSSGGGGCGAASTASA  160
6397   81    SIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSGGGGGCGAASTASA  160
AF1    81    SIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSSGGGGCGAASTASA  160
U11    81    SIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSSGGGGCGAASTASA  160
3301   81    SIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSSGGGGCGAASTASA  160
Merlin 81    SIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSGGGGGCGAASTASA  160

ZMB240 161   VCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGAGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQ  240
AD169  161   VCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGAGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQ  240
6397   161   VCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGAGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQ  240
AF1    161   VCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGAGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQ  240
U11    161   VCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGAGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQ  240
3301   161   VCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGAGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQ  240
Merlin 161   VCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGASGGGGGGGSGGQHSSDSANGLLRDPRLMNRQ  240

ZMB240 241   KERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAV  320
AD169  241   KERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAV  320
6397   241   KERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGCMSLPLDTSEAVAFLNYSSSSSAV  320
AF1    241   KERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAV  320
U11    241   KERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAV  320
3301   241   KERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAV  320
Merlin 241   KERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAV  320

ZMB240 321   SSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPGSAHSAGETVQESEAAATAAAAGLM  400
AD169  321   SSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPSSAHSAGETVQESEAAATAAAAGLM  400
6397   321   SSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPSSAHSAGETVQESEAAATAAAAGLM  400
AF1    321   SSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPSSAHSAGETVQESEAAATAAAAGLM  400
U11    321   SSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPSSAHSAGETVQESEAAATAAAAGLM  400
3301   321   SSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPSSAHSAGETVQESEAAATAAAAGLM  400
Merlin 321   SSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPSSAHSAGETVQESEAAATAAAAGLM  400

ZMB240 401   MMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIQSLTKMLNDCKEKRLCDLPLVSSR  480
AD169  401   MMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIQSLTKMLNDCKEKRLCDLPLVSSR  480
6397   401   MMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIHSLTKMLNDCKEKRLCDLPLVSSR  480
AF1    401   MMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIQSLTKMLNDCKEKRLCDLPLVSSR  480
U11    401   MMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIQSLTKMLNDCKEKRLCDLPLVSSR  480
3301   401   MMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIQSLTKMLNDCKEKRLCDLPLVSSR  480
Merlin 401   MMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIQSLTKMLNDCKEKRLCDLPLVSSR  480

ZMB240 481   LLPETSGGTVVVNHSSVARTAAAVSTAGVGPPAAACPPLVTTGVVPSGSVAGVAPVAAAVETPVAPPRPVCEIKPYVVNP  560
AD169  481   LLPETSGGTVVVNHSSVARTAAAVSAAGVGPPAAACPPLVTTGVVPSGSVAGVAPVAAAIETPAAPPRPVCEIKPYVVNP  560
6397   481   LLPETSGGTVVVNHSSVARTAAAVSTAGVGPPAAACPPLVTTGVVPSGSVAGVAPVAAAVETPAAPPRPVCEIKPYVVNP  560
AF1    481   LLPETSGGTVVVNHSSVARTAAAVSTAGVGPPAAACPPLVTTGVVPSGSVAGVAPVAAAVETPAAPPRPVCEIKPYVVNP  560
U11    481   LLPETSGGTVVVNHSSVARTAAAVSTAGVGPPAAACPPLVTTGVVPSGSVAGVAPVAAAVETPAAPPRPVCEIKPYVVNP  560
3301   481   LLPETSGGTVVVNHSSVARTAAAVSTAGVGPPAAACPPLVTTGVVPSGSVAGVAPVAAAVETPAAPPRPVCEIKPYVVNP  560
Merlin 481   LLPETSGGTVVVNHSSVARTAAAVSTAGVGPPAVACPPLVTTGVVPSGSVAGVAPVAAAVETPAAPPRPVCEIKPYVVNP  560

ZMB240 561   VVATAAAASNSSSSSSAPLPPPPPPPSGRRGRARNNTR-GGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHEDEEDNDP  639
AD169  561   VVATAAAASNSSSSSSAPLPPPPPPSGGRRGRARNNTR-GGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHEDEEDNDP  639
6397   561   VVATAAAASNSSSSSSAPLPPPPPPPGGRRGRARNNTR-GGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHEDEEDNDP  639
AF1    561   VVATAAAASNSSSSSSAPLPPPPPPPGGRRGRARNNTR-GGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHEDEEDNDP  639
U11    561   VVATAAADSNSSSSSSAPLPPPPPPPGGRRGRARNNTR-GGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHEDEEDNDP  639
3301   561   VVATAAAASNSSSSSSAPLPPPPPPPGGRRGRARNNTR-GGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHEDEEDNDP  639
Merlin 561   VVATAAAASNSSSSSSAPLPPPPPPPGGRRGRARNNTRGGGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHED-EDNDP  639

ZMB240 640   LLRLSQVAGSGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD  684
AD169  640   LLRLSQVAGNGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD  684
6397   640   LLRLSQVAGSGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD  684
AF1    640   LLRLSQVAGSGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD  684
U11    640   LLRLSQVAGSGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD  684
3301   640   LLRLSQVAGSGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD  684
Merlin 640   LLRLSQVAGSGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD  684
```

Figure 5.34. ZMB240 UL112 amino acid alignment. The ZMB240 UL112 product has a deletion at position 599 and an insertion after position 634 relative to Merlin (AY446894.2). These changes are also present in 34 full genome HCMV strains (table 5.10), five of which are included in this alignment.

```
ZMB240 1     MVMFSQDHVQIVYGSTRICKSLAPANKRKTHRTIVVAPRRGFLRIPPDGQDVNHVKIVPTT-SSSLAPPRDDERRPTPPL  79
Towne  1     MVMFSQDHVQIVYGSTRICKSLAPANKRKTHRTIVVAPRRGFLRIPPDGQDVNHVKIVPTT-SSSLAPPRDDERRPTPPL  79
Merlin 1     MVMFSQDHVQIVYGSTRICKSLAPANKRKTHRTIVVAPRRGFLRIPPDGQDVNHVKIVPTTTSSSLAPPRDDERRPTPPL  80

ZMB240 80    RPPLTVYPYGTSLIRRSARDAKLRSKLIVFHITRPALGQHPQNPGISGPAAMDHSEFLTSFRREVDRQTVLTAESAPATV  159
Towne  80    RPPLTVYPYGTSLIRRSARDAKLRSKLIVFHITRPALGQHPQNPGISGPAAMDHSEFLTSFRREVDRQTVLTAESAPATV  159
Merlin 81    RPPLTVYPYGTSLIRRSARDAKLRSKLIVFHITRPALGQHPQNPGISGPAAMDHSEFLTSFRREVDRQTVLTAESAPATV  160

ZMB240 160   EVCLGDALPGGVMGGGGLPAGVGSASAAVAAAAAAVAGVPVAANPVMPATATVTTPPMIDLTSHHRPLTLFTPASAAAAP  239
Towne  160   EVCLGDALPGGVMGGGGLPAGVGSASAAVAAAAAAVAGVPVAANPVMPATATVTTPPMIDLTSHHRPLTLFTPASAAAAP  239
Merlin 161   EVCLGDALPGGVMGGGGLPAGVGSASAAVAAAAAAVAGVPVAANPVMPATATVTTPPMIDLTSHHRPLTLFTPASAAAAP  240

ZMB240 240   AVATNGGNATYILPADCRYAPLFASKYKYVFEEVSRLMRLHDSTAVQLQISASCGNAFQALKSALLKLHNVTVLAGQQLI  319
Towne  240   AVATNGGNATYILPADCRYAPLFASKYKYVFEEVSRLMRLHDSTAVQLQISASCGNAFQALKSALLKLHNVTVLAGQQLI  319
Merlin 241   AVATNGGNATYILPADCRYAPLFASKYKYVFEEVSRLMRLHDSTAVQLQISASCGNAFQALKSALLKLHNVTVLAGQQLI  320

ZMB240 320   TQTMPHTPQAVATFKFFHQDPNRVLDCIRPVVPRSTSYHETGVYQMWVSGATKKDLFDAVTLCASIVEKQPDVFNINVSL  399
Towne  320   TQTMPHTPQAVATFKFFHQDPNRVLDCIRPVVPRSTSYHETGVYQMWVSGATKKDLFDAVTLCASIVEKQPDVFNINVSL  399
Merlin 321   TQTMPHTPQAVATFKFFHQDPNRVLDCIRPVVPRSTSYHETGVYQMWVSGATKKDLFDAVTLCASIVEKQPDVFNINVSL  400

ZMB240 400   LTYPSIAAPHLPLYNEFTSFRLPTS  424
Towne  400   LTYPSIAAPHLPLYNEFTSFRLPTS  424
Merlin 401   LTYPSIAAPHLPLYNEFTSFRLPTS  425
```

Figure 5.35. ZMB240 protein UL117 amino acid alignment. Like HCMV strain Towne, ZMB240 protein UL117 has an amino acid deletion at position 607 (highlighted by the red border) relative to HCMV strain Merlin (AY446894.2).

```
ZMB240      1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTFTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
Toledo      1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTFTTSINTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
PAV11       1    MCPALAITLAAALLSNTHPGMGSSTKSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/17/2010  1    MYPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/26/2010  1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/4/2011   1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/15/2012  1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
CZ/3/2012   1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/11/2012  1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/30/2011  1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/7/2012   1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/16/2012  1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/13/2010  1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
BE/21/2010  1    MCPALAIALAAALLSNTHPGMGSSTKSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
U8          1    MCPALAIALAAALLSNTHPGMGSSTTSAVTSP-NITVTSTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVATTTQ 79
Merlin      1    MCSVLAIALVVALLGDMHPGVKSSTTSAVTSPSNTTVTSTTSISTSNNVSSAVTTTVQTST-------SSASTSVIATTQ 73

ZMB240      80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFIQVGYLSAFPSDDKGKLHLSYNATAQE 159
Toledo      80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
PAV11       80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/17/2010  80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/26/2010  80   KEGHLYNVNCETSYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/4/2011   80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/15/2012  80   KEGHLYNVNCETSYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
CZ/3/2012   80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/11/2012  80   KEGHLYNVNCETSYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/30/2011  80   KEGHLYNVNCETSYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/7/2012   80   KEGHLYNVNCETSYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/16/2012  80   KEGHLYNVNCETSYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/13/2010  80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
BE/21/2010  80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
U8          80   KEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFTQVGYLSAFPSDDKGKLHLSYNATAQE 159
Merlin      74   KEGHLYTVNCEASYSYDQVSLNATCKVILLNNTKNPDILSVTCYARTDCKGPFTQVGYLSAFPSNDKGKLHLSYNATAQE 153
```

```
ZMB240     160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
Toledo     160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
PAV11      160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/17/2010 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/26/2010 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/4/2011  160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/15/2012 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
CZ/3/2012  160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/11/2012 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/30/2011 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/7/2012  160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/16/2012 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/13/2010 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
BE/21/2010 160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
U8         160 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 239
Merlin     154 LLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCLLNGSSLYHPNSTVHLHQG 233

ZMB240     240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
Toledo     240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
PAV11      240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/17/2010 240 DQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/26/2010 240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/4/2011  240 DQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFDDRLLAYGVLAFLVFMVIILLYV 319
BE/15/2012 240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
CZ/3/2012  240 DQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/11/2012 240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/30/2011 240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/7/2012  240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/16/2012 240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/13/2010 240 DQLIPPWNISNVTYNGQRLREFVFYINGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
BE/21/2010 240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
U8         240 NQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 319
Merlin     234 DQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDPLFEDRLLAYGVLAFLVFMVIILLYV 313

ZMB240     320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
Toledo     320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
PAV11      320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/17/2010 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/26/2010 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/4/2011  320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/15/2012 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
CZ/3/2012  320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/11/2012 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/30/2011 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/7/2012  320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/16/2012 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/13/2010 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
BE/21/2010 320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
U8         320  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  351
Merlin     314  TYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW  345
```

Figure 5.36. ZMB240 UL119 amino acid alignment. ZMB240 protein UL119 has an amino acid deletion at position 35 and a block of six residues inserted after position 61 (highlighted by the red borders) relative to HCMV strain Merlin (AY446894.2). These changes are also present in 14 other full genome HCMV strains as shown here.

5.4.3 Non-coding RNAs

In addition to mRNAs that encode proteins, HCMV also encodes numerous non-coding RNAs (ncRNAs). ncRNAs include long non-coding RNAs (lncRNAs) and microRNAs (miRNAs). The ZMB240 genome has at least 30 non-coding RNA regions (table 5.9) which comprise 4 lncRNAs and 26 microRNAs.

5.4.3.1 Long non-coding RNAs (lncRNAs)

Long non-coding RNAs (lncRNAs) are found in every life form as a loosely classified diverse group of abundantly transcribed RNA molecules. Generally, lncRNAs are described as RNA transcripts >200 nucleotides in length, that do not encode proteins (Quinn and Chang, 2016, Mattick and Rinn, 2015). Broadly, lncRNAs are important regulators of gene expression, with a wide range of functions in cellular and developmental processes including imprinting genomic loci, shaping chromosome conformation and allosterically regulating enzymatic activity (Quinn and Chang, 2016). HCMV lncRNAs include RNA2.7, RNA1.2, RNA4.9 and RNA5.0 (Kulesza and Shenk, 2004, McDonough et al., 1985, Rawlinson and Barrell, 1993, Hutchinson and Tocci, 1986, Gatherer et al., 2011). Although they are polyadenylated and can be expressed at very high levels in lytic infection (Kulesza and Shenk, 2004, Gatherer et al., 2011), so far these lncRNAs do not appear to play protein coding functions. There is however a report suggesting the possibility of polypeptides being produced from lncRNAs through non-conventional translation (Ingolia et al., 2014). Functions of the HCMV lncRNAs are largely unclear, but in general involve targeting of specific gene promoters mediating transcriptional silencing. This includes the inhibition of apoptosis (RNA2.7) (Reeves et al., 2007), virulence (RNA5.0) (Kulesza and Shenk, 2006), and transcriptional repression of viral IE gene expression during latency (RNA4.9) (Noriega et al., 2014, Rossetto et al., 2013). Recently, a therapeutic role for RNA2.7 in Parkinson’s disease has been proposed (Poole et al., 2016). All four classic HCMV lncRNA regions are present in the genomes of both HCMV strains Merlin and ZMB240.

5.4.3.2 MicroRNAs (miRNAs)

MicroRNAs (miRNAs) are a family of single-stranded noncoding RNA molecules of approximatively 20-23 nucleotides length present in many multicellular organisms (Bartel, 2004, He and Hannon, 2004, Dunn et al., 2005). By binding to specific mRNA targets whose translation is consequently inhibited, miRNAs play vital roles in the post-transcriptional regulation of gene expression (Ambros, 2004). HCMV is known to also encode at least 26 miRNAs (Dunn et al., 2005, Grey et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012) as curated in the microRNA database, miRBase release 21 (http://www.mirbase.org) (Griffiths-Jones, 2004, Griffiths-Jones et al., 2006, Griffiths-Jones et al., 2008, Kozomara and Griffiths-Jones,

2011, Kozomara and Griffiths-Jones, 2014). HCMV miRNAs are unique among human herpesviruses in that they are spread throughout the genome, unlike in alpha- and gammaherpesviruses, where the miRNA genes are clustered within defined locations within the genome (Buck et al., 2007, Dolken et al., 2007, Dunn et al., 2005, Pfeffer et al., 2005). HCMV miRNAs perform roles in complex processes including regulation of host cell metabolism, immune evasion, and establishment and maintenance of latency (Grey et al., 2007, Stern-Ginossar et al., 2007, Stern-Ginossar et al., 2009, Lau et al., 2016a, Lau et al., 2016b, Poole et al., 2011, Shen et al., 2014). Some HCMV miRNAs have been shown to target both viral and host cell mRNAs (Stern-Ginossar et al., 2009). Interest in miRNAs has heightened in recent times due to their potential for use in diagnostics and as therapeutic targets, as reviewed by Ng (Ng et al., 2015). Here we investigated the HCMV miRNA sequences and show in table 5.11 that all 26 sequences are absolutely conserved in ZMB240. Of these 13 miRNAs are located within intergenic regions, and of 11 occurring within coding regions, nine are antisense to the coding region while two (hcmv-miR-UL29-5p and hcmv-miR-UL29-3p) are in the same sense as the coding region. A further two miRNAs (hcmv-miR-UL36-3p and hcmv-miR-UL36-5p) are encoded within an intron of the UL36 gene. We note that the current annotation of strain Merlin requires updating, as it is based on the outdated miRBase release 16 with only 17 miRNA sequences.

Micro RNA Accession No.	ID	Sequence	Length	Orientation	Coordinates in ZMB240	References
MIMAT0001574	hcmv-miR-UL22A-5p	TAACTAGCCTTCCCGTGAGA	20	Forward	28122–28141	(Dunn et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001575	hcmv-miR-UL22A-3p	TCACCAGAATGCTAGTTTGTAG	22	Forward	28159–28180	(Dunn et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0004754	hcmv-miR-UL36-3p	TTTCCAGGTGTTTTCAACGTG	22	Reverse	c49977–49998	(Grey et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001576	hcmv-miR-UL36-5p	TCGTTGAAGACACCTGGAAAGA	22	Reverse	c50021–50042	(Grey et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0029122	hcmv-miR-UL59	GTTCTCTCGCTCGTCATGCCGT	22	Forward	96028–96049	(Meshesha et al., 2016)
MIMAT0029123	hcmv-miR-UL69	CCAGAGGCTAAGCCGAAACCG	21	Forward	98330–98350	(Meshesha et al., 2016)
MIMAT0003342	hcmv-miR-UL70-5p	TGCGTCTCGGCCTCGTCCAGA	21	Forward	104511–104531	(Grey et al., 2005, Meshesha et al., 2016)
MIMAT0003343	hcmv-miR-UL70-3p	GGGGATGGGCTGGCGCGCGG	20	Forward	104552–104571	(Grey et al., 2005, Meshesha et al., 2016)
MIMAT0026552	hcmv-miR-UL112-5p	CCTCCGGATCACATGGTTACTCA	23	Forward	164588–164610	(Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001577	hcmv-miR-UL112-3p	AAGTGACGGTGAGATCCAGGCT	22	Forward	164625–164646	(Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001578	hcmv-miR-UL148D	TCGTCCTCCCCTTCTTCACCG	21	Forward	193646–193666	(Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0003341	hcmv-miR-US4-5p	TGGACGTGCAGGGGGATGTCTG	22	Forward	201747–201768	(Grey et al., 2005, Meshesha et al., 2016, Stark et al., 2012)
MIMAT0026628	hcmv-miR-US4-3p	TGACAGCCCGCTACACCTCT	20	Forward	201786–201805	(Grey et al., 2005, Meshesha et al., 2016, Stark et al., 2012)
MIMAT0001579	hcmv-miR-US5-1	TGACAAGCCTGACGAGAGCGT	21	Forward	202692–202712	(Grey et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0026553	hcmv-miR-US5-2-5p	CTTTCGCCACACCTATCCTGAAA	24	Forward	202784–202806	(Landgraf et al., 2007, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001580	hcmv-miR-US5-2-3p	TATGATAGGTGTGACGATGTCT	22	Forward	202820–202841	(Landgraf et al., 2007, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0028165	hcmv-miR-US22-5p	TGTTTCAGCGTGTGTCCGCGGG	22	Forward	216476–216497	(Stark et al., 2012)
MIMAT0028166	hcmv-miR-US22-3p	TCGCCGGCCGCGCTGTAACCAGG	23	Forward	216514–216536	(Stark et al., 2012)
MIMAT0004755	hcmv-miR-US25-1-3p	GTCCGAACGCTAGGTCGGTTCT	21	Reverse	c221795–221816	(Dunn et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001581	hcmv-miR-US25-1-5p	AACCGCTCAGTGGCTCGGACC	21	Reverse	c221838–221858	(Dunn et al., 2005, Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001583	hcmv-miR-US25-2-3p	ATCCACTTGGAGAGCTCCCGCGG	24	Reverse	c221999–222021	(Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0001582	hcmv-miR-US25-2-5p	AGCGGTCTGTTCAGGTGGATGA	22	Reverse	c222058–222079	(Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0004756	hcmv-miR-US33-3p	TCACGGTCCGAGCACATCCAA	21	Reverse	c227024–227044	(Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0028167	hcmv-miR-US29-5p	TGGATGTGCTCGGACCGTGACG	22	Forward	227025–227046	(Stark et al., 2012)
MIMAT0001584	hcmv-miR-US33-5p	GATTGTGCCCGGACCGTGGGCG	22	Reverse	c227060–227081	(Landgraf et al., 2007, Meshesha et al., 2016, Pfeffer et al., 2005, Stark et al., 2012)
MIMAT0028168	hcmv-miR-US29-3p	CCCACGGTCCGGGCACAATCA	21	Forward	227062–227082	(Stark et al., 2012)

Table 5.11. HCMV microRNAs found in the ZMB240 genome. All 26 HCMV miRNA sequences characterised in miRBase release 21 (http://www.mirbase.org) were found conserved in the ZMB240 genome. c, complement.

5.4.4 Review of Anti-HCMV Drug Resistance Mutations in ZMB240

Finally, we examined whether the ZMB240 sequence contained mutations associated with anti-HCMV drug resistance. Although the use of anti-HCMV drugs in the Zambian population is presently almost non-existent, conditions such as renal failure and liver failure are becoming more common in the local population and there is an increasing drive for access to organ transplant services. It is therefore essential that HCMV management and related drug resistance issues be considered. For ZMB240 we checked three key HCMV genes – UL97, UL54, and UL27 – that are associated with emergence of resistance to commonly used anti-HCMV drugs.

5.4.4.1 UL97

The UL97 gene product is a serine/threonine kinase incorporated in the mature virion (van Zeijl et al., 1997, Prichard et al., 2005) and has autophosphorylation capacity (He et al., 1997, Michel and Mertens, 2004). The kinase is vital for the initial phosphorylation of GCV and ACV, and is selectively inhibited by MBV. Numerous drug resistance mutations can develop in UL97, as reviewed by Lurain and Chou (Lurain and Chou, 2010) and Campos et al (Campos et al., 2016). These mutations are incorporated in the Mutation Resistance Analyzer (MRA) database (Chevillotte et al., 2010), an online drug resistance genotyping tool available at http://www.informatik.uni-ulm.de/ni/mitarbeiter/HKestler/hcmv. ZMB240 UL97 was screened using the MRA and no drug resistance mutations were found. However two polymorphisms not associated with drug resistance, and two other changes not present in the MRA database were revealed (figure 5.38-A).

5.4.4.2 UL54

UL54 encodes the DNA polymerase, which in conjunction with other components of the viral replication complex, is responsible for viral DNA replication. The polymerase has both nucleotide polymerization and proofreading (3’-5’-exonuclease) activity, and is the ultimate target of the anti-HCMV drugs GCV/vGCV, CDV, and FOS. Drug resistance chiefly results from mutations that prevent drug binding to the enzyme or favour the enzyme’s exonuclease activity (and hence removal of the incorporated drug) (Gilbert and Boivin, 2005, Lurain and Chou, 2010). Screening of ZMB240 UL54 using the MRA

showed no drug resistance mutations, but five polymorphisms not linked to drug resistance were found (Figure 5.38-B).

5.4.4.3 UL27

Mutations in UL27 have are associated with relatively low-level resistance to MBV. Classical mutations (relative to the HCMV strain AD169, FJ527563.1) include R233S, W362R, A406V combined with C415stop (Chou et al., 2004), V353A, W153R, L193F, A269T, V353E, L426F, E22stop, W362stop, 218delC, and 301-311del (Chou, 2009). None of these mutations was present in ZMB240. Several polymorphisms in UL27 have also been observed which are not associated with conferring drug resistance. They include N289D, D298G, N300G, P307L, V310A, D351N, and I367V (Chou, 2009). Of these D298G and I367V occur in ZMB240, as highlighted in figure 5.37.

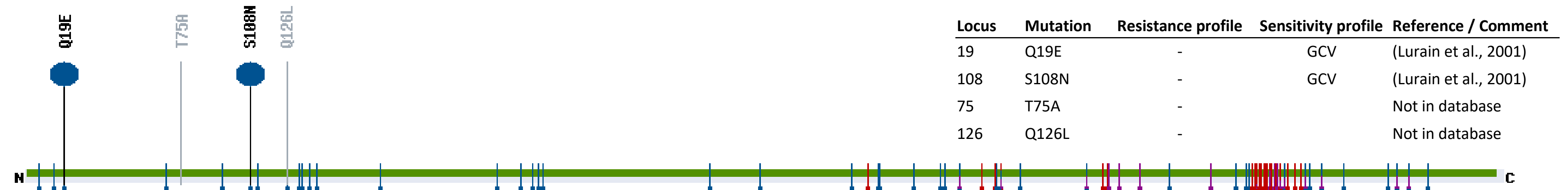
```
MNPVDQPPPPLPTQQPEEQAKEDHDDGDERLFRDPLTTYEYLDDCRDDEEFCHQFLRAYL
         10        20        30        40        50        60
TPIRNRQEAVRAGLLCRTPEDLAAAGGQKKKTPAPKHPKHAMVYIRRSCLVHSACATAHG
         70        80        90       100       110       120
KYDIRGLTLESDLAVWAALRGVPLPPDPQHFRWLNAGAFRRLVHEAQYLPEISRAAKRIA
        130       140       150       160       170       180
LAVATGQYVVCTLLDYKTFGTRTHYLRQLCSMTEELYLRLDGTLCLFLEPEERELIGRCL
        190       200       210       220       230       240
PAALCRGLPVKYRTHRAAVFFHATFMARAEAALKDLYAAFCECGDGRDNGGNHDGNHDGN
        250       260       270       280       290       300
DHSSLSPSAVASHHSRLEHAELRLERNRHLGAFHLPAIRHLTAGDVARVQDSVSRDLGFA
        310       320       330       340       350       360
DWSQTLVDDYFLLPAGWACANPRRGYAMYLASNAVLALRIIRLLRASIRHEYTACIRMLS
        370       380       390       400       410       420
GDVQRLIRLFKGEAALLRKGLAQNPVQRRELSRFRKHVHDLKRIRFTEDTFVETFCDFLE
        430       440       450       460       470       480
LVQRIPDYRSVSLRIKRELLCLHVFKLRRGCRAPPTPETARVQRLLWHSLRHGDAPQDRT
        490       500       510       520       530       540
RLPQFSSALSDAELSNHANRCRRKAPLELGPAVVAAPGPSVRYRAHIQKFERLHVRRFRP
        550       560       570       580       590       600
HEVGGHAT
```

Figure 5.37. Polymorphisms in ZMB240 UL27. The full-length amino acid sequence shows two polymorphisms relative to HCMV strain AD169 (FJ527563.1): D298G and I367V, which are not associated with resistance to Maribavir.

5.4.4.4 Summary

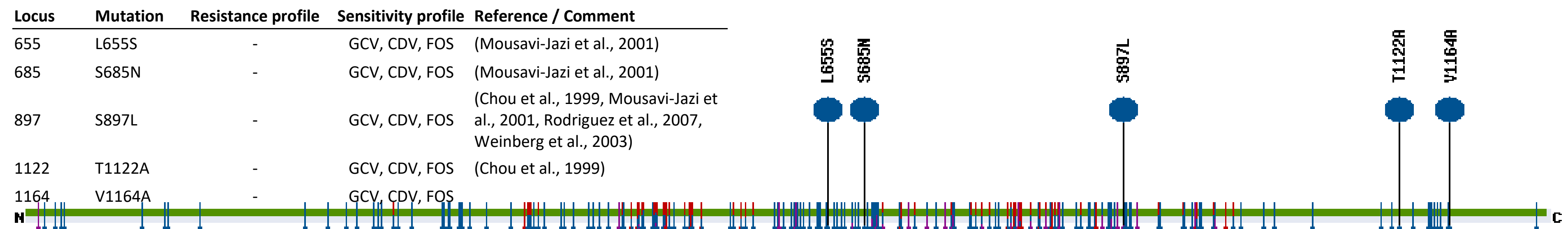
None of the three genes associated with antiviral drug resistance in HCMV showed any SNPS indicative of baseline drug resistance in the ZMB240 consensus genome.

[A] UL 97



Locus	Mutation	Resistance profile	Sensitivity profile	Reference / Comment
19	Q19E	-	GCV	(Lurain et al., 2001)
108	S108N	-	GCV	(Lurain et al., 2001)
75	T75A	-		Not in database
126	Q126L	-		Not in database

[B] UL 54



Locus	Mutation	Resistance profile	Sensitivity profile	Reference / Comment
655	L655S	-	GCV, CDV, FOS	(Mousavi-Jazi et al., 2001)
685	S685N	-	GCV, CDV, FOS	(Mousavi-Jazi et al., 2001)
897	S897L	-	GCV, CDV, FOS	(Chou et al., 1999, Mousavi-Jazi et al., 2001, Rodriguez et al., 2007, Weinberg et al., 2003)
1122	T1122A	-	GCV, CDV, FOS	(Chou et al., 1999)
1164	V1164A	-	GCV, CDV, FOS	

— alignment — mutation not in database — mutation associated with drug resistance
— mutation with unclear phenotype — mutations associated with genetic polymorphism
— single frameshift mutation — double frameshift mutation — detected stop codon

Figure 5.38. Polymorphisms in ZMB240 UL97 (Serine/Threonine Kinase) and UL54 (DNA polymerase) gene sequences. Mutation maps and drug resistance/sensitivity profile tables adapted from screening reports by the Mutation Resistance Analyzer (MRA) genotyping database, http://www.informatik.uni-ulm.de/ni/mitarbeiter/HKestler/hcmv.

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
RL1	1551	2483	Forward	Protein RL1	310	34905	
RL5A	5588	5860	Reverse	Protein RL5A	90	9706	5 amino acid insertions (various positions) and 10 residues deleted (various positions)
RL8A	8036	8308	Reverse	Protein RL8A	90	10168	
RL9A	8374	8517	Reverse	Protein RL9A	47	5285	
RL10	8724	9236	Forward	Envelope glycoprotein RL10	170	19032	
RL11	9265	9969	Forward	Membrane glycoprotein RL11	234	26629	
RL12	9973	11172	Forward	Membrane protein RL12	399	45369	13 amino acid insertions (various positions) and 28 residues deleted (various positions)
RL13	11282	12205	Forward	Membrane protein RL13	307	34279	22 amino acid insertions (various positions) and 8 residues deleted (various positions)
UL1	12320	12967	Forward	Membrane protein UL1	215	24378	8 amino acids inserted at positions 16 (X2), 35 (X1), 44 (X3), 189 (X2); 11 residues deleted at 140–150
UL2	13361	13543	Reverse	Protein UL2	60	6869	Single amino acid insertion after position 12
UL4	13971	14420	Forward	Envelope glycoprotein UL4	149	17140	
UL5	14511	15008	Forward	Protein UL5	165	18758	Single amino acid deletion at position 28
UL6	15119	15961	Forward	Membrane protein UL6	280	30922	
UL7	16021	16689	Forward	Membrane glycoprotein UL7	222	24435	
UL8	16021	17097	Forward	Membrane glycoprotein UL8	324	35826	
UL9	17094	17789	Forward	Membrane glycoprotein UL9	231	26621	2 amino acid insertions: at positions -1 and after 76; block of 4 residues deleted at positions 62–65
UL10	17904	18668	Forward	Membrane protein UL10	254	29037	
UL11	18770	19588	Forward	Membrane glycoprotein UL11	272	30823	
UL13	19786	21204	Forward	Protein UL13	472	54403	Single amino acid deletion at position 396
UL14	21354	22337	Forward	Membrane protein UL14	327	36707	
UL15A	22350	22658	Forward	Protein UL15A	102	11199	
UL16	22879	23571	Forward	Membrane glycoprotein UL16	230	26161	
UL17	23679	23993	Forward	Protein UL17	104	12637	

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
UL18	24102	25205	Forward	Membrane glycoprotein UL18	367	41753	Single amino acid deletion at position 34
UL19	25202	25498	Forward	Protein UL19	98	11379	
UL20	25762	26784	Forward	Membrane protein UL20	340	38449	
UL21A	26929	27300	Reverse	Protein UL21A	123	14216	
UL22A	27569	27972	Forward	Glycoprotein UL22A	106	11575	Single amino acid insertion after position 44
UL23	28338	29192	Reverse	Tegument protein UL23	284	32939	
UL24	29409	30485	Reverse	Tegument protein UL24	358	40205	
UL25	30533	32503	Forward	Tegument protein UL25	656	73540	
UL26	32690	33256	Reverse	Tegument protein UL26	188	21127	
UL27	33312	35138	Reverse	Protein UL27	608	69207	
UL29	35235	37486	Reverse	Protein UL29	701	79053	Single amino acid insertion after position 25
UL30	37616	37981	Reverse	Protein UL30	121	13986	
UL30A	38027	38266	Reverse	Protein UL30A	79	9469	Has the alternative ACG initiation codon; same as Merlin
UL31	38460	40247	Forward	Protein UL31	595	65661	
UL32	40325	43471	Reverse	Tegument protein pp150	1048	112766	2 amino acid deletions at 431 and 720 and an insertion after position 899
UL33	43542	44898	Forward	Envelope glycoprotein UL33	411	46073	Single amino acid deletion at position 179
UL34	45266	46489	Forward	Protein UL34	407	45445	
UL35	46568	48493	Forward	Tegument protein UL35	641	72631	
UL36	48719	50256	Reverse	Tegument protein vICA	476	54950	
UL37	50390	53195	Reverse	Envelope glycoprotein UL37	490	56274	Three amino acid insertions after position 340 and a deletion at position 400
UL38	51617	52612	Reverse	Protein UL38	331	36822	
UL40	53709	54374	Reverse	Membrane glycoprotein UL40	221	24473	
UL41A	54536	54772	Reverse	Protein UL41A	78	8953	
UL42	54878	55246	Reverse	Protein UL42	122	13399	Three amino acid deletions at positions 116, 121, and 122

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
UL43	55310	56581	Reverse	Tegument protein UL43	423	47734	
UL44	56628	57929	Reverse	DNA polymerase processivity subunit	433	46233	
UL45	58070	60790	Reverse	Ribonucleotide reductase subunit 1	906	101562	
UL46	60936	61808	Reverse	Capsid triplex subunit 1	290	33027	
UL47	61807	64758	Forward	Tegument protein UL37	983	109990	
UL48	64755	71477	Forward	Large tegument protein	2240	253115	
UL48A	71596	71823	Reverse	Small capsid protein	75	8495	Single amino acid deletion at position 318
UL49	71816	73528	Reverse	Protein UL49	570	63868	
UL50	73485	74678	Reverse	Nuclear egress membrane protein	397	42895	Single amino acid deletion at position 217
UL51	74700	75173	Reverse	DNA packaging protein UL33	157	17025	
UL52	75211	77217	Forward	DNA packaging protein UL32	668	74146	
UL53	77210	78340	Forward	Nuclear egress lamina protein	376	42303	
UL54	78318	82046	Reverse	DNA polymerase catalytic subunit	1242	137160	Five polymorphisms are present: L655S, S685N, S897L, V1164A, (also present in Merlin) and T1122A (absent in Merlin). None confers resistance to anti-HCMV drugs.
UL55	82188	84911	Reverse	Envelope glycoprotein B	907	101954	Genotype 1
UL56	84874	87426	Reverse	DNA packaging terminase subunit 2	850	95757	
UL57	87993	91700	Reverse	Single-stranded DNA-binding protein	1235	133880	
UL69	99626	101854	Reverse	Multifunctional expression regulator	742	82456	
UL70	101954	104794	Reverse	Helicase-primase primase subunit	946	107825	
UL71	104810	105895	Forward	Tegument protein UL51	361	39870	
UL72	105976	107142	Reverse	Deoxyuridine triphosphatase	388	43475	
UL73	107158	107565	Forward	Envelope glycoprotein N	135	14386	Genotype gN4d
UL74	107537	108940	Reverse	Envelope glycoprotein O	467	54127	Genotype gO5; First 5 amino acids are deleted, leaving alternative internal initiation site
UL74A	108249	109157	Forward	Envelope glycoprotein 24	70	7629	

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
UL75	109329	111557	Reverse	Envelope glycoprotein H	742	84230	Genotype 2
UL76	111752	112729	Forward	Nuclear protein UL24	325	36015	
UL77	112332	114239	Forward	DNA packaging tegument protein UL25	635	70521	Single amino acid deletion at position 142
UL78	114328	115623	Forward	Envelope protein UL78	431	47369	
UL79	115678	116565	Reverse	Protein UL79	295	33872	
UL80	116602	118728	Forward	Capsid maturation protease	708	73862	
UL80.5	117607	118728	Forward	Capsid scaffold protein	373	38222	
UL82	118888	120567	Reverse	Tegument protein pp71	559	62002	
UL83	120756	122441	Reverse	Tegument protein pp65	561	62883	
UL84	122712	124475	Reverse	Protein UL84	587	65379	
UL85	124507	125427	Reverse	Capsid triplex subunit 2	306	34595	
UL86	125589	129701	Reverse	Major capsid protein	1370	153786	
UL87	129761	132586	Forward	Protein UL87	941	104851	
UL88	132583	133872	Forward	Tegument protein UL88	429	47631	
UL89	133869	139795	Reverse	DNA packaging terminase subunit 1	674	77035	
UL91	135241	135576	Forward	Protein UL91	111	12013	
UL92	135546	136151	Forward	Protein UL92	201	22511	
UL93	136117	137901	Forward	DNA packaging tegument protein UL17	594	68421	
UL94	137759	138796	Forward	Tegument protein UL16	345	38354	
UL95	139794	141377	Forward	Protein UL95	527	56911	Block of 6 amino acids deleted at position 382-387
UL96	141374	141757	Forward	Tegument protein UL14	127	14465	
UL97	141878	144001	Forward	Tegument serine/threonine protein kinase	707	78248	4 polymorphisms present: Q19E and S108N curated in MRA database, while T75A and Q126L currently not in MRA. None is known to confer resistance to anti-HCMV drugs.
UL98	144095	145849	Forward	Deoxyribonuclease	584	65331	

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
UL99	145786	146358	Forward	Myristylated tegument protein	190	20933	
UL100	146616	147737	Reverse	Envelope glycoprotein M	373	42817	2 amino acid insertions after positions 290 and 361
UL102	147915	150533	Forward	Helicase-primase subunit	872	93934	Single amino acid deletion at position 805
UL103	150697	151446	Reverse	Tegument protein UL7	249	28636	
UL104	151394	153487	Reverse	Capsid portal protein	697	78521	
UL105	153315	156185	Forward	Helicase-primase helicase subunit	956	106500	
UL111A	161069	161758	Forward	Interleukin-10	176	20108	
UL112	161983	164194	Forward	Protein UL112	684	70270	Single amino acid deletion at position 599 and an insertion after position 634
UL114	164366	165118	Reverse	Uracil-DNA glycosylase	250	28353	
UL115	165090	165926	Reverse	Envelope glycoprotein L	278	30765	Genotype 3
UL116	165926	166867	Reverse	Protein UL116	313	34203	
UL117	166864	168138	Reverse	Protein UL117	424	45491	Single amino acid deletion at position 607
UL119	168251	169393	Reverse	Membrane glycoprotein UL119	351	39330	Single amino acid deletion at position 35, and block of 7 residues inserted after position 61
UL120	169448	170041	Reverse	Membrane protein UL120	197	22267	Block of 5 amino acids deleted at positions 27–31, and a single residue inserted after position 145
UL121	170092	170634	Reverse	Membrane protein UL121	180	20199	
UL122	170762	174164	Reverse	Regulatory protein IE2	580	63017	
UL123	172404	174164	Reverse	Regulatory protein IE1	491	55038	
UL124	174197	174655	Forward	Membrane protein UL124	152	15855	2 amino acid insertions: 1 after position 48 and the other after position 52
UL128	176263	177021	Reverse	Envelope protein UL128	171	19717	Intact UL128, with the classical 'C' variation at nucleotide position 634 of mature transcript
UL130	177061	177705	Reverse	Envelope glycoprotein UL130	214	24652	
UL131A	177726	178223	Reverse	Envelope protein UL131A	129	14989	
UL132	178329	179141	Reverse	Envelope glycoprotein UL132	270	29991	
UL148	179218	180168	Reverse	Membrane protein UL148	316	36475	
UL147A	180237	180464	Reverse	Membrane protein UL147A	75	8291	

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
UL147	180467	180946	Reverse	Chemokine vCXCL2	159	18826	
UL146	181010	181387	Reverse	Chemokine vCXCL1	125	14307	Block of 5 amino acids inserted after position 105; Genotype 3
UL145	181668	182060	Reverse	Protein UL145	130	14586	
UL144	182301	182831	Reverse	Membrane glycoprotein UL144	176	19469	
UL142	183472	184392	Reverse	Membrane glycoprotein UL142	306	35011	Single amino acid insertion after position 211
UL141	184465	185481	Reverse	Membrane glycoprotein UL141	338	38929	
UL140	185780	186355	Reverse	Protein UL140	191	21489	
UL139	186537	186944	Reverse	Membrane glycoprotein UL139	135	14025	3 amino acids deleted at positions 34–36; Genotype 4
UL138	187507	188016	Reverse	Protein UL138	169	19282	
UL136	188099	188821	Reverse	Protein UL136	240	27121	
UL135	188912	189838	Reverse	Protein UL135	308	33325	
UL133	190015	190788	Reverse	Protein UL133	257	27689	
UL148A	190884	191123	Reverse	Protein UL148A	79	8926	Single amino acid deletion at position 6
UL148B	191253	191495	Reverse	Protein UL148B	80	8869	
UL148C	191582	191815	Forward	Protein UL148C	77	8527	
UL148D	192196	192384	Forward	Protein UL148D	62	7065	
UL150	192400	194319	Reverse	Protein UL150	639	70148	
UL150A	193259	194185	Forward	Protein UL150A	271	29382	
IRS1	196313	198889	Forward	Tegument protein IRS1	858	92077	Block of 11 amino acids inserted after position 368
US1	198913	199383	Reverse	Protein US1	156	17791	
US2	199701	200300	Reverse	Membrane glycoprotein US2	199	23112	
US3	200715	201275	Reverse	Membrane glycoprotein US3	186	21515	
US6	201987	202538	Reverse	Membrane glycoprotein US6	183	20541	Single amino acid insertion after position 35
US7	202952	203629	Reverse	Membrane glycoprotein US7	225	26350	

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
US8	203832	204512	Reverse	Membrane glycoprotein US8	226	26528	Single amino acid deletion at position 29
US9	204537	205235	Reverse	Membrane glycoprotein US9	232	26247	Deletion of 15 amino acids at positions 233–247
US10	205618	206172	Reverse	Membrane glycoprotein US10	184	20659	Single amino acid deletion at position 18
US11	206248	206895	Reverse	Membrane glycoprotein US11	215	25264	
US12	207081	207926	Reverse	Membrane protein US12	281	32499	
US13	208006	208791	Reverse	Membrane protein US13	261	29460	
US14	208860	209792	Reverse	Membrane protein US14	310	34278	
US15	209837	210625	Reverse	Membrane protein US15	262	29085	
US16	210685	211614	Reverse	Membrane protein US16	309	34689	
US17	211759	212640	Reverse	Membrane protein US17	293	31935	
US18	212908	213732	Reverse	Membrane protein US18	274	30194	
US19	213870	214592	Reverse	Membrane protein US19	240	26393	
US20	214639	215403	Reverse	Membrane protein US20	254	28494	
US21	215510	216241	Reverse	Membrane protein US21	243	27015	
US22	216418	218148	Reverse	Tegument protein US22	576	65003	
US23	218262	220040	Reverse	Protein US23	592	68885	
US24	220136	221641	Reverse	Tegument protein US24	501	58026	
US26	222275	224086	Reverse	Protein US26	603	69978	
US27	224452	225540	Forward	Envelope glycoprotein US27	362	41984	2 amino acid deletions at positions 10 and 11
US28	225748	226812	Forward	Envelope protein US28	354	41020	
US29	226975	228363	Forward	Membrane protein US29	462	50983	
US30	228167	229216	Forward	Membrane protein US30	349	39184	
US31	229331	229816	Forward	Protein US31	161	18939	
US32	229934	230485	Forward	Protein US32	183	22057	

Gene	Start	End	Orientation	Product	AA length	MW	Comparison to HCMV Strain Merlin (AY446894.2)
US33A	230687	230860	Forward	Protein US33A	57	6874	
US34	231029	231520	Forward	Protein US34	163	17741	
US34A	231514	231708	Forward	Protein US34A	64	8140	
TRS1	232702	235104	Reverse	Tegument protein TRS1	800	85023	12 amino acid insertions: block of 11 after position 368; 1 after 788

Table 5.12. HCMV strain ZMB240 gene list. Comparisons of the 169 genes products in ZMB240 to reference strain Merlin (AY446894.2) are highlighted in the last column. CDS, Coding DNA Sequence; bp, base pairs; AA, amino acid; MW, molecular weight in daltons; MRA, Mutation Resistance Analyzer database (Chevillotte et al., 2010). The amino acid length excludes the STOP codon.

5.5 DISCUSSION

This chapter analyses burden of infection with HCMV in breast milk, comparing with viral load as well as effects of HIV infection. In order to do this, methods were established for the detection and quantification of HCMV genotype variants. ZMB240 – a complete HCMV genome sequenced from breast milk is also characterised. The genotype detection and enumeration method devised here utilises molecular 'barcodes' specific to each HCMV gO and gN genotype, combined with perl Scripts to interrogate NGS FASTQ reads. The method was tested on archived sequences in GenBank release 211 (National Center for Biotechnology Information (NCBI), 2015) and showed high sensitivity of 99.4% for gO and 98.6% for gN. This however may in fact be an underestimate, since sequences not assigned a genotype appeared to have unresolved ambiguous bases and possible compilation errors, although two potential recombinants were detected as detailed in Section 5.2.3. The method was subsequently applied to NGS reads sequenced directly from 21 clinical (breast milk) samples unexposed to biases introduced by tissue culture. We showed that in these samples, collected from both HIV-positive and negative women, multiple genotypes were present indicating mixed infections. Overall, all eight HCMV gO genotypes were detected among the HIV-positive samples, while among the HIV-negatives gO2b was not detected. Although this could be an effect of the relatively small sample, it is consistent with Sanger sequencing results that showed differences in the genotypes present in the two HIV groups (see Chapter 4). Interestingly, for individuals where milk samples were available and sequenced independently from both breasts, there was uniformity in genotype proportions, suggesting a common source of virus into the mammary gland possibly from reactivation from latency established in another body compartment. Interestingly, the viral load and genotypes increased from week 4 postpartum, and this is also a documented time of myeloid cellular infiltration (Maschmann et al., 2015). Dynamics of genotype proportions were investigated by correlating the genotype prevalence to the viral load kinetics over the first 16 postpartum weeks. Three HIV positive women had samples collected at more than one time-point and were therefore used for this analysis, which showed that between week 4 and week 12 there was expansion in the number of genotypes in milk, and by week 16 the picture returned to the week 4 state. The period between weeks 4 and 12 coincides with the end of the puerperium and there is recovery from the relative immunosuppressed pregnancy state (Groer et al., 2015, Groer et al.,

2005). The expansion of HCMV genotypes during this period may reflect the enhanced immune pressure on the virus leading to a phenomenon akin to “quasispecies” seen in RNA viruses, as has been shown previously in HCMV (Renzette et al., 2011). This genotype expansion may also represent new strains imported into the breast compartment from enhanced trafficking of macrophages and other monocytic cells which has been shown to peak around weeks 9 – 11 (Maschmann et al., 2015). The subsequent return to the week 4 genotypes points to an equilibrium state being reached as immune control is enhanced, as reflected by the fall in HCMV viral load by week 16. Humoral (antibody) responses could play a vital role in shaping this picture, as has been recently demonstrated in studies of HIV where antibody responses remained directed at the initial infecting subtype (Cornelissen et al., 2016).

The final section of this chapter characterises the sequence of ZMB240, a complete HCMV genome derived from milk from an HIV-negative mother. This is to our knowledge the first complete HCMV genome from Africa. Also significant is the fact that it represents virus involved in the natural route of infection via breast milk. With a typical type E herpesviral genome architecture, 236,211 bp length, and 57.6% GC content, the ZMB240 genome is similar to other complete HCMV sequences (see appendix table 5.1). ZMB240 has a complement of 169 genes, and other genetic elements such as microRNAs and regulatory factors found in other HCMV strains. The only HCMV gene absent is UL6, which in ZMB240, like several other strains (see section 5.2), is initiated by the alternative TTG start codon. Core, beta, and betagamma genes, as well as all 26 currently verified HCMV microRNAs, are conserved in ZMB240. Although some gene products show differences compared to the reference strain Merlin (AY446894.2), these differences have been documented previously in several other HCMV strains, as detailed in section 5.2. Similar to the Merlin strain, ZMB240 carries the gO genotype 5 and gN genotype 4d. It was noted that ZMB240 gO5 is initiated from an alternative start site 5 residues downstream relative to strain Merlin gO5 (figure 5.29). This phenomenon has been demonstrated in several gO5 sequences such as HCMV strain U8 (accession GU179288.1) (figure 5.29) and others (Bates et al., 2008). An important difference between ZMB240 and HCMV strain Merlin is that unlike the sequence of Merlin deposited in GenBank (accession number AY446894.2) ZMB240 contains intact locus UL128, which is a critical determinant of viral tropism as it encodes part of the

pentameric complex necessary for virus entry into non-fibroblast cells. HCMV strain Merlin later re-sequenced from the original clinical source material (urine) was also found to have intact UL128 (Dolan et al., 2004). Similar to other clinical strains and unlike laboratory-adapted HCMV strains, the ULb’ region is intact in ZMB240 and located at the right-hand end of the long unique region. This “virulence region” crucially encodes up to 20 genes essential for viral pathogenesis, immune evasion, virus dissemination, and latency. ULb’ is dispensable for growth and, interestingly, strains lacking ULb’ replicate with high efficiency and increased yields of infectious progeny virus *in vitro* compared to clinical strains with intact ULb’ (Wang et al., 2005, Goodrum et al., 2007, Cha et al., 1996, Dolan et al., 2004, Murphy et al., 2003). Further, there are no anti-HCMV drug resistance mutations present in ZMB240, consistent with the fact that use of anti-HCMV agents is currently virtually non-existent in the Zambian population, from where ZMB240 originated. Currently, HCMV genomes are predominantly isolated from donors with pathology, commonly from urine from congenitally infected children, or from virus reactivations in immunosuppressed transplantation patients, or drug resistance variants. Overall, ZMB240 represents a reference clinical strain from the natural transmission medium – breast milk – from a healthy HIV-negative donor, therefore the first reference strain from a healthy donor representing a transmission population.

CHAPTER 6:

SUMMARY AND CONCLUSIONS

This thesis aimed to investigate and establish which route(s) of infection were associated with adverse childhood growth and development effects shown in previous studies among infants in Zambia, Sub-Saharan Africa , as well as the role of maternal HIV infection in the transmission of HCMV in this setting. Non-invasive approaches for diagnosing HCMV were also trialled in examining congenital infection via saliva, and postnatal transmission via breast milk. Importantly, our studies compared both HIV-positive and negative groups inorder to delineate the role of maternal HIV in HCMV transmission dynamics.

Our investigation of Congenital HCMV in a normal labour ward utilised PCR on neonatal saliva collected within the first hour of birth and before onset of breast milk feeding. The use of saliva to diagnose infection was warmly accepted by parents, even during this culturally sensitive period of life. Congenital HCMV was 1% prevalence (1/100), similar to surveys elsewhere. We were able to obtain nucleotide sequence and using Sanger sequencing determined the the gO genotype to be gO1c. This shows that use of saliva is sensitive for both detection and genotyping of HCMV, as well as the related betaherpesviries HHV-6A/B, as demonstrated in our validation experiments (section 3.4). Our finding of 1% prevalence of congenital HCMV was lower than the 3.8% prevalence found by Mwaanza and co-workers among high risk neonates in the UTH NICU (Mwaanza et al., 2014). That study showed that HCMV prevalence was linked to maternal HIV seropositivity, but it did not account for possible confounding effects on HCMV prevalence of breast milk feeding (which mothers are encouraged to do in NICU). Another study published during this thesis found 2.6% congenital HCMV prevalence among HIV positive infants in South Africa (Manicklal et al., 2014), but no comparisons were made with infants unexposed to HIV. Similarly, in Zimbabwe Gumbo and colleagues found congenital HCMV was 11% among treatmant-naïve HIV-infected infants (Gumbo et al., 2014); no comparison was made to HIV-unfected infants. Although our congenital HCMV study was limited in the number of infants recruited due to industrial action by healthcare staff during the field study, we took precautions to

control for possible confounders including breast milk feeding by collecting samples within one hour of birth and before onset of feeding. Furthermore, we recruited both HIV-exposed and unexposed neonates. It would be interesting to study congenital HCMV further in LW in a larger neonatal sample.

In our examination of breast milk, we found HCMV infection common among both HIV-positive and negative mothers. Virus reactivation in milk occurred early in the postpartum period; infection was detected as early as day three postpartum, and HCMV DNA loads were comparable in both groups of mothers. However HCMV loads increased twice as quickly in HIV-positive mothers as it did in HIV-negative mothers, reaching peak levels in both groups at week 4 postpartum (figure 4.5). Interestingly, studies in Europe and Asia have also reported HCMV levels peaking in milk at week 4 (Hamprecht et al., 2008, Jim et al., 2009, Yasuda et al., 2003). In our study, peak HCMV levels (at W4) were significantly higher in the HIV-positive group (p=.026), and remained higher to the last time point, week 16, where significantly more HIV-positive women remained with detectable HCMV in milk compared to their HIV-negative counterparts (p<.001). Assuming steady state HCMV dynamics, we estimated that it would take about 26 weeks for HCMV to reach undetectable levels from peak loads in the HIV-positive mothers. The high and prolonged levels of virus secretion correspond to the greater odds for HCMV infection for infants of HIV-positive mothers breasfed longer than 6 months or those of HIV-negative mothers breastfed beyond 18 months (table 4.6). Clearly, maternal HIV increases the load and prolongs duration of HCMV secretion, thereby increasing HCMV risk and exposure to infants.

Although all gO genotypes were found in breast milk by Sanger sequencing, no genotype was associated with higher HCMV levels. Nevertheless, mixed-genotype infections were evident and therefore further examined by NGS deep sequencing to ascertain the burden of infection. This required a new analytical approach, which we devised by developing molecular barcodes specific to each gO genotype. Coupled with custom *perl* scripts, the barcodes were used to enumerate gO genotypes, including minor variants to as low as 1%, in NGS FASTA reads. In 21 breast milk-derived DNA samples analysed by NGS and our barcode method, we found upto seven genotypes per sample (table 5.8). In three HIV-positive mothers with longitudinal samples analysed by NGS, we noted

a pattern whereby there was an expansion of genotypes present in milk between W4 and W12 before a return to the W4 genotype(s) by W16 (figures 4.15 – 5.17). The genotype expansion may be due to superinfection, which commonly occurs, particularly if there are other young children in the family. It may also be due to enhanced immune activation and cellular infiltration into breast tissue similar to an immune reconstitution syndrome (IRIS) phenomenon as mothers during the peuperium undergo physiological changes returning to the pre-pregnancy state.

Finally, this thesis characterised ZM240, a complete HCMV genome derived from DNA prepared directly from breast milk from a healthy HIV-negative mother. ZM240 has the architecture of a typical herpesviral class E genome and at 236,211bp length and 57.6% GC content is similar to other full-length HCMV genomes publicly available in GenBank. It contains the full gene complement but one, all regulatory elements, including noncoding RNAs, and is free of anti-HCMV drug resistance mutations. To our knowledge, ZM240 is the first complete HCMV genome from a natural route of transmission, breast milk, in a healthy individual, and from Africa.

As endorsed by the World Health Organisation (WHO), exclusive breastfeeding for the first six months (Kramer and Kakuma, 2012) and continuation thereafter following introduction of complementary foods, remains of nutritional and immunological benefit for infants (http://www.who.int/nutrition/topics/infantfeeding_recommendation/en/). This is of particular importance where alternative feeding options are unavailable or cannot be safely practiced, for instance because of non-availability of safe water supply or due to compromised hygiene conditions. However given the detrimental effects of breastmilk-transmitted HCMV highlighted in this thesis, including on child health and development, and long-term effects on quality of life, geriatric illness, and lifespan, breastfeeding poses a real danger. A balanced approach is therefore necessary to retain the value and benefits of breast milk while limiting HCMV transmission especially in communities where breastfeeding is widely practiced, and therefore presents higher risk of infant infection. Strategies could include use of anti-HCMV drugs in mothers, although this would first require safety trials, since so far no anti-HCMV drugs are licenced for use during lactation. Treatment of milk by ultrashort heat treatment (Goelz et al., 2009, Stock et al., 2015, Hamprecht et al., 2004a) or irradiation (Christen et al.,

2013, Christen, 2014) are also options that are already being applied in controlled hospital settings for premature and very low birthweight infants. Additionally, vaccines are in development by various researchers to improve maternal immunity against HCMV (Boppana and Britt, 2014, Fu et al., 2014, Wang and Fu, 2014). Use of hyperimmune globulin, which interestingly is focused on antibodies to the gH/gL complex (Fouts et al., 2012), is another strategy for limiting infection and possible virus transmission. Other control efforts are targeted at addressing HCMV transmission via saliva and urine from young children, who serve as infection reservoirs in families and communities. These efforts include the promotion of good personal hygiene, particularly hand hygiene (Stowell et al., 2014, Cannon and Davis, 2005).

FUTURE WORK

In this thesis, we assessed postnatal and congenital HCMV infection together with effects of maternal HIV exposure in maternal and neonatal cohorts in Zambia. The results showed low levels of congenital infection in the general new-born labour ward, 1%, compared to high levels of early postnatal infection by the age of 18 months, which were increased in children who were maternally HIV exposed. In order to characterise transmission at this early age, secretion in breast milk was examined, since this had been shown a major route for postnatal infection in breastfeeding populations. Viral loads, genotypes, and genomes were examined in DNA from breast milk from HIV–infected and uninfected mothers. Further, we devised a tool for assessing burden of infection via detecting genotype mixtures in NGS deep sequencing reads.

For future studies, it would be interesting to employ the novel method developed here which would enable examining the burden of infection, either already present in the mother, or from repeated infections introduced from young siblings. Young children have been shown to secrete HCMV for several years (Adler et al., 2004, Noyola et al., 2000), and in addition to breast milk, HCMV secretion in children's saliva is likely a major source, although urine could also contribute depending on level of hygiene. With this genotyping method, all hypervariable genotype tags could be included in the script to provide a definitive examination of this issue of relevance for any intervention. The thesis developed a novel method based on molecular tags unique to each gO and gN

genotype. Further work could develop similar tags for detecting and quantifying other hypervariable HCMV genes, including members of the RL11 and US12 families, which are some of the most hypervariable in the genome (Dolan et al., 2004, Murphy et al., 2003). This could be useful in rapidly constructing a viral 'haplotype' profile for each sequenced sample and in complementing approaches such as principal component analysis to explore and define relationships between the various hypervariable HCMV genes and even those with minor changes.

It is also of interest to advance the concept of molecular tags further for application into HCMV diagnostics, which currently remain a critical bottleneck in many areas outside the developed world. The molecular tags have potential for incorporation into simplified low-cost assays. For instance, the molecular tags could potentially be fluorescently labelled and used as probes in real time qPCR-based assay formats to examine relative viral loads and kinetics over time, for example in response to vaccination or repeated infections. This can not only greatly improve diagnosis capability, but also serve various other applications, such as genotype profiling, and measurement of disease burden. Such multi-faceted assays would be particularly useful in settings of constrained resources where access to full-scale NGS facilities may not be available.

Another study area that merits future consideration, and where the molecular tags have potential for application, is in the systematic tracking of HCMV genotypes transmitted by mothers to their infants and how these evolve over time in infants. Such a study would require a cohort of mother-infant pairs, and would help provide answers to the yet still open question of whether infants do re-infect their mothers during early infancy when there is intimate nursing, and how this affects virus dynamics and evolution in both mother and infant.

A main limitation in our studies was that the maternal HCMV serostatus was unknown, and samples from other compartments such as serum or saliva were unavailable. However, in epidemiological follow up, almost 100% children from HIV positive mothers were HCMV positive, therefore the adult population is likely close to universal HCMV. It would have been interesting to examine HCMV genotypes secreted from saliva. Although this does not affect our primary goals and findings, availability of other sample

types would have been useful to compare the viral load and genotype kinetics in various body compartments to the findings in milk. This would certainly shed more light on how virus in breast milk relates to that in other tissues.

In this thesis, ZMB240, the first complete genome from the natural transmission source, breast milk, in Africa was presented and extensively discussed. Future work can include this as a reference, since all other publicly available full HCMV genomes are from pathological samples and from non-African nations. Another genome was assembled from an HIV-positive mother and will be submitted to NCBI in addition to ZMB240. Further work would include comparison of all assembled genomes from this cohort to these reference genomes. This will not only add to the growing body of HCMV genome data from diverse geographical regions of the world, important for vaccine considerations, but also help define a fuller picture of the HCMVs circulating in settings where transmission occurs naturally as well as potential effects of maternal HIV exposure to viral load, genotype diversity and pathogenesis.

CONCLUSIONS

This study has demonstrated that at 1% the birth prevalence of HCMV in Zambia is similar to other world regions. We have also shown that HCMV infection in breast milk is common in both HIV-infected and uninfected mothers. The principal mode of early postnatal HCMV transmission from both HIV-infected and uninfected mothers is breastfeeding, with peak HCMV levels in milk occurring at one month postpartum. The thesis demonstrated there were significantly higher loads in breast milk from HIV-infected mothers ($p=.026$) compared to HIV-negative, showing potential mechanism of effects of maternal HIV on postnatal transmission. Further, HCMV shedding is significantly prolonged beyond the fourth month in HIV-infected mothers ($p<.001$). All eight HCMV gO genotypes could be detected in breast milk, and there was evidence for mixed genotype infections, signifying reinfection and changes in maternal immune control over time. We did not find any link between HCMV load and any particular genotype. In our studies in this thesis, non-invasive samples including saliva and umbilical tissue were acceptable locally by parents of neonates for HCMV diagnosis, and together with breast milk, were successfully utilised to detect and sequence HCMV.

Additionally, we devised a new tool for detecting mixed genotypes to aid assessment of infection burden in NGS deep sequencing reads. The study characterised ZMB240, a genome sequenced from DNA prepared directly from breast milk from a healthy HIV-negative mother, representing virus in a natural transmission setting. We showed this genome to be of similar size, GC content, and architecture to published HCMV genomes from other world regions. However, unlike laboratory-adapted strains in which several genes are routinely deleted or mutated through tissue culture passaging, ZMB240 contains the full complement of HCMV genes (except for RL6, a member of the accordion gene family RL11), all known HCMV ncRNAs, and regulatory elements, and is free of anti-HCMV drug resistance mutations. On the whole, this thesis shows that maternal HIV infection increases HCMV load, burden, and duration of secretion in breast milk, and thereby can escalate the risk of early infant HCMV infection.

REFERENCES

AARNISALO, J., ILONEN, J., VAINIONPAA, R., VOLANEN, I., KAITOSAARI, T. & SIMELL, O. 2003. Development of antibodies against cytomegalovirus, varicella-zoster virus and herpes simplex virus in Finland during the first eight years of life: a prospective study. *Scand J Infect Dis,* 35**,** 750-3.

ABLASHI, D., AGUT, H., ALVAREZ-LAFUENTE, R., CLARK, D. A., DEWHURST, S., DILUCA, D., FLAMAND, L., FRENKEL, N., GALLO, R., GOMPELS, U. A., HOLLSBERG, P., JACOBSON, S., LUPPI, M., LUSSO, P., MALNATI, M., MEDVECZKY, P., MORI, Y., PELLETT, P. E., PRITCHETT, J. C., YAMANISHI, K. & YOSHIKAWA, T. 2013. Classification of HHV-6A and HHV-6B as distinct viruses. *Arch Virol*.

ACOSTA, E. P., BRUNDAGE, R. C., KING, J. R., SANCHEZ, P. J., SOOD, S., AGRAWAL, V., HOMANS, J., JACOBS, R. F., LANG, D., ROMERO, J. R., GRIFFIN, J., CLOUD, G., WHITLEY, R., KIMBERLIN, D. W., NATIONAL INSTITUTE OF, A. & INFECTIOUS DISEASES COLLABORATIVE ANTIVIRAL STUDY, G. 2007. Ganciclovir population pharmacokinetics in neonates following intravenous administration of ganciclovir and oral administration of a liquid valganciclovir formulation. *Clin Pharmacol Ther,* 81**,** 867-72.

ADAMS, M. J. & CARSTENS, E. B. 2012. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012). *Arch Virol,* 157**,** 1411-22.

ADAMS, O., KREMPE, C., KOGLER, G., WERNET, P. & SCHEID, A. 1998. Congenital infections with human herpesvirus 6. *J Infect Dis,* 178**,** 544-6.

ADJEI, A., ARMAH, H. & NARTER-OLAGA, E. 2006. Seroprevalence of cytomegalovirus among some voluntary blood donors at the 37 military hospital, accra, ghana. *Ghana Med J,* 40**,** 99-104.

ADLER, S. P., FINNEY, J. W., MANGANELLO, A. M. & BEST, A. M. 2004. Prevention of child-to-mother transmission of cytomegalovirus among pregnant women. *J Pediatr,* 145**,** 485-91.

ADLER, S. P., STARR, S. E., PLOTKIN, S. A., HEMPFLING, S. H., BUIS, J., MANNING, M. L. & BEST, A. M. 1995. Immunity induced by primary human cytomegalovirus infection protects against secondary infection among women of childbearing age. *J Infect Dis,* 171**,** 26-32.

AKTER, P., CUNNINGHAM, C., MCSHARRY, B. P., DOLAN, A., ADDISON, C., DARGAN, D. J., HASSAN-WALKER, A. F., EMERY, V. C., GRIFFITHS, P. D., WILKINSON, G. W. & DAVISON, A. J. 2003. Two novel spliced genes in human cytomegalovirus. *J Gen Virol,* 84**,** 1117-22.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

AMBROS, V. 2004. The functions of animal microRNAs. *Nature,* 431**,** 350-5.

ARBUCKLE, J. H., MEDVECZKY, M. M., LUKA, J., HADLEY, S. H., LUEGMAYR, A., ABLASHI, D., LUND, T. C., TOLAR, J., DE MEIRLEIR, K., MONTOYA, J. G., KOMAROFF, A. L., AMBROS, P. F. & MEDVECZKY, P. G. 2010. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. *Proc Natl Acad Sci U S A,* 107**,** 5563-8.

ARCANGELETTI, M. C., VASILE SIMONE, R., RODIGHIERO, I., DE CONTO, F., MEDICI, M. C., MARTORANA, D., CHEZZI, C. & CALDERARO, A. 2015. Combined genetic variants of human cytomegalovirus envelope glycoproteins as congenital infection markers. *Virol J,* 12**,** 202.

ARRODE, G. & DAVRINCHE, C. 2003. Dendritic cells and HCMV cross-presentation. *Curr Top Microbiol Immunol,* 276**,** 277-94.

ASAHI-OZAKI, Y., SATO, Y., KANNO, T., SATA, T. & KATANO, H. 2006. Quantitative analysis of Kaposi sarcoma-associated herpesvirus (KSHV) in KSHV-associated diseases. *J Infect Dis,* 193**,** 773-82.

ASANUMA, H., NUMAZAKI, K., NAGATA, N., HOTSUBO, T., HORINO, K. & CHIBA, S. 1996. Role of milk whey in the transmission of human cytomegalovirus infection by breast milk. *Microbiol Immunol,* 40**,** 201-4.

ASSEFA, S., KEANE, T. M., OTTO, T. D., NEWBOLD, C. & BERRIMAN, M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics,* 25**,** 1968-9.

AVERY, R. K., MARTY, F. M., STRASFELD, L., LEE, I., ARRIETA, A., CHOU, S., TATAROWICZ, W. & VILLANO, S. 2010. Oral maribavir for treatment of refractory or resistant cytomegalovirus infections in transplant recipients. *Transpl Infect Dis,* 12**,** 489-96.

BALTIMORE, D. 1971. Expression of animal virus genomes. *Bacteriol Rev,* 35**,** 235-41.

BANKIER, A. T., BECK, S., BOHNI, R., BROWN, C. M., CERNY, R., CHEE, M. S., HUTCHISON, C. A., 3RD, KOUZARIDES, T., MARTIGNETTI, J. A., PREDDIE, E. & ET AL. 1991. The DNA sequence of the human cytomegalovirus genome. *DNA Seq,* 2**,** 1-12.

BARAMI, K. 2010. Oncomodulatory mechanisms of human cytomegalovirus in gliomas. *Journal of Clinical Neuroscience,* 17**,** 819-823.

BARTEL, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* 116**,** 281-97.

BATE, S. L., DOLLARD, S. C. & CANNON, M. J. 2010. Cytomegalovirus seroprevalence in the United States: the national health and nutrition examination surveys, 1988-2004. *Clin Infect Dis,* 50**,** 1439-47.

BATES, M., MONZE, M., BIMA, H., KAPAMBWE, M., CLARK, D., KASOLO, F. C. & GOMPELS, U. A. 2009. Predominant human herpesvirus 6 variant A infant infections in an HIV-1 endemic region of Sub-Saharan Africa. *J Med Virol,* 81**,** 779-89.

BATES, M., MONZE, M., BIMA, H., KAPAMBWE, M., KASOLO, F. C., GOMPELS, U. A. & GROUP, C. S. 2008. High human cytomegalovirus loads and diverse linked variable

genotypes in both HIV-1 infected and exposed, but uninfected, children in Africa. *Virology,* 382**,** 28-36.

BATTIWALLA, M., PAPLHAM, P., ALMYROUDIS, N. G., MCCARTHY, A., ABDELHALIM, A., ELEFANTE, A., SMITH, P., BECKER, J., MCCARTHY, P. L. & SEGAL, B. H. 2007. Leflunomide failure to control recurrent cytomegalovirus infection in the setting of renal failure after allogeneic stem cell transplantation. *Transpl Infect Dis,* 9**,** 28-32.

BAYER, C., VARANI, S., WANG, L., WALTHER, P., ZHOU, S., STRASCHEWSKI, S., BACHEM, M., SODERBERG-NAUCLER, C., MERTENS, T. & FRASCAROLI, G. 2013. Human cytomegalovirus infection of M1 and M2 macrophages triggers inflammation and autologous T-cell proliferation. *J Virol,* 87**,** 67-79.

BEGO, M. G., KEYES, L. R., MACIEJEWSKI, J. & ST JEOR, S. C. 2011. Human cytomegalovirus latency-associated protein LUNA is expressed during HCMV infections in vivo. *Arch Virol,* 156**,** 1847-51.

BHATTACHARJEE, B., RENZETTE, N. & KOWALIK, T. F. 2012. Genetic analysis of cytomegalovirus in malignant gliomas. *J Virol,* 86**,** 6815-24.

BIRON, K. K. 2006. Antiviral drugs for cytomegalovirus diseases. *Antiviral Res,* 71**,** 154-63.

BISHOP, R. K., VALLE OSEGUERA, C. A. & SPENCER, J. V. 2015. Human Cytomegalovirus interleukin-10 promotes proliferation and migration of MCF-7 breast cancer cells. *Cancer Cell Microenviron,* 2.

BODEUS, M., KABAMBA-MUKADI, B., ZECH, F., HUBINONT, C., BERNARD, P. & GOUBAU, P. 2010. Human cytomegalovirus in utero transmission: follow-up of 524 maternal seroconversions. *J Clin Virol,* 47**,** 201-2.

BOETZER, M., HENKEL, C. V., JANSEN, H. J., BUTLER, D. & PIROVANO, W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics,* 27**,** 578-9.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BOPPANA, S. B. & BRITT, W. J. 2014. Recent approaches and strategies in the generation of antihuman cytomegalovirus vaccines. *Methods Mol Biol,* 1119**,** 311-48.

BOPPANA, S. B. & FOWLER, K. B. 2007. Persistence in the population: epidemiology and transmisson. *In:* ARVIN, A., CAMPADELLI-FIUME, G., MOCARSKI, E., MOORE, P. S., ROIZMAN, B., WHITLEY, R. & YAMANISHI, K. (eds.) *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Cambridge: Cambridge University Press.

BOPPANA, S. B., FOWLER, K. B., BRITT, W. J., STAGNO, S. & PASS, R. F. 1999. Symptomatic congenital cytomegalovirus infection in infants born to mothers with preexisting immunity to cytomegalovirus. *Pediatrics,* 104**,** 55-60.

BOPPANA, S. B., RIVERA, L. B., FOWLER, K. B., MACH, M. & BRITT, W. J. 2001. Intrauterine transmission of cytomegalovirus to infants of women with preconceptional immunity. *N Engl J Med,* 344**,** 1366-71.

BOPPANA, S. B., ROSS, S. A., SHIMAMURA, M., PALMER, A. L., AHMED, A., MICHAELS, M. G., SANCHEZ, P. J., BERNSTEIN, D. I., TOLAN, R. W., JR., NOVAK, Z., CHOWDHURY, N., BRITT, W. J., FOWLER, K. B., NATIONAL INSTITUTE ON, D. & OTHER COMMUNICATION DISORDERS, C. S. 2011. Saliva polymerase-chain-reaction assay for cytomegalovirus screening in newborns. *N Engl J Med,* 364**,** 2111-8.

BOWDEN, R. & SAYERS, M. 1990. The risk of transmitting cytomegalovirus infection by fresh frozen plasma. *Transfusion,* 30**,** 762-3.

BRADLEY, A. J., KOVACS, I. J., GATHERER, D., DARGAN, D. J., ALKHARSAH, K. R., CHAN, P. K., CARMAN, W. F., DEDICOAT, M., EMERY, V. C., GEDDES, C. C., GERNA, G., BEN-ISMAEIL, B., KAYE, S., MCGREGOR, A., MOSS, P. A., PUSZTAI, R., RAWLINSON, W. D., SCOTT, G. M., WILKINSON, G. W., SCHULZ, T. F. & DAVISON, A. J. 2008. Genotypic analysis of two hypervariable human cytomegalovirus genes. *J Med Virol,* 80**,** 1615-23.

BRITT, W. J. & BOPPANA, S. 2004. Human cytomegalovirus virion proteins. *Hum Immunol,* 65**,** 395-402.

BUCK, A. H., SANTOYO-LOPEZ, J., ROBERTSON, K. A., KUMAR, D. S., RECZKO, M. & GHAZAL, P. 2007. Discrete clusters of virus-encoded micrornas are associated with complementary strands of the genome and the 7.2-kilobase stable intron in murine cytomegalovirus. *J Virol,* 81**,** 13761-70.

BUSTIN, S. A., BEAULIEU, J. F., HUGGETT, J., JAGGI, R., KIBENGE, F. S., OLSVIK, P. A., PENNING, L. C. & TOEGEL, S. 2010. MIQE precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Mol Biol,* 11**,** 74.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTWER, C. T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55**,** 611-22.

BUXMANN, H., STACKELBERG, O. M., SCHLOSSER, R. L., ENDERS, G., GONSER, M., MEYER-WITTKOPF, M., HAMPRECHT, K. & ENDERS, M. 2012. Use of cytomegalovirus hyperimmunoglobulin for prevention of congenital cytomegalovirus disease: a retrospective analysis. *J Perinat Med,* 40**,** 439-46.

CAMPOS, A. B., RIBEIRO, J., BOUTOLLEAU, D. & SOUSA, H. 2016. Human cytomegalovirus antiviral drug resistance in hematopoietic stem cell transplantation: current state of the art. *Rev Med Virol,* 26**,** 161-82.

CANNON, M. J. & DAVIS, K. F. 2005. Washing our hands of the congenital cytomegalovirus disease epidemic. *BMC Public Health,* 5**,** 70.

CANNON, M. J., HYDE, T. B. & SCHMID, D. S. 2011. Review of cytomegalovirus shedding in bodily fluids and relevance to congenital cytomegalovirus infection. *Reviews in Medical Virology,* 21**,** 240-255.

CANNON, M. J., SCHMID, D. S. & HYDE, T. B. 2010. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol,* 20**,** 202-13.

CANNON, M. J., STOWELL, J. D., CLARK, R., DOLLARD, P. R., JOHNSON, D., MASK, K., STOVER, C., WU, K., AMIN, M., HENDLEY, W., GUO, J., SCHMID, D. S. & DOLLARD, S. C. 2014. Repeated measures study of weekly and daily cytomegalovirus shedding patterns in saliva and urine of healthy cytomegalovirus-seropositive children. *BMC Infect Dis,* 14**,** 569.

CHA, T. A., TOM, E., KEMBLE, G. W., DUKE, G. M., MOCARSKI, E. S. & SPAETE, R. R. 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J Virol,* 70**,** 78-83.

CHAN, G., NOGALSKI, M. T. & YUROCHKO, A. D. 2009. Activation of EGFR on monocytes is required for human cytomegalovirus entry and mediates cellular motility. *Proc Natl Acad Sci U S A,* 106**,** 22369-74.

CHANDLER, S. H., HOLMES, K. K., WENTWORTH, B. B., GUTMAN, L. T., WIESNER, P. J., ALEXANDER, E. R. & HANDSFIELD, H. H. 1985. The epidemiology of cytomegaloviral infection in women attending a sexually transmitted disease clinic. *J Infect Dis,* 152**,** 597-605.

CHANDLER, S. H. & MCDOUGALL, J. K. 1986. Comparison of restriction site polymorphisms among clinical isolates and laboratory strains of human cytomegalovirus. *J Gen Virol,* 67 (Pt 10)**,** 2179-92.

CHEE, M. S., BANKIER, A. T., BECK, S., BOHNI, R., BROWN, C. M., CERNY, R., HORSNELL, T., HUTCHISON, C. A., KOUZARIDES, T., MARTIGNETTI, J. A., PREDDIE, E., SATCHWELL, S. C., TOMLINSON, P., WESTON, K. M. & BARRELL, B. G. 1990. Analysis of the Protein-Coding Content of the Sequence of Human Cytomegalovirus Strain AD169. *In:* MCDOUGALL, J. K. (ed.) *Cytomegaloviruses.* Berlin, Heidelberg: Springer Berlin Heidelberg.

CHEERAN, M. C., LOKENSGARD, J. R. & SCHLEISS, M. R. 2009. Neuropathogenesis of congenital cytomegalovirus infection: disease mechanisms and prospects for intervention. *Clin Microbiol Rev,* 22**,** 99-126.

CHENG, J., KE, Q., JIN, Z., WANG, H., KOCHER, O., MORGAN, J. P., ZHANG, J. & CRUMPACKER, C. S. 2009. Cytomegalovirus Infection Causes an Increase of Arterial Blood Pressure. *PLOS Pathogens,* 5**,** e1000427.

CHEUNG, A. K., ABENDROTH, A., CUNNINGHAM, A. L. & SLOBEDMAN, B. 2006. Viral gene expression during the establishment of human cytomegalovirus latent infection in myeloid progenitor cells. *Blood,* 108**,** 3691-9.

CHEVILLOTTE, M., VON EINEM, J., MEIER, B. M., LIN, F. M., KESTLER, H. A. & MERTENS, T. 2010. A new tool linking human cytomegalovirus drug resistance mutations to resistance phenotypes. *Antiviral Res,* 85**,** 318-27.

CHIAVARINI, M., BRAGETTI, P., SENSINI, A., CENCI, E., CASTRONARI, R., ROSSI, M. J., FANTAUZZI, A. & MINELLI, L. 2011. Breastfeeding and transmission of cytomegalovirus to preterm infants. Case report and kinetic of CMV-DNA in breast milk. *Ital J Pediatr,* 37**,** 6.

CHINTU, C., MUDENDA, V., LUCAS, S., NUNN, A., LISHIMPI, K., MASWAHU, D., KASOLO, F., MWABA, P., BHAT, G., TERUNUMA, H., ZUMLA, A. & GROUP, U.-U. P. P. P.-M. S. 2002. Lung diseases at necropsy in African children dying from respiratory illnesses: a descriptive necropsy study. *Lancet,* 360**,** 985-90.

CHIUPPESI, F., WUSSOW, F., JOHNSON, E., BIAN, C., ZHUO, M., RAJAKUMAR, A., BARRY, P. A., BRITT, W. J., CHAKRABORTY, R. & DIAMOND, D. J. 2015. Vaccine-Derived Neutralizing Antibodies to the Human Cytomegalovirus gH/gL Pentamer Potently Block Primary Cytotrophoblast Infection. *J Virol,* 89**,** 11884-98.

CHOU, S. 2008. Cytomegalovirus UL97 mutations in the era of ganciclovir and maribavir. *Rev Med Virol,* 18**,** 233-46.

CHOU, S. 2009. Diverse cytomegalovirus UL27 mutations adapt to loss of viral UL97 kinase activity under maribavir. *Antimicrob Agents Chemother,* 53**,** 81-5.

CHOU, S., LURAIN, N. S., WEINBERG, A., CAI, G. Y., SHARMA, P. L. & CRUMPACKER, C. S. 1999. Interstrain variation in the human cytomegalovirus DNA polymerase sequence and its effect on genotypic diagnosis of antiviral drug resistance. Adult AIDS Clinical Trials Group CMV Laboratories. *Antimicrob Agents Chemother,* 43**,** 1500-2.

CHOU, S., MAROUSEK, G. I., SENTERS, A. E., DAVIS, M. G. & BIRON, K. K. 2004. Mutations in the human cytomegalovirus UL27 gene that confer resistance to maribavir. *J Virol,* 78**,** 7124-30.

CHRISTEN, L. 2014. Apparatus and methods for pasteurization of human milk. Google Patents.

CHRISTEN, L., LAI, C. T., HARTMANN, B., HARTMANN, P. E. & GEDDES, D. T. 2013. The effect of UV-C pasteurization on bacteriostatic properties and immunological proteins of donor human milk. *PLoS One,* 8**,** e85867.

CIFERRI, C., CHANDRAMOULI, S., DONNARUMMA, D., NIKITIN, P. A., CIANFROCCO, M. A., GERREIN, R., FEIRE, A. L., BARNETT, S. W., LILJA, A. E., RAPPUOLI, R., NORAIS, N., SETTEMBRE, E. C. & CARFI, A. 2015a. Structural and biochemical studies of HCMV gH/gL/gO and Pentamer reveal mutually exclusive cell entry complexes. *Proc Natl Acad Sci U S A,* 112**,** 1767-72.

CIFERRI, C., CHANDRAMOULI, S., LEITNER, A., DONNARUMMA, D., CIANFROCCO, M. A., GERREIN, R., FRIEDRICH, K., AGGARWAL, Y., PALLADINO, G., AEBERSOLD, R., NORAIS, N., SETTEMBRE, E. C. & CARFI, A. 2015b. Antigenic Characterization of the

HCMV gH/gL/gO and Pentamer Cell Entry Complexes Reveals Binding Sites for Potently Neutralizing Human Antibodies. *PLoS Pathog,* 11**,** e1005230.

COAQUETTE, A., BOURGEOIS, A., DIRAND, C., VARIN, A., CHEN, W. & HERBEIN, G. 2004. Mixed cytomegalovirus glycoprotein B genotypes in immunocompromised patients. *Clin Infect Dis,* 39**,** 155-61.

COLLIN, S. M., CHISENGA, M. M., KASONKA, L., HAWORTH, A., YOUNG, C., FILTEAU, S. & MURRAY, S. F. 2006. Factors associated with postpartum physical and mental morbidity among women with known HIV status in Lusaka, Zambia. *AIDS Care,* 18**,** 812-20.

COMPTON, T., NOWLIN, D. M. & COOPER, N. R. 1993. Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. *Virology,* 193**,** 834-41.

CORNELISSEN, M., EULER, Z., VAN DEN KERKHOF, T. L., VAN GILS, M. J., BOESER-NUNNINK, B. D., KOOTSTRA, N. A., ZORGDRAGER, F., SCHUITEMAKER, H., PRINS, J. M., SANDERS, R. W. & VAN DER KUYL, A. C. 2016. The Neutralizing Antibody Response in an Individual with Triple HIV-1 Infection Remains Directed at the First Infecting Subtype. *AIDS Res Hum Retroviruses,* 32**,** 1135-1142.

COURIVAUD, C., BAMOULID, J., CHALOPIN, J.-M., GAIFFE, E., TIBERGHIEN, P., SAAS, P. & DUCLOUX, D. 2013. Cytomegalovirus Exposure and Cardiovascular Disease in Kidney Transplant Recipients. *Journal of Infectious Diseases*.

CROUGH, T. & KHANNA, R. 2009. Immunobiology of human cytomegalovirus: from bench to bedside. *Clin Microbiol Rev,* 22**,** 76-98, Table of Contents.

CUI, X., LEE, R., ADLER, S. P. & MCVOY, M. A. 2013. Antibody inhibition of human cytomegalovirus spread in epithelial cell cultures. *J Virol Methods,* 192**,** 44-50.

CZANK, C., PRIME, D. K., HARTMANN, B., SIMMER, K. & HARTMANN, P. E. 2009. Retention of the immunological proteins of pasteurized human milk in relation to pasteurizer design and practice. *Pediatr Res,* 66**,** 374-9.

DAHL, H., FJAERTOFT, G., NORSTED, T., WANG, F. Z., MOUSAVI-JAZI, M. & LINDE, A. 1999. Reactivation of human herpesvirus 6 during pregnancy. *J Infect Dis,* 180**,** 2035-8.

DAIBATA, M. & MIYOSHI, I. 1999. Presence of human herpesvirus 6 DNA in cord blood cells. *J Infect Dis,* 179**,** 1046-7.

DAIMINGER, A., BADER, U. & ENDERS, G. 2005. Pre- and periconceptional primary cytomegalovirus infection: risk of vertical transmission and congenital disease. *BJOG,* 112**,** 166-72.

DAR, L., PATI, S. K., PATRO, A. R., DEORARI, A. K., RAI, S., KANT, S., BROOR, S., FOWLER, K. B., BRITT, W. J. & BOPPANA, S. B. 2008. Congenital cytomegalovirus infection in a highly seropositive semi-urban population in India. *Pediatr Infect Dis J,* 27**,** 841-3.

DARGAN, D. J., DOUGLAS, E., CUNNINGHAM, C., JAMIESON, F., STANTON, R. J., BALUCHOVA, K., MCSHARRY, B. P., TOMASEC, P., EMERY, V. C., PERCIVALLE, E., SARASINI, A., GERNA, G., WILKINSON, G. W. & DAVISON, A. J. 2010. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J Gen Virol,* 91**,** 1535-46.

DAVISON, A. J. 2007. Comparative analysis of the genomes. *In:* ARVIN, A., CAMPADELLI-FIUME, G., MOCARSKI, E., MOORE, P. S., ROIZMAN, B., WHITLEY, R. & YAMANISHI, K. (eds.) *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Cambridge: Cambridge University Press.

DAVISON, A. J., AKTER, P., CUNNINGHAM, C., DOLAN, A., ADDISON, C., DARGAN, D. J., HASSAN-WALKER, A. F., EMERY, V. C., GRIFFITHS, P. D. & WILKINSON, G. W. 2003a. Homology between the human cytomegalovirus RL11 gene family and human adenovirus E3 genes. *J Gen Virol,* 84**,** 657-63.

DAVISON, A. J., DOLAN, A., AKTER, P., ADDISON, C., DARGAN, D. J., ALCENDOR, D. J., MCGEOCH, D. J. & HAYWARD, G. S. 2003b. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol,* 84**,** 17-28.

DE BOLLE, L., NAESENS, L. & DE CLERCQ, E. 2005. Update on human herpesvirus 6 biology, clinical features, and therapy. *Clin Microbiol Rev,* 18**,** 217-45.

DE ORY, F., RAMIREZ, R., GARCIA COMAS, L., LEON, P., SAGUES, M. J. & SANZ, J. C. 2004. Is there a change in cytomegalovirus seroepidemiology in Spain? *Eur J Epidemiol,* 19**,** 85-9.

DE VRIES, J. J., VAN DER EIJK, A. A., WOLTHERS, K. C., RUSMAN, L. G., PAS, S. D., MOLENKAMP, R., CLAAS, E. C., KROES, A. C. & VOSSEN, A. C. 2012. Real-time PCR versus viral culture on urine as a gold standard in the diagnosis of congenital cytomegalovirus infection. *J Clin Virol,* 53**,** 167-70.

DE WOLF, F., SPIJKERMAN, I., SCHELLEKENS, P. T., LANGENDAM, M., KUIKEN, C., BAKKER, M., ROOS, M., COUTINHO, R., MIEDEMA, F. & GOUDSMIT, J. 1997. AIDS prognosis based on HIV-1 RNA, CD4+ T-cell count and function: markers with reciprocal predictive value over time after seroconversion. *AIDS,* 11**,** 1799-806.

DEPLEDGE, D. P., PALSER, A. L., WATSON, S. J., LAI, I. Y., GRAY, E. R., GRANT, P., KANDA, R. K., LEPROUST, E., KELLAM, P. & BREUER, J. 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One,* 6**,** e27805.

DERHOVANESSIAN, E., LARBI, A. & PAWELEC, G. 2009. Biomarkers of human immunosenescence: impact of Cytomegalovirus infection. *Curr Opin Immunol,* 21**,** 440-5.

DHURUVASAN, K., SIVASUBRAMANIAN, G. & PELLETT, P. E. 2011. Roles of host and viral microRNAs in human cytomegalovirus biology. *Virus Res,* 157**,** 180-92.

DOHNER, K. & SODEIK, B. 2005. The role of the cytoskeleton during viral infection. *Curr Top Microbiol Immunol,* 285**,** 67-108.

DOLAN, A., CUNNINGHAM, C., HECTOR, R. D., HASSAN-WALKER, A. F., LEE, L., ADDISON, C., DARGAN, D. J., MCGEOCH, D. J., GATHERER, D., EMERY, V. C., GRIFFITHS, P. D., SINZGER, C., MCSHARRY, B. P., WILKINSON, G. W. & DAVISON, A. J. 2004. Genetic content of wild-type human cytomegalovirus. *J Gen Virol,* 85**,** 1301-12.

DOLKEN, L., PEROT, J., COGNAT, V., ALIOUA, A., JOHN, M., SOUTSCHEK, J., RUZSICS, Z., KOSZINOWSKI, U., VOINNET, O. & PFEFFER, S. 2007. Mouse cytomegalovirus microRNAs dominate the cellular small RNA profile during lytic infection and show features of posttranscriptional regulation. *J Virol,* 81**,** 13771-82.

DOWD, J. B., AIELLO, A. E. & ALLEY, D. E. 2009. Socioeconomic disparities in the seroprevalence of cytomegalovirus infection in the US population: NHANES III. *Epidemiol Infect,* 137**,** 58-65.

DOYLE, M., ATKINS, J. T. & RIVERA-MATOS, I. R. 1996. Congenital cytomegalovirus infection in infants infected with human immunodeficiency virus type 1. *Pediatr Infect Dis J,* 15**,** 1102-6.

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**,** 14223-8.

DUNN, W., TRANG, P., ZHONG, Q., YANG, E., VAN BELLE, C. & LIU, F. 2005. Human cytomegalovirus expresses novel microRNAs during productive viral infection. *Cell Microbiol,* 7**,** 1684-95.

DUNNE, W. M., JR. & JEVON, M. 1993. Examination of human breast milk for evidence of human herpesvirus 6 by polymerase chain reaction. *J Infect Dis,* 168**,** 250.

DUROSE, J. B., LI, J., CHIEN, S. & SPECTOR, D. H. 2012. Infection of vascular endothelial cells with human cytomegalovirus under fluid shear stress reveals preferential entry and spread of virus in flow conditions simulating atheroprone regions of the artery. *J Virol,* 86**,** 13745-55.

ELDE, N. C., CHILD, S. J., EICKBUSH, M. T., KITZMAN, J. O., ROGERS, K. S., SHENDURE, J., GEBALLE, A. P. & MALIK, H. S. 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell,* 150**,** 831-841.

EMERY, V. C. 2011. Could a vaccine against immune-evading cytomegalovirus become a reality? *Expert Rev Vaccines,* 10**,** 1109-11.

EMERY, V. C., ZUCKERMAN, M., JACKSON, G., AITKEN, C., OSMAN, H., PAGLIUCA, A., POTTER, M., PEGGS, K., CLARK, A., BRITISH COMMITTEE FOR STANDARDS IN, H., BRITISH SOCIETY OF, B., MARROW, T. & NETWORK, U. K. V. 2013. Management of cytomegalovirus infection in haemopoietic stem cell transplantation. *Br J Haematol,* 162**,** 25-39.

ENDERS, G., DAIMINGER, A., BADER, U., EXLER, S. & ENDERS, M. 2011. Intrauterine transmission and clinical outcome of 248 pregnancies with primary cytomegalovirus infection in relation to gestational age. *J Clin Virol,* 52**,** 244-6.

ENDO, T., GOTO, K., ITO, K., SUGIURA, T., TERABE, K., CHO, S., NISHIYAMA, M., SUGIYAMA, K. & TOGARI, H. 2009. Detection of congenital cytomegalovirus infection using umbilical cord blood samples in a screening survey. *J Med Virol,* 81**,** 1773-6.

EWASCHUK, J. B., UNGER, S., HARVEY, S., O'CONNOR, D. L. & FIELD, C. J. 2011a. Effect of pasteurization on immune components of milk: implications for feeding preterm infants. *Appl Physiol Nutr Metab,* 36**,** 175-82.

EWASCHUK, J. B., UNGER, S., O'CONNOR, D. L., STONE, D., HARVEY, S., CLANDININ, M. T. & FIELD, C. J. 2011b. Effect of pasteurization on selected immune components of donated human breast milk. *J Perinatol,* 31**,** 593-8.

FEIRE, A. L., KOSS, H. & COMPTON, T. 2004. Cellular integrins function as entry receptors for human cytomegalovirus via a highly conserved disintegrin-like domain. *Proc Natl Acad Sci U S A,* 101**,** 15470-5.

FLAMAND, L., KOMAROFF, A. L., ARBUCKLE, J. H., MEDVECZKY, P. G. & ABLASHI, D. V. 2010. Review, part 1: Human herpesvirus-6-basic biology, diagnostic testing, and antiviral efficacy. *J Med Virol,* 82**,** 1560-8.

FONSECA, R. F., KAWAMURA, M. T., OLIVEIRA, J. A., TEIXEIRA, A., ALVES, G. & CARVALHO, M. D. G. D. C. 2012. The prevalence of human cytomegalovirus DNA in gliomas of Brazilian patients. *Memórias do Instituto Oswaldo Cruz,* 107**,** 953-954.

FORMAN, M., VALSAMAKIS, A. & ARAV-BOGER, R. 2012. Dried urine spots for detection and quantification of cytomegalovirus in newborns. *Diagn Microbiol Infect Dis,* 73**,** 326-9.

FOULON, I., NAESSENS, A., FOULON, W., CASTEELS, A. & GORDTS, F. 2008. Hearing loss in children with congenital cytomegalovirus infection in relation to the maternal trimester in which the maternal primary infection occurred. *Pediatrics,* 122**,** e1123-7.

FOUTS, A. E., CHAN, P., STEPHAN, J. P., VANDLEN, R. & FEIERBACH, B. 2012. Antibodies against the gH/gL/UL128/UL130/UL131 complex comprise the majority of the anti-cytomegalovirus (anti-CMV) neutralizing antibody response in CMV hyperimmune globulin. *J Virol,* 86**,** 7444-7.

FOWLER, K. B. & BOPPANA, S. B. 2006. Congenital cytomegalovirus (CMV) infection and hearing deficit. *J Clin Virol,* 35**,** 226-31.

FOWLER, K. B., DAHLE, A. J., BOPPANA, S. B. & PASS, R. F. 1999. Newborn hearing screening: will children with hearing loss caused by congenital cytomegalovirus infection be missed? *J Pediatr,* 135**,** 60-4.

FOWLER, K. B., STAGNO, S. & PASS, R. F. 2003. Maternal immunity and prevention of congenital cytomegalovirus infection. *JAMA,* 289**,** 1008-11.

FOWLER, K. B., STAGNO, S., PASS, R. F., BRITT, W. J., BOLL, T. J. & ALFORD, C. A. 1992. The outcome of congenital cytomegalovirus infection in relation to maternal antibody status. *N Engl J Med,* 326**,** 663-7.

FU, T. M., AN, Z. & WANG, D. 2014. Progress on pursuit of human cytomegalovirus vaccines for prevention of congenital infection and disease. *Vaccine,* 32**,** 2525-33.

FUJISAKI, H., TANAKA-TAYA, K., TANABE, H., HARA, T., MIYOSHI, H., OKADA, S. & YAMANISHI, K. 1998. Detection of human herpesvirus 7 (HHV-7) DNA in breast milk by polymerase chain reaction and prevalence of HHV-7 antibody in breast-fed and bottle-fed children. *J Med Virol,* 56**,** 275-9.

GANTT, S., CARLSSON, J., SHETTY, A. K., SEIDEL, K. D., QIN, X., MUTSVANGWA, J., MUSINGWINI, G., WOELK, G., ZIJENAH, L. S., KATZENSTEIN, D. A. & FRENKEL, L. M. 2008. Cytomegalovirus and Epstein-Barr virus in breast milk are associated with HIV-1 shedding but not with mastitis. *AIDS,* 22**,** 1453-60.

GANTT, S., HUANG, M. L., MAGARET, A., BUNTS, L., SELKE, S., WALD, A., ROSENTHAL, P. J., DORSEY, G. & CASPER, C. 2013. An artesunate-containing antimalarial treatment regimen did not suppress cytomegalovirus viremia. *J Clin Virol,* 58**,** 276-8.

GATHERER, D., SEIRAFIAN, S., CUNNINGHAM, C., HOLTON, M., DARGAN, D. J., BALUCHOVA, K., HECTOR, R. D., GALBRAITH, J., HERZYK, P., WILKINSON, G. W. & DAVISON, A. J. 2011. High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci U S A,* 108**,** 19755-60.

GERMI, R., MARIETTE, C., ALAIN, S., LUPO, J., THIEBAUT, A., BRION, J. P., EPAULARD, O., SAINT RAYMOND, C., MALVEZZI, P. & MORAND, P. 2014. Success and failure of artesunate treatment in five transplant recipients with disease caused by drug-resistant cytomegalovirus. *Antiviral Res,* 101**,** 57-61.

GESKUS, R. B., PRINS, M., HUBERT, J.-B., MIEDEMA, F., BERKHOUT, B., ROUZIOUX, C., DELFRAISSY, J.-F. & MEYER, L. 2007. The HIV RNA setpoint theory revisited. *Retrovirology,* 4**,** 65.

GIBSON, W. 2008. Structure and formation of the cytomegalovirus virion. *Curr Top Microbiol Immunol,* 325**,** 187-204.

GILBERT, C. & BOIVIN, G. 2005. Human cytomegalovirus resistance to antiviral drugs. *Antimicrob Agents Chemother,* 49**,** 873-83.

GINDES, L., TEPERBERG-OIKAWA, M., SHERMAN, D., PARDO, J. & RAHAV, G. 2008. Congenital cytomegalovirus infection following primary maternal infection in the third trimester. *BJOG,* 115**,** 830-5.

GOELZ, R., HIHN, E., HAMPRECHT, K., DIETZ, K., JAHN, G., POETS, C. & ELMLINGER, M. 2009. Effects of different CMV-heat-inactivation-methods on growth factors in human breast milk. *Pediatr Res,* 65**,** 458-61.

GOELZ, R., MEISNER, C., BEVOT, A., HAMPRECHT, K., KRAEGELOH-MANN, I. & POETS, C. F. 2013. Long-term cognitive and neurological outcome of preterm infants with postnatally acquired CMV infection through breast milk. *Arch Dis Child Fetal Neonatal Ed,* 98**,** F430-3.

GOMPELS, U. A., LARKE, N., SANZ-RAMOS, M., BATES, M., MUSONDA, K., MANNO, D., SIAME, J., MONZE, M., FILTEAU, S. & GROUP, C. S. 2012. Human cytomegalovirus infant infection adversely affects growth and development in maternally HIV-exposed and unexposed infants in Zambia. *Clin Infect Dis,* 54**,** 434-42.

GOODRUM, F., JORDAN, C. T., HIGH, K. & SHENK, T. 2002. Human cytomegalovirus gene expression during infection of primary hematopoietic progenitor cells: a model for latency. *Proc Natl Acad Sci U S A,* 99**,** 16255-60.

GOODRUM, F., REEVES, M., SINCLAIR, J., HIGH, K. & SHENK, T. 2007. Human cytomegalovirus sequences expressed in latently infected individuals promote a latent infection in vitro. *Blood,* 110**,** 937-45.

GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet,* 17**,** 333-51.

GORZER, I., KERSCHNER, H., JAKSCH, P., BAUER, C., SEEBACHER, G., KLEPETKO, W. & PUCHHAMMER-STOCKL, E. 2008. Virus load dynamics of individual CMV-genotypes in lung transplant recipients with mixed-genotype infections. *J Med Virol,* 80**,** 1405-14.

GRAHAME-CLARKE, C. 2005. Human cytomegalovirus, endothelial function and atherosclerosis. *Herpes: the journal of the IHMF,* 12**,** 42-45.

GRETCH, D. R., GEHRZ, R. C. & STINSKI, M. F. 1988a. Characterization of a human cytomegalovirus glycoprotein complex (gcI). *J Gen Virol,* 69 (Pt 6)**,** 1205-15.

GRETCH, D. R., KARI, B., RASMUSSEN, L., GEHRZ, R. C. & STINSKI, M. F. 1988b. Identification and characterization of three distinct families of glycoprotein complexes in the envelopes of human cytomegalovirus. *J Virol,* 62**,** 875-81.

GREY, F., ANTONIEWICZ, A., ALLEN, E., SAUGSTAD, J., MCSHEA, A., CARRINGTON, J. C. & NELSON, J. 2005. Identification and characterization of human cytomegalovirus-encoded microRNAs. *J Virol,* 79**,** 12095-9.

GREY, F., MEYERS, H., WHITE, E. A., SPECTOR, D. H. & NELSON, J. 2007. A human cytomegalovirus-encoded microRNA regulates expression of multiple viral genes involved in replication. *PLoS Pathog,* 3**,** e163.

GRIFFITHS-JONES, S. 2004. The microRNA Registry. *Nucleic Acids Res,* 32**,** D109-11.

GRIFFITHS-JONES, S., GROCOCK, R. J., VAN DONGEN, S., BATEMAN, A. & ENRIGHT, A. J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res,* 34**,** D140-4.

GRIFFITHS-JONES, S., SAINI, H. K., VAN DONGEN, S. & ENRIGHT, A. J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res,* 36**,** D154-8.

GRIFFITHS, P. D. 2002. Opening the Gates for vaccine development in the real world. *Rev Med Virol,* 12**,** 67-8.

GRIFFITHS, P. D. 2012. Burden of disease associated with human cytomegalovirus and prospects for elimination by universal immunisation. *Lancet Infect Dis,* 12**,** 790-8.

GRIFFITHS, P. D. 2014. Can the world afford to eliminate congenital rubella? *Rev Med Virol*.

GRIFFITHS, P. D., STANTON, A., MCCARRELL, E., SMITH, C., OSMAN, M., HARBER, M., DAVENPORT, A., JONES, G., WHEELER, D. C., O'BEIRNE, J., THORBURN, D., PATCH, D., ATKINSON, C. E., PICHON, S., SWENY, P., LANZMAN, M., WOODFORD, E., ROTHWELL, E., OLD, N., KINYANJUI, R., HAQUE, T., ATABANI, S., LUCK, S., PRIDEAUX, S., MILNE, R. S., EMERY, V. C. & BURROUGHS, A. K. 2011. Cytomegalovirus glycoprotein-B vaccine with MF59 adjuvant in transplant recipients: a phase 2 randomised placebo-controlled trial. *Lancet,* 377**,** 1256-63.

GRIFFITHS, P. D. & WALTER, S. 2005. Cytomegalovirus. *Curr Opin Infect Dis,* 18**,** 241-5.

GROER, M. E., JEVITT, C. & JI, M. 2015. Immune changes and dysphoric moods across the postpartum. *Am J Reprod Immunol,* 73**,** 193-8.

GROER, M. W., DAVIS, M. W., SMITH, K., CASEY, K., KRAMER, V. & BUKOVSKY, E. 2005. Immunity, Inflammation and Infection in Post-partum Breast and Formula Feeders. *American Journal of Reproductive Immunology,* 54**,** 222-231.

GROSJEAN, J., TRAPES, L., HANTZ, S., MENGELLE, C., VIREY, B., UNDREINER, F., MESSAGER, V., DENIS, F., MARIN, B. & ALAIN, S. 2014. Human cytomegalovirus quantification in toddlers saliva from day care centers and emergency unit: a feasibility study. *J Clin Virol,* 61**,** 371-7.

GUMBO, H., CHASEKWA, B., CHURCH, J. A., NTOZINI, R., MUTASA, K., HUMPHREY, J. H. & PRENDERGAST, A. J. 2014. Congenital and Postnatal CMV and EBV Acquisition in HIV-Infected Zimbabwean Infants. *PLoS ONE,* 9**,** e114870.

HAHN, G., JORES, R. & MOCARSKI, E. S. 1998. Cytomegalovirus remains latent in a common precursor of dendritic and myeloid cells. *Proc Natl Acad Sci U S A,* 95**,** 3937-42.

HAHN, G., REVELLO, M. G., PATRONE, M., PERCIVALLE, E., CAMPANINI, G., SARASINI, A., WAGNER, M., GALLINA, A., MILANESI, G., KOSZINOWSKI, U., BALDANTI, F. & GERNA, G. 2004. Human cytomegalovirus UL131-128 genes are indispensable for virus growth in endothelial cells and virus transfer to leukocytes. *J Virol,* 78**,** 10023-33.

HALL, C. B., CASERTA, M. T., SCHNABEL, K., SHELLEY, L. M., MARINO, A. S., CARNAHAN, J. A., YOO, C., LOFTHUS, G. K. & MCDERMOTT, M. P. 2008. Chromosomal integration of human herpesvirus 6 is the major mode of congenital human herpesvirus 6 infection. *Pediatrics,* 122**,** 513-20.

HALL, C. B., CASERTA, M. T., SCHNABEL, K. C., BOETTRICH, C., MCDERMOTT, M. P., LOFTHUS, G. K., CARNAHAN, J. A. & DEWHURST, S. 2004. Congenital infections with human herpesvirus 6 (HHV6) and human herpesvirus 7 (HHV7). *J Pediatr,* 145**,** 472-7.

HALL, C. B., CASERTA, M. T., SCHNABEL, K. C., SHELLEY, L. M., CARNAHAN, J. A., MARINO, A. S., YOO, C. & LOFTHUS, G. K. 2010. Transplacental congenital human herpesvirus 6 infection caused by maternal chromosomally integrated virus. *J Infect Dis,* 201**,** 505-7.

HAMPRECHT, K., GOELZ, R. & MASCHMANN, J. 2005. Breast milk and cytomegalovirus infection in preterm infants. *Early Hum Dev,* 81**,** 989-96.

HAMPRECHT, K., MASCHMANN, J., JAHN, G., POETS, C. F. & GOELZ, R. 2008. Cytomegalovirus transmission to preterm infants during lactation. *J Clin Virol,* 41**,** 198-205.

HAMPRECHT, K., MASCHMANN, J., MULLER, D., DIETZ, K., BESENTHAL, I., GOELZ, R., MIDDELDORP, J. M., SPEER, C. P. & JAHN, G. 2004a. Cytomegalovirus (CMV) inactivation in breast milk: reassessment of pasteurization and freeze-thawing. *Pediatr Res,* 56**,** 529-35.

HAMPRECHT, K., MASCHMANN, J., MÜLLER, D., DIETZ, K., BESENTHAL, I., GOELZ, R., MIDDELDORP, J. M., SPEER, C. P. & JAHN, G. 2004b. Cytomegalovirus (CMV) inactivation in breast milk: reassessment of pasteurization and freeze-thawing. *Pediatr Res,* 56**,** 529-35.

HAMPRECHT, K., MASCHMANN, J., VOCHEM, M., DIETZ, K., SPEER, C. P. & JAHN, G. 2001. Epidemiology of transmission of cytomegalovirus from mother to preterm infant by breastfeeding. *Lancet,* 357**,** 513-8.

HAMPRECHT, K., VOCHEM, M., BAUMEISTER, A., BONIEK, M., SPEER, C. P. & JAHN, G. 1998. Detection of cytomegaloviral DNA in human milk cells and cell free milk whey by nested PCR. *J Virol Methods,* 70**,** 167-76.

HAMPRECHT, K., WITZEL, S., MASCHMANN, J., DIETZ, K., BAUMEISTER, A., MIKELER, E., GOELZ, R., SPEER, C. P. & JAHN, G. 2003. Rapid detection and quantification of cell free cytomegalovirus by a high-speed centrifugation-based microculture assay: comparison to longitudinally analyzed viral DNA load and pp67 late transcript during lactation. *J Clin Virol,* 28**,** 303-16.

HARGETT, D. & SHENK, T. E. 2010. Experimental human cytomegalovirus latency in CD14+ monocytes. *Proc Natl Acad Sci U S A,* 107**,** 20039-44.

HASSAN-WALKER, A. F., OKWUADI, S., LEE, L., GRIFFITHS, P. D. & EMERY, V. C. 2004. Sequence variability of the alpha-chemokine UL146 from clinical strains of human cytomegalovirus. *J Med Virol,* 74**,** 573-9.

HAWKINS, C. & CROUL, S. 2011. Viruses and human brain tumors: cytomegalovirus enters the fray. *The Journal of Clinical Investigation,* 121**,** 3831-3833.

HAYES, K., DANKS, D. M., GIBAS, H. & JACK, I. 1972. Cytomegalovirus in human milk. *N Engl J Med,* 287**,** 177-8.

HE, L. & HANNON, G. J. 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet,* 5**,** 522-31.

HE, Z., HE, Y. S., KIM, Y., CHU, L., OHMSTEDE, C., BIRON, K. K. & COEN, D. M. 1997. The human cytomegalovirus UL97 protein is a protein kinase that autophosphorylates on serines and threonines. *J Virol,* 71**,** 405-11.

HECKER, M., QIU, D., MARQUARDT, K., BEIN, G. & HACKSTEIN, H. 2004. Continuous cytomegalovirus seroconversion in a large group of healthy blood donors. *Vox Sang,* 86**,** 41-4.

HEINEMAN, T. C., SCHLEISS, M., BERNSTEIN, D. I., SPAETE, R. R., YAN, L., DUKE, G., PRICHARD, M., WANG, Z., YAN, Q., SHARP, M. A., KLEIN, N., ARVIN, A. M. & KEMBLE, G. 2006. A phase 1 study of 4 live, recombinant human cytomegalovirus Towne/Toledo chimeric vaccines. *J Infect Dis,* 193**,** 1350-60.

HERRIOT, R. & GRAY, E. S. 1994. Images in clinical medicine. Owl's-eye cells. *N Engl J Med,* 331**,** 649.

HO, M. 2008. The history of cytomegalovirus and its diseases. *Med Microbiol Immunol,* 197**,** 65-73.

HOBOM, U., BRUNE, W., MESSERLE, M., HAHN, G. & KOSZINOWSKI, U. H. 2000. Fast screening procedures for random transposon libraries of cloned herpesvirus genomes: mutational analysis of human cytomegalovirus envelope glycoprotein genes. *J Virol,* 74**,** 7720-9.

HOSTETLER, K. Y. 2009. Alkoxyalkyl prodrugs of acyclic nucleoside phosphonates enhance oral antiviral activity and reduce toxicity: current state of the art. *Antiviral Res,* 82**,** A84-98.

HUMAR, A., KUMAR, D., GILBERT, C. & BOIVIN, G. 2003. Cytomegalovirus (CMV) glycoprotein B genotypes and response to antiviral therapy, in solid-organ-transplant recipients with CMV disease. *J Infect Dis,* 188**,** 581-4.

HUTCHINSON, N. I. & TOCCI, M. J. 1986. Characterization of a major early gene from the human cytomegalovirus long inverted repeat; predicted amino acid sequence of a 30-kDa protein encoded by the 1.2-kb mRNA. *Virology,* 155**,** 172-82.

ICTV. 2016. *Virus Taxonomy: 2015 Release. EC 47, London, UK, July 2015; Email ratification 2016 (MSL #30)* [Online]. International Committee on Taxonomy of Viruses (ICTV). Available: http://www.ictvonline.org/virusTaxonomy.asp [Accessed June 2 2016].

INGOLIA, N. T., BRAR, G. A., STERN-GINOSSAR, N., HARRIS, M. S., TALHOUARNE, G. J., JACKSON, S. E., WILLS, M. R. & WEISSMAN, J. S. 2014. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep,* 8**,** 1365-79.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-45.

INVESTEC. 2015. *It's all in the genes* [Online]. Available: http://privateoffice.investec.co.uk/research-and-insights/insights/vision_next_generation_sequencing.html [Accessed March 20 2016].

ISAACSON, M. K., FEIRE, A. L. & COMPTON, T. 2007. Epidermal growth factor receptor is not required for human cytomegalovirus entry or signaling. *J Virol,* 81**,** 6241-7.

IWASENKO, J. M., HOWARD, J., ARBUCKLE, S., GRAF, N., HALL, B., CRAIG, M. E. & RAWLINSON, W. D. 2011. Human cytomegalovirus infection is detected frequently in stillbirths and is associated with fetal thrombotic vasculopathy. *J Infect Dis,* 203**,** 1526-33.

JACOBSON, M. A., ADLER, S. P., SINCLAIR, E., BLACK, D., SMITH, A., CHU, A., MOSS, R. B. & WLOCH, M. K. 2009. A CMV DNA vaccine primes for memory immune responses to live-attenuated CMV (Towne strain). *Vaccine,* 27**,** 1540-8.

JIM, W. T., SHU, C. H., CHIU, N. C., CHANG, J. H., HUNG, H. Y., PENG, C. C., KAO, H. A., WEI, T. Y., CHIANG, C. L. & HUANG, F. Y. 2009. High cytomegalovirus load and prolonged virus excretion in breast milk increase risk for viral acquisition by very low birth weight infants. *Pediatr Infect Dis J,* 28**,** 891-4.

JIM, W. T., SHU, C. H., CHIU, N. C., KAO, H. A., HUNG, H. Y., CHANG, J. H., PENG, C. C., HSIEH, W. S., LIU, K. C. & HUANG, F. Y. 2004. Transmission of cytomegalovirus from mothers to preterm infants by breast milk. *Pediatr Infect Dis J,* 23**,** 848-51.

KABANOVA, A., MARCANDALLI, J., ZHOU, T., BIANCHI, S., BAXA, U., TSYBOVSKY, Y., LILLERI, D., SILACCI-FREGNI, C., FOGLIERINI, M., FERNANDEZ-RODRIGUEZ, B. M., DRUZ, A., ZHANG, B., GEIGER, R., PAGANI, M., SALLUSTO, F., KWONG, P. D., CORTI, D., LANZAVECCHIA, A. & PEREZ, L. 2016. Platelet-derived growth factor-alpha receptor is the cellular receptor for human cytomegalovirus gHgLgO trimer. *Nat Microbiol,* 2016.

KAINTH, M. K. & CASERTA, M. T. 2011. Molecular diagnostic tests for human herpesvirus 6. *Pediatr Infect Dis J,* 30**,** 604-5.

KAPTEIN, S. J., EFFERTH, T., LEIS, M., RECHTER, S., AUEROCHS, S., KALMER, M., BRUGGEMAN, C. A., VINK, C., STAMMINGER, T. & MARSCHALL, M. 2006. The anti-malaria drug artesunate inhibits replication of cytomegalovirus in vitro and in vivo. *Antiviral Res,* 69**,** 60-9.

KARI, B. & GEHRZ, R. 1992. A human cytomegalovirus glycoprotein complex designated gC-II is a major heparin-binding component of the envelope. *J Virol,* 66**,** 1761-4.

KARI, B., LUSSENHOP, N., GOERTZ, R., WABUKE-BUNOTI, M., RADEKE, R. & GEHRZ, R. 1986. Characterization of monoclonal antibodies reactive to several biochemically distinct human cytomegalovirus glycoprotein complexes. *J Virol,* 60**,** 345-52.

KASONKA, L., MAKASA, M., MARSHALL, T., CHISENGA, M., SINKALA, M., CHINTU, C., KASEBA, C., KASOLO, F., GITAU, R., TOMKINS, A., MURRAY, S. & FILTEAU, S. 2006. Risk factors for subclinical mastitis among HIV-infected and uninfected women in Lusaka, Zambia. *Paediatr Perinat Epidemiol,* 20**,** 379-91.

KAUL, D. R., STOELBEN, S., COBER, E., OJO, T., SANDUSKY, E., LISCHKA, P., ZIMMERMANN, H. & RUBSAMEN-SCHAEFF, H. 2011. First report of successful treatment of multidrug-resistant cytomegalovirus disease with the novel anti-CMV compound AIC246. *Am J Transplant,* 11**,** 1079-84.

KAYE, S., MILES, D., ANTOINE, P., BURNY, W., OJUOLA, B., KAYE, P., ROWLAND-JONES, S., WHITTLE, H., VAN DER SANDE, M. & MARCHANT, A. 2008. Virological and immunological correlates of mother-to-child transmission of cytomegalovirus in The Gambia. *J Infect Dis,* 197**,** 1307-14.

KENNESON, A. & CANNON, M. J. 2007. Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. *Rev Med Virol,* 17**,** 253-76.

KEYES, L. R., HARGETT, D., SOLAND, M., BEGO, M. G., ROSSETTO, C. C., ALMEIDA-PORADA, G. & ST JEOR, S. 2012. HCMV protein LUNA is required for viral reactivation from latently infected primary CD14(+) cells. *PLoS One,* 7**,** e52827.

KHARFAN-DABAJA, M. A., BOECKH, M., WILCK, M. B., LANGSTON, A. A., CHU, A. H., WLOCH, M. K., GUTERWILL, D. F., SMITH, L. R., ROLLAND, A. P. & KENNEY, R. T. 2012. A novel therapeutic cytomegalovirus DNA vaccine in allogeneic haemopoietic stem-cell transplantation: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Infect Dis,* 12**,** 290-9.

KIMBERLIN, D. W., JESTER, P. M., SANCHEZ, P. J., AHMED, A., ARAV-BOGER, R., MICHAELS, M. G., ASHOURI, N., ENGLUND, J. A., ESTRADA, B., JACOBS, R. F., ROMERO, J. R., SOOD, S. K., WHITWORTH, M. S., ABZUG, M. J., CASERTA, M. T., FOWLER, S., LUJAN-ZILBERMANN, J., STORCH, G. A., DEBIASI, R. L., HAN, J. Y., PALMER, A., WEINER, L. B., BOCCHINI, J. A., DENNEHY, P. H., FINN, A., GRIFFITHS, P. D., LUCK, S., GUTIERREZ, K., HALASA, N., HOMANS, J., SHANE, A. L., SHARLAND, M., SIMONSEN, K., VANCHIERE, J. A., WOODS, C. R., SABO, D. L., ABAN, I., KUO, H., JAMES, S. H., PRICHARD, M. N., GRIFFIN, J., GILES, D., ACOSTA, E. P., WHITLEY, R. J., NATIONAL INSTITUTE OF, A. & INFECTIOUS DISEASES COLLABORATIVE ANTIVIRAL STUDY, G. 2015. Valganciclovir for symptomatic congenital cytomegalovirus disease. *N Engl J Med,* 372**,** 933-43.

KIMBERLIN, D. W., LIN, C. Y., SANCHEZ, P. J., DEMMLER, G. J., DANKNER, W., SHELTON, M., JACOBS, R. F., VAUDRY, W., PASS, R. F., KIELL, J. M., SOONG, S. J., WHITLEY, R. J., NATIONAL INSTITUTE OF, A. & INFECTIOUS DISEASES COLLABORATIVE ANTIVIRAL STUDY, G. 2003. Effect of ganciclovir therapy on hearing in symptomatic congenital cytomegalovirus disease involving the central nervous system: a randomized, controlled trial. *J Pediatr,* 143**,** 16-25.

KIRCHER, M. & KELSO, J. 2010. High-throughput DNA sequencing--concepts and limitations. *Bioessays,* 32**,** 524-36.

KONDO, K., KANESHIMA, H. & MOCARSKI, E. S. 1994. Human cytomegalovirus latent infection of granulocyte-macrophage progenitors. *Proc Natl Acad Sci U S A,* 91**,** 11879-83.

KONDO, K., KONDO, T., OKUNO, T., TAKAHASHI, M. & YAMANISHI, K. 1991. Latent human herpesvirus 6 infection of human monocytes/macrophages. *J Gen Virol,* 72 (Pt 6)**,** 1401-8.

KOTHARI, A., RAMACHANDRAN, V. G., GUPTA, P., SINGH, B. & TALWAR, V. 2002. Seroprevalence of cytomegalovirus among voluntary blood donors in Delhi, India. *J Health Popul Nutr,* 20**,** 348-51.

KOVACS, A., SCHLUCHTER, M., EASLEY, K., DEMMLER, G., SHEARER, W., LA RUSSA, P., PITT, J., COOPER, E., GOLDFARB, J., HODES, D., KATTAN, M. & MCINTOSH, K. 1999. Cytomegalovirus infection and HIV-1 disease progression in infants born to HIV-1-infected women. Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted HIV Infection Study Group. *N Engl J Med,* 341**,** 77-84.

KOYANO, S., ARAKI, A., HIRANO, Y., FUJIEDA, K., SUZUTANI, T., YAGYU, K., MURONO, K. & INOUE, N. 2004. Retrospective diagnosis of congenital cytomegalovirus infection using dried umbilical cords. *Pediatr Infect Dis J,* 23**,** 481-2.

KOYANO, S., INOUE, N., NAGAMORI, T., YAN, H., ASANUMA, H., YAGYU, K., OSAKI, M., SEIWA, C. & FUJIEDA, K. 2009. Dried umbilical cords in the retrospective diagnosis of congenital cytomegalovirus infection as a cause of developmental delays. *Clin Infect Dis,* 48**,** e93-5.

KOZOMARA, A. & GRIFFITHS-JONES, S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res,* 39**,** D152-7.

KOZOMARA, A. & GRIFFITHS-JONES, S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res,* 42**,** D68-73.

KRAMER, M. S. & KAKUMA, R. 2012. Optimal duration of exclusive breastfeeding. *Cochrane Database Syst Rev***,** CD003517.

KULESZA, C. A. & SHENK, T. 2004. Human cytomegalovirus 5-kilobase immediate-early RNA is a stable intron. *J Virol,* 78**,** 13182-9.

KULESZA, C. A. & SHENK, T. 2006. Murine cytomegalovirus encodes a stable intron that facilitates persistent replication in the mouse. *Proc Natl Acad Sci U S A,* 103**,** 18302-7.

LA TORRE, R., NIGRO, G., MAZZOCCO, M., BEST, A. M. & ADLER, S. P. 2006. Placental enlargement in women with primary maternal cytomegalovirus infection is associated with fetal and neonatal disease. *Clin Infect Dis,* 43**,** 994-1000.

LANDGRAF, P., RUSU, M., SHERIDAN, R., SEWER, A., IOVINO, N., ARAVIN, A., PFEFFER, S., RICE, A., KAMPHORST, A. O., LANDTHALER, M., LIN, C., SOCCI, N. D., HERMIDA, L., FULCI, V., CHIARETTI, S., FOA, R., SCHLIWKA, J., FUCHS, U., NOVOSEL, A., MULLER, R. U., SCHERMER, B., BISSELS, U., INMAN, J., PHAN, Q., CHIEN, M., WEIR, D. B.,

CHOKSI, R., DE VITA, G., FREZZETTI, D., TROMPETER, H. I., HORNUNG, V., TENG, G., HARTMANN, G., PALKOVITS, M., DI LAURO, R., WERNET, P., MACINO, G., ROGLER, C. E., NAGLE, J. W., JU, J., PAPAVASILIOU, F. N., BENZING, T., LICHTER, P., TAM, W., BROWNSTEIN, M. J., BOSIO, A., BORKHARDT, A., RUSSO, J. J., SANDER, C., ZAVOLAN, M. & TUSCHL, T. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell,* 129**,** 1401-14.

LANDOLFO, S., GARIGLIO, M., GRIBAUDO, G. & LEMBO, D. 2003. The human cytomegalovirus. *Pharmacol Ther,* 98**,** 269-97.

LAU, B., POOLE, E., KRISHNA, B., MONTANUY, I., WILLS, M. R., MURPHY, E. & SINCLAIR, J. 2016a. Corrigendum: The Expression of Human Cytomegalovirus MicroRNA MiR-UL148D during Latent Infection in Primary Myeloid Cells Inhibits Activin A-triggered Secretion of IL-6. *Sci Rep,* 6**,** 33771.

LAU, B., POOLE, E., KRISHNA, B., SELLART, I., WILLS, M. R., MURPHY, E. & SINCLAIR, J. 2016b. The Expression of Human Cytomegalovirus MicroRNA MiR-UL148D during Latent Infection in Primary Myeloid Cells Inhibits Activin A-triggered Secretion of IL-6. *Sci Rep,* 6**,** 31205.

LAU, P. K., WOODS, M. L., RATANJEE, S. K. & JOHN, G. T. 2011. Artesunate is ineffective in controlling valganciclovir-resistant cytomegalovirus infection. *Clin Infect Dis,* 52**,** 279.

LEMMERMANN, N. A., KRMPOTIC, A., PODLECH, J., BRIZIC, I., PRAGER, A., ADLER, H., KARBACH, A., WU, Y., JONJIC, S., REDDEHASE, M. J. & ADLER, B. 2015. Non-redundant and redundant roles of cytomegalovirus gH/gL complexes in host organ entry and intra-tissue spread. *PLoS Pathog,* 11**,** e1004640.

LEN, O., GAVALDÀ, J., MARÍA AGUADO, J., BORRELL, N., CERVERA, C., MIGUEL CISNEROS, J., CUERVAS-MONS, V., GURGUÍ, M., MARTIN-DÁVILA, P., MONTEJO, M., MUÑOZ, P., BOU, G., CARRATALÀ, J., TORRE-CISNEROS, J., PAHISSA, A. & RESITRA, O. B. O. 2008. Valganciclovir as Treatment for Cytomegalovirus Disease in Solid Organ Transplant Recipients. *Clinical Infectious Diseases,* 46**,** 20-27.

LEPILLER, Q., TRIPATHY, M. K., DI MARTINO, V., KANTELIP, B. & HERBEIN, G. 2011. Increased HCMV seroprevalence in patients with hepatocellular carcinoma. *Virology Journal,* 8**,** 485.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, H. & DURBIN, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics,* 26**,** 589-95.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, X., QIAN, D., JU, F. & WANG, B. 2015. Upregulation of Toll-like receptor 2 expression in colorectal cancer infected by human cytomegalovirus. *Oncol Lett,* 9**,** 365-370.

LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25**,** 1451-2.

LIESNARD, C., DONNER, C., BRANCART, F., GOSSELIN, F., DELFORGE, M. L. & RODESCH, F. 2000. Prenatal diagnosis of congenital cytomegalovirus infection: prospective study of 237 pregnancies at risk. *Obstet Gynecol,* 95**,** 881-8.

LILLERI, D., KABANOVA, A., REVELLO, M. G., PERCIVALLE, E., SARASINI, A., GENINI, E., SALLUSTO, F., LANZAVECCHIA, A., CORTI, D. & GERNA, G. 2013. Fetal human cytomegalovirus transmission correlates with delayed maternal antibodies to gH/gL/pUL128-130-131 complex during primary infection. *PLoS One,* 8**,** e59863.

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.

LOPO, S., VINAGRE, E., PALMINHA, P., PAIXAO, M. T., NOGUEIRA, P. & FREITAS, M. G. 2011. Seroprevalence to cytomegalovirus in the Portuguese population, 2002-2003. *Euro Surveill,* 16.

LUKASHCHUK, V., MCFARLANE, S., EVERETT, R. D. & PRESTON, C. M. 2008. Human cytomegalovirus protein pp71 displaces the chromatin-associated factor ATRX from nuclear domain 10 at early stages of infection. *J Virol,* 82**,** 12543-54.

LURAIN, N. S. & CHOU, S. 2010. Antiviral drug resistance of human cytomegalovirus. *Clin Microbiol Rev,* 23**,** 689-712.

LURAIN, N. S., FOX, A. M., LICHY, H. M., BHORADE, S. M., WARE, C. F., HUANG, D. D., KWAN, S. P., GARRITY, E. R. & CHOU, S. 2006. Analysis of the human cytomegalovirus genomic region from UL146 through UL147A reveals sequence hypervariability, genotypic stability, and overlapping transcripts. *Virol J,* 3**,** 4.

LURAIN, N. S., WEINBERG, A., CRUMPACKER, C. S., CHOU, S. & ADULT, A. C. T. G.-C. M. V. L. 2001. Sequencing of cytomegalovirus UL97 gene for genotypic antiviral resistance testing. *Antimicrob Agents Chemother,* 45**,** 2775-80.

MACAGNO, A., BERNASCONI, N. L., VANZETTA, F., DANDER, E., SARASINI, A., REVELLO, M. G., GERNA, G., SALLUSTO, F. & LANZAVECCHIA, A. 2010. Isolation of human monoclonal antibodies that potently neutralize human cytomegalovirus infection by targeting different epitopes on the gH/gL/UL128-131A complex. *J Virol,* 84**,** 1005-13.

MACH, M., KROPFF, B., DAL MONTE, P. & BRITT, W. 2000. Complex formation by human cytomegalovirus glycoproteins M (gpUL100) and N (gpUL73). *J Virol,* 74**,** 11881-92.

MACH, M., KROPFF, B., KRYZANIAK, M. & BRITT, W. 2005. Complex formation by glycoproteins M and N of human cytomegalovirus: structural and functional aspects. *J Virol,* 79**,** 2160-70.

MANICKLAL, S., EMERY, V. C., LAZZAROTTO, T., BOPPANA, S. B. & GUPTA, R. K. 2013. The "silent" global burden of congenital cytomegalovirus. *Clin Microbiol Rev,* 26**,** 86-102.

MANICKLAL, S., VAN NIEKERK, A. M., KROON, S. M., HUTTO, C., NOVAK, Z., PATI, S. K., CHOWDHURY, N., HSIAO, N. Y. & BOPPANA, S. B. 2014. Birth Prevalence of Congenital Cytomegalovirus Among Infants of HIV-Infected Women on Prenatal Antiretroviral Prophylaxis in South Africa. *Clin Infect Dis*.

MARGERIDON-THERMET, S., SHULMAN, N. S., AHMED, A., SHAHRIAR, R., LIU, T., WANG, C., HOLMES, S. P., BABRZADEH, F., GHARIZADEH, B., HANCZARUK, B., SIMEN, B. B., EGHOLM, M. & SHAFER, R. W. 2009. Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. *J Infect Dis,* 199**,** 1275-85.

MARINDA, E., HUMPHREY, J. H., ILIFF, P. J., MUTASA, K., NATHOO, K. J., PIWOZ, E. G., MOULTON, L. H., SALAMA, P., WARD, B. J. & GROUP, Z. S. 2007. Child mortality according to maternal and infant HIV status in Zimbabwe. *Pediatr Infect Dis J,* 26**,** 519-26.

MARTY, F. M. & BOECKH, M. 2011. Maribavir and human cytomegalovirus-what happened in the clinical trials and why might the drug have failed? *Curr Opin Virol,* 1**,** 555-62.

MARTY, F. M., LJUNGMAN, P., PAPANICOLAOU, G. A., WINSTON, D. J., CHEMALY, R. F., STRASFELD, L., YOUNG, J. A., RODRIGUEZ, T., MAERTENS, J., SCHMITT, M., EINSELE, H., FERRANT, A., LIPTON, J. H., VILLANO, S. A., CHEN, H., BOECKH, M. & MARIBAVIR -300 CLINICAL STUDY, G. 2011. Maribavir prophylaxis for prevention of cytomegalovirus disease in recipients of allogeneic stem-cell transplants: a phase 3, double-blind, placebo-controlled, randomised trial. *Lancet Infect Dis,* 11**,** 284-92.

MASCHMANN, J., GOELZ, R., WITZEL, S., STRITTMATTER, U., STEINMASSL, M., JAHN, G. & HAMPRECHT, K. 2015. Characterization of human breast milk leukocytes and their potential role in cytomegalovirus transmission to newborns. *Neonatology,* 107**,** 213-9.

MASCHMANN, J., HAMPRECHT, K., WEISSBRICH, B., DIETZ, K., JAHN, G. & SPEER, C. P. 2006. Freeze-thawing of breast milk does not prevent cytomegalovirus transmission to a preterm infant. *Arch Dis Child Fetal Neonatal Ed,* 91**,** F288-90.

MATTES, F. M., HAINSWORTH, E. G., HASSAN-WALKER, A. F., BURROUGHS, A. K., SWENY, P., GRIFFITHS, P. D. & EMERY, V. C. 2005. Kinetics of cytomegalovirus load decrease in solid-organ transplant recipients after preemptive therapy with valganciclovir. *J Infect Dis,* 191**,** 89-92.

MATTICK, C., DEWIN, D., POLLEY, S., SEVILLA-REYES, E., PIGNATELLI, S., RAWLINSON, W., WILKINSON, G., DAL MONTE, P. & GOMPELS, U. A. 2004. Linkage of human cytomegalovirus glycoprotein gO variant groups identified from worldwide clinical isolates with gN genotypes, implications for disease associations and evidence for N-terminal sites of positive selection. *Virology,* 318**,** 582-97.

MATTICK, J. S. & RINN, J. L. 2015. Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol,* 22**,** 5-7.

MCCLURE, E. M., DUDLEY, D. J., REDDY, U. M. & GOLDENBERG, R. L. 2010. Infectious causes of stillbirth: a clinical perspective. *Clin Obstet Gynecol,* 53**,** 635-45.

MCDONOUGH, S. H., STAPRANS, S. I. & SPECTOR, D. H. 1985. Analysis of the major transcripts encoded by the long repeat of human cytomegalovirus strain AD169. *J Virol,* 53**,** 711-8.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20**,** 1297-303.

MCVOY, M. A. 2013. Cytomegalovirus Vaccines. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America,* 57**,** S196-S199.

MELENDEZ, D. P. & RAZONABLE, R. R. 2015. Letermovir and inhibitors of the terminase complex: a promising new class of investigational antiviral drugs against human cytomegalovirus. *Infect Drug Resist,* 8**,** 269-77.

MENDELSON, M., MONARD, S., SISSONS, P. & SINCLAIR, J. 1996. Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors. *J Gen Virol,* 77 (Pt 12)**,** 3099-102.

MESHESHA, M. K., BENTWICH, Z., SOLOMON, S. A. & AVNI, Y. S. 2016. In vivo expression of human cytomegalovirus (HCMV) microRNAs during latency. *Gene,* 575**,** 101-7.

MICHEL, D. & MERTENS, T. 2004. The UL97 protein kinase of human cytomegalovirus and homologues in other herpesviruses: impact on virus and host. *Biochim Biophys Acta,* 1697**,** 169-80.

MITSUYA, Y., VARGHESE, V., WANG, C., LIU, T. F., HOLMES, S. P., JAYAKUMAR, P., GHARIZADEH, B., RONAGHI, M., KLEIN, D., FESSEL, W. J. & SHAFER, R. W. 2008. Minority human immunodeficiency virus type 1 variants in antiretroviral-naive persons with reverse transcriptase codon 215 revertant mutations. *J Virol,* 82**,** 10747-55.

MOCARSKI JR, E. S. 2007. Betaherpes viral genes and their functions. *In:* ARVIN, A., CAMPADELLI-FIUME, G., MOCARSKI, E., MOORE, P. S., ROIZMAN, B., WHITLEY, R. & YAMANISHI, K. (eds.) *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Cambridge: Cambridge University Press.

MOCARSKI JR, E. S. & PASS, R. F. 2008. Human Cytomegalovirus: General Features. *In:* MAHY, B. W. J. & VAN_REGENMORTEL, M. H. V. (eds.) *Encyclopedia of Virology.* Third ed. Oxford: Oxford Academic Press.

MOCARSKI JR, E. S., SHENK, T. & PASS, R. 2007. Cytomegalovirus. *In:* KNIPE, D. M. & HOWLEY, P. M. (eds.) *Field's Virology.* Fifth ed. Philadelphia: Lippincott Williams & Wilkins.

MORISSETTE, G. & FLAMAND, L. 2010. Herpesviruses and chromosomal integration. *J Virol,* 84**,** 12100-9.

MOUSAVI-JAZI, M., HOKEBERG, I., SCHLOSS, L., ZWEYGBERG-WIRGART, B., GRILLNER, L., LINDE, A. & BRYTTING, M. 2001. Sequence analysis of UL54 and UL97 genes and evaluation of antiviral susceptibility of human cytomegalovirus isolates obtained from kidney allograft recipients before and after treatment. *Transpl Infect Dis,* 3**,** 195-202.

MURPHY, E., YU, D., GRIMWOOD, J., SCHMUTZ, J., DICKSON, M., JARVIS, M. A., HAHN, G., NELSON, J. A., MYERS, R. M. & SHENK, T. E. 2003. Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proc Natl Acad Sci U S A,* 100**,** 14976-81.

MURPHY, J. C., FISCHLE, W., VERDIN, E. & SINCLAIR, J. H. 2002. Control of cytomegalovirus lytic gene expression by histone acetylation. *EMBO J,* 21**,** 1112-20.

MURRELL, I., TOMASEC, P., WILKIE, G. S., DARGAN, D. J., DAVISON, A. J. & STANTON, R. J. 2013. Impact of sequence variation in the UL128 locus on production of human cytomegalovirus in fibroblast and epithelial cells. *J Virol,* 87**,** 10489-500.

MURTHY, S., HAYWARD, G. S., WHEELAN, S., FORMAN, M. S., AHN, J. H., PASS, R. F. & ARAV-BOGER, R. 2011. Detection of a single identical cytomegalovirus (CMV) strain in recently seroconverted young women. *PLoS One,* 6**,** e15949.

MUSONDA, K. G., NYONDA, M., FILTEAU, S., KASONKA, L., MONZE, M. & GOMPELS, U. A. 2016. Increased Cytomegalovirus Secretion and Risks of Infant Infection by Breastfeeding Duration From Maternal Human Immunodeficiency Virus Positive Compared to Negative Mothers in Sub-Saharan Africa. *J Pediatric Infect Dis Soc,* 5**,** 138-46.

MUSSI-PINHATA, M. M., YAMAMOTO, A. Y., MOURA BRITO, R. M., DE LIMA ISAAC, M., DE CARVALHO E OLIVEIRA, P. F., BOPPANA, S. & BRITT, W. J. 2009. Birth prevalence and natural history of congenital cytomegalovirus infection in a highly seroimmune population. *Clin Infect Dis,* 49**,** 522-8.

MWAANZA, N., CHILUKUTU, L., TEMBO, J., KABWE, M., MUSONDA, K., KAPASA, M., CHABALA, C., SINYANGWE, S., MWABA, P., ZUMLA, A. & BATES, M. 2014. High Rates of Congenital Cytomegalovirus Infection Linked With Maternal HIV Infection Among Neonatal Admissions at a Large Referral Center in Sub-Saharan Africa. *Clin Infect Dis,* 58**,** 728-35.

NASSETTA, L., KIMBERLIN, D. & WHITLEY, R. 2009. Treatment of congenital cytomegalovirus infection: implications for future therapeutic strategies. *J Antimicrob Chemother,* 63**,** 862-7.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI). 2015. *Release Notes For GenBank Release 211* [Online]. Available: https://www.ncbi.nlm.nih.gov/genbank/release/211 [Accessed December 18 2015].

NG, K. R., LI, J. Y. & GLEADLE, J. M. 2015. Human cytomegalovirus encoded microRNAs: hitting targets. *Expert Rev Anti Infect Ther,* 13**,** 1469-79.

NIGRO, G., ADLER, S. P., LA TORRE, R., BEST, A. M. & CONGENITAL CYTOMEGALOVIRUS COLLABORATING, G. 2005. Passive immunization during pregnancy for congenital cytomegalovirus infection. *N Engl J Med,* 353**,** 1350-62.

NIGRO, G., LA TORRE, R., PENTIMALLI, H., TAVERNA, P., LITUANIA, M., DE TEJADA, B. M. & ADLER, S. P. 2008. Regression of fetal cerebral abnormalities by primary cytomegalovirus infection following hyperimmunoglobulin therapy. *Prenat Diagn,* 28**,** 512-7.

NORIEGA, V. M., HAYE, K. K., KRAUS, T. A., KOWALSKY, S. R., GE, Y., MORAN, T. M. & TORTORELLA, D. 2014. Human cytomegalovirus modulates monocyte-mediated innate immune responses during short-term experimental latency in vitro. *J Virol,* 88**,** 9391-405.

NOVAKOVA, V., HAMPRECHT, K., MULLER, A. M., ARELLANO-GALINDO, J., EHLEN, M. & HORNEFF, G. 2014. Severe postnatal CMV colitis with an extensive colonic stenosis in a 2-month-old male immunocompetent term infant infected via breast milk. *J Clin Virol,* 59**,** 259-63.

NOYOLA, D. E., DEMMLER, G. J., NELSON, C. T., GRIESSER, C., WILLIAMSON, W. D., ATKINS, J. T., ROZELLE, J., TURCICH, M., LLORENTE, A. M., SELLERS-VINSON, S., REYNOLDS, A., BALE, J. F., JR., GERSON, P., YOW, M. D. & HOUSTON CONGENITAL, C. M. V. L. S. G. 2001. Early predictors of neurodevelopmental outcome in symptomatic congenital cytomegalovirus infection. *J Pediatr,* 138**,** 325-31.

NOYOLA, D. E., DEMMLER, G. J., WILLIAMSON, W. D., GRIESSER, C., SELLERS, S., LLORENTE, A., LITTMAN, T., WILLIAMS, S., JARRETT, L. & YOW, M. D. 2000. Cytomegalovirus urinary excretion and long term outcome in children with congenital cytomegalovirus infection. Congenital CMV Longitudinal Study Group. *Pediatr Infect Dis J,* 19**,** 505-10.

O'CONNOR, C. M. & SHENK, T. 2012. Human cytomegalovirus pUL78 G protein-coupled receptor homologue is required for timely cell entry in epithelial cells but not fibroblasts. *J Virol,* 86**,** 11425-33.

OLIVER, S. E., CLOUD, G. A., SANCHEZ, P. J., DEMMLER, G. J., DANKNER, W., SHELTON, M., JACOBS, R. F., VAUDRY, W., PASS, R. F., SOONG, S. J., WHITLEY, R. J., KIMBERLIN, D. W. & NATIONAL INSTITUTE OF ALLERGY, I. D. C. A. S. G. 2009. Neurodevelopmental outcomes following ganciclovir therapy in symptomatic congenital cytomegalovirus infections involving the central nervous system. *J Clin Virol,* 46 Suppl 4**,** S22-6.

OTTO, T. D., DILLON, G. P., DEGRAVE, W. S. & BERRIMAN, M. 2011. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res,* 39**,** e57.

PANG, X., HUMAR, A. & PREIKSAITIS, J. K. 2008. Concurrent genotyping and quantitation of cytomegalovirus gB genotypes in solid-organ-transplant recipients by use of a real-time PCR assay. *J Clin Microbiol,* 46**,** 4004-10.

PAPADOPOULOS, J. S. & AGARWALA, R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics,* 23**,** 1073-9.

PASS, R. F. 2009. Development and evidence for efficacy of CMV glycoprotein B vaccine with MF59 adjuvant. *J Clin Virol,* 46 Suppl 4**,** S73-6.

PASS, R. F., FOWLER, K. B., BOPPANA, S. B., BRITT, W. J. & STAGNO, S. 2006. Congenital cytomegalovirus infection following first trimester maternal infection: symptoms at birth and outcome. *J Clin Virol,* 35**,** 216-20.

PASS, R. F., ZHANG, C., EVANS, A., SIMPSON, T., ANDREWS, W., HUANG, M. L., COREY, L., HILL, J., DAVIS, E., FLANIGAN, C. & CLOUD, G. 2009. Vaccine prevention of maternal cytomegalovirus infection. *N Engl J Med,* 360**,** 1191-9.

PATERSON, D. A., DYER, A. P., MILNE, R. S., SEVILLA-REYES, E. & GOMPELS, U. A. 2002. A role for human cytomegalovirus glycoprotein O (gO) in cell fusion and a new hypervariable locus. *Virology,* 293**,** 281-94.

PAWELEC, G. 2012. Hallmarks of human "immunosenescence": adaptation or dysregulation? *Immun Ageing,* 9**,** 15.

PAWELEC, G., DERHOVANESSIAN, E., LARBI, A., STRINDHALL, J. & WIKBY, A. 2009. Cytomegalovirus and human immunosenescence. *Rev Med Virol,* 19**,** 47-56.

PEEK, R., VERBRAAK, F., BRUINENBERG, M., VAN DER LELIJ, A., VAN DEN HORN, G. & KIJLSTRA, A. 1998. Cytomegalovirus glycoprotein B genotyping in ocular fluids and blood of AIDS patients with cytomegalovirus retinitis. *Invest Ophthalmol Vis Sci,* 39**,** 1183-7.

PEREIRA, L., PETITT, M., FONG, A., TSUGE, M., TABATA, T., FANG-HOOVER, J., MAIDJI, E., ZYDEK, M., ZHOU, Y., INOUE, N., LOGHAVI, S., PEPKOWITZ, S., KAUVAR, L. M. & OGUNYEMI, D. 2014. Intrauterine growth restriction caused by underlying congenital cytomegalovirus infection. *J Infect Dis*.

PFEFFER, S., SEWER, A., LAGOS-QUINTANA, M., SHERIDAN, R., SANDER, C., GRASSER, F. A., VAN DYK, L. F., HO, C. K., SHUMAN, S., CHIEN, M., RUSSO, J. J., JU, J., RANDALL, G., LINDENBACH, B. D., RICE, C. M., SIMON, V., HO, D. D., ZAVOLAN, M. & TUSCHL, T. 2005. Identification of microRNAs of the herpesvirus family. *Nat Methods,* 2**,** 269-76.

PHIRI, W., KASONKA, L., COLLIN, S., MAKASA, M., SINKALA, M., CHINTU, C., KASOLO, F., KASEBA, C., TOMKINS, A. M. & FILTEAU, S. M. 2006. Factors influencing breast milk HIV RNA viral load among Zambian women. *AIDS Res Hum Retroviruses,* 22**,** 607-14.

PIGNATELLI, S. & DAL MONTE, P. 2009. Epidemiology of human cytomegalovirus strains through comparison of methodological approaches to explore gN variants. *New Microbiol,* 32**,** 1-10.

PIGNATELLI, S., DAL MONTE, P., ROSSINI, G. & LANDINI, M. P. 2004. Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Rev Med Virol,* 14**,** 383-410.

PIGNATELLI, S., LAZZAROTTO, T., GATTO, M. R., DAL MONTE, P., LANDINI, M. P., FALDELLA, G. & LANARI, M. 2010. Cytomegalovirus gN genotypes distribution among congenitally infected newborns and their relationship with symptoms at birth and sequelae. *Clin Infect Dis,* 51**,** 33-41.

PLACHTER, B., SINZGER, C. & JAHN, G. 1996. Cell types involved in replication and distribution of human cytomegalovirus. *Adv Virus Res,* 46**,** 195-261.

POOLE, E., KUAN, W. L., BARKER, R. & SINCLAIR, J. 2016. The human cytomegalovirus non-coding Beta2.7 RNA as a novel therapeutic for Parkinson's disease - Translational research with no translation. *Virus Res,* 212**,** 64-9.

POOLE, E., MCGREGOR DALLAS, S. R., COLSTON, J., JOSEPH, R. S. & SINCLAIR, J. 2011. Virally induced changes in cellular microRNAs maintain latency of human cytomegalovirus in CD34(+) progenitors. *J Gen Virol,* 92**,** 1539-49.

POOLE, E., WALTHER, A., RAVEN, K., BENEDICT, C. A., MASON, G. M. & SINCLAIR, J. 2013. The myeloid transcription factor GATA-2 regulates the viral UL144 gene during human cytomegalovirus latency in an isolate-specific manner. *J Virol,* 87**,** 4261-71.

POOLE, E., WILLS, M. & SINCLAIR, J. 2014. Human Cytomegalovirus Latency: Targeting Differences in the Latently Infected Cell with a View to Clearing Latent Infection. *New Journal of Science,* 2014**,** 10.

PREIKSAITIS, J. K., BROWN, L. & MCKENZIE, M. 1988. Transfusion-acquired cytomegalovirus infection in neonates. A prospective study. *Transfusion,* 28**,** 205-9.

PRICE, N. B. & PRICHARD, M. N. 2011. Progress in the development of new therapies for herpesvirus infections. *Curr Opin Virol,* 1**,** 548-54.

PRICHARD, M. N., BRITT, W. J., DAILY, S. L., HARTLINE, C. B. & KERN, E. R. 2005. Human cytomegalovirus UL97 Kinase is required for the normal intranuclear distribution of pp65 and virion morphogenesis. *J Virol,* 79**,** 15494-502.

PUCHHAMMER-STOCKL, E. & GORZER, I. 2011. Human cytomegalovirus: an enormous variety of strains and their possible clinical significance in the human host. *Future Virology,* 6**,** 259-271.

PUCHHAMMER-STOCKL, E., GORZER, I., ZOUFALY, A., JAKSCH, P., BAUER, C. C., KLEPETKO, W. & POPOW-KRAUPP, T. 2006. Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation,* 81**,** 187-94.

QUINN, J. J. & CHANG, H. Y. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet,* 17**,** 47-62.

RADFORD, A. D., CHAPMAN, D., DIXON, L., CHANTREY, J., DARBY, A. C. & HALL, N. 2012. Application of next-generation sequencing technologies in virology. *J Gen Virol,* 93**,** 1853-68.

RAHAV, G., GABBAY, R., ORNOY, A., SHECHTMAN, S., ARNON, J. & DIAV-CITRIN, O. 2007. Primary versus nonprimary cytomegalovirus infection during pregnancy, Israel. *Emerg Infect Dis,* 13**,** 1791-3.

RASMUSSEN, L. & COWAN, C. M. 2003. Neutralizing antibody to gB2 human cytomegalovirus does not prevent reactivation in patients with human immunodeficiency virus infection. *J Gen Virol,* 84**,** 1853-7.

RASMUSSEN, L., GEISSLER, A., COWAN, C., CHASE, A. & WINTERS, M. 2002. The genes encoding the gCIII complex of human cytomegalovirus exist in highly diverse combinations in clinical isolates. *J Virol,* 76**,** 10841-8.

RASMUSSEN, L., MULLENAX, J., NELSON, R. & MERIGAN, T. C. 1985. Viral polypeptides detected by a complement-dependent neutralizing murine monoclonal antibody to human cytomegalovirus. *J Virol,* 55**,** 274-80.

RASMUSSEN, L., NELSON, M., NEFF, M. & MERIGAN, T. C. 1988. Characterization of two different human cytomegalovirus glycoproteins which are targets for virus neutralizing antibody. *Virology,* 163**,** 308-18.

RAWLINSON, W. D. & BARRELL, B. G. 1993. Spliced transcripts of human cytomegalovirus. *J Virol,* 67**,** 5502-13.

REAP, E. A., DRYGA, S. A., MORRIS, J., RIVERS, B., NORBERG, P. K., OLMSTED, R. A. & CHULAY, J. D. 2007a. Cellular and humoral immune responses to alphavirus replicon vaccines expressing cytomegalovirus pp65, IE1, and gB proteins. *Clin Vaccine Immunol,* 14**,** 748-55.

REAP, E. A., MORRIS, J., DRYGA, S. A., MAUGHAN, M., TALARICO, T., ESCH, R. E., NEGRI, S., BURNETT, B., GRAHAM, A., OLMSTED, R. A. & CHULAY, J. D. 2007b. Development and preclinical evaluation of an alphavirus replicon particle vaccine for cytomegalovirus. *Vaccine,* 25**,** 7441-9.

REEVES, M. B. 2011. Chromatin-mediated regulation of cytomegalovirus gene expression. *Virus Res,* 157**,** 134-43.

REEVES, M. B., DAVIES, A. A., MCSHARRY, B. P., WILKINSON, G. W. & SINCLAIR, J. H. 2007. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science,* 316**,** 1345-8.

REEVES, M. B., LEHNER, P. J., SISSONS, J. G. & SINCLAIR, J. H. 2005a. An in vitro model for the regulation of human cytomegalovirus latency and reactivation in dendritic cells by chromatin remodelling. *J Gen Virol,* 86**,** 2949-54.

REEVES, M. B., MACARY, P. A., LEHNER, P. J., SISSONS, J. G. & SINCLAIR, J. H. 2005b. Latency, chromatin remodeling, and reactivation of human cytomegalovirus in the dendritic cells of healthy carriers. *Proc Natl Acad Sci U S A,* 102**,** 4140-5.

REEVES, M. B., MURPHY, J., GREAVES, R., FAIRLEY, J., BREHM, A. & SINCLAIR, J. 2006. Autorepression of the human cytomegalovirus major immediate-early

promoter/enhancer at late times of infection is mediated by the recruitment of chromatin remodeling enzymes by IE86. *J Virol,* 80**,** 9998-10009.

REEVES, M. B. & SINCLAIR, J. H. 2008. Aspects of Human Cytomegalovirus Latency and Reactivation. *In:* SHENK, T. E. & STINSKI, M. F. (eds.) *Human Cytomegalovirus.* Berlin, Heidelberg: Springer Berlin Heidelberg.

REEVES, M. B. & SINCLAIR, J. H. 2010. Analysis of latent viral gene expression in natural and experimental latency models of human cytomegalovirus and its correlation with histone modifications at a latent promoter. *J Gen Virol,* 91**,** 599-604.

REEVES, M. B., WOODHALL, D., COMPTON, T. & SINCLAIR, J. 2010. Human cytomegalovirus IE72 protein interacts with the transcriptional repressor hDaxx to regulate LUNA gene expression during lytic infection. *J Virol,* 84**,** 7185-94.

RENZETTE, N., BHATTACHARJEE, B., JENSEN, J. D., GIBSON, L. & KOWALIK, T. F. 2011. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog,* 7**,** e1001344.

RENZETTE, N., GIBSON, L., BHATTACHARJEE, B., FISHER, D., SCHLEISS, M. R., JENSEN, J. D. & KOWALIK, T. F. 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. *PLoS Genet,* 9**,** e1003735.

REVELLO, M. G. & GERNA, G. 2002. Diagnosis and management of human cytomegalovirus infection in the mother, fetus, and newborn infant. *Clin Microbiol Rev,* 15**,** 680-715.

REVELLO, M. G. & GERNA, G. 2004. Pathogenesis and prenatal diagnosis of human cytomegalovirus infection. *J Clin Virol,* 29**,** 71-83.

RIEDER, F. & STEININGER, C. 2014. Cytomegalovirus vaccine: phase II clinical trial results. *Clin Microbiol Infect,* 20 Suppl 5**,** 95-102.

RILEY, H. D., JR. 1997. History of the cytomegalovirus. *South Med J,* 90**,** 184-90.

ROBERTS, E. T., HAAN, M. N., DOWD, J. B. & AIELLO, A. E. 2010. Cytomegalovirus Antibody Levels, Inflammation, and Mortality Among Elderly Latinos Over 9 Years of Follow-up. *American Journal of Epidemiology,* 172**,** 363-371.

RODRIGUEZ, J., CASPER, K., SMALLWOOD, G., STIEBER, A., FASOLA, C., LEHNEMAN, J. & HEFFRON, T. 2007. Resistance to combined ganciclovir and foscarnet therapy in a liver transplant recipient with possible dual-strain cytomegalovirus coinfection. *Liver Transpl,* 13**,** 1396-400.

ROIZMAN, B. & PELLETT, P. E. 2001. The family Herpesviridae: a brief introduction. *In:* KNIPE, D. M. & HOWLEY, P. M. (eds.) *Fields Virology.* 4th ed. Philadelphia: Lippincott, Williams and Wilkins.

ROIZMAN, B. & PELLETT, P. E. 2007. The family Herpesviridae: a brief introduction. *In:* KNIPE, D. M. & HOWLEY, P. M. (eds.) *Fields Virology.* 5th ed. Philadelphia: Lippincott, Williams and Wilkins.

ROSENTHAL, L. S., FOWLER, K. B., BOPPANA, S. B., BRITT, W. J., PASS, R. F., SCHMID, S. D., STAGNO, S. & CANNON, M. J. 2009. Cytomegalovirus shedding and delayed sensorineural hearing loss: results from longitudinal follow-up of children with congenital infection. *Pediatr Infect Dis J,* 28**,** 515-20.

ROSS, S. A. & BOPPANA, S. B. 2005. Congenital cytomegalovirus infection: outcome and diagnosis. *Semin Pediatr Infect Dis,* 16**,** 44-9.

ROSS, S. A., NOVAK, Z., PATI, S. & BOPPANA, S. B. 2011. Diagnosis of Cytomegalovirus Infections. *Infectious disorders drug targets,* 11**,** 466-474.

ROSSETTO, C. C., TARRANT-ELORZA, M. & PARI, G. S. 2013. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog,* 9**,** e1003366.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-5.

RYCKMAN, B. J., CHASE, M. C. & JOHNSON, D. C. 2008a. 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**,** 14118-23.

RYCKMAN, B. J., JARVIS, M. A., DRUMMOND, D. D., NELSON, J. A. & JOHNSON, D. C. 2006. Human cytomegalovirus entry into epithelial and endothelial cells depends on genes UL128 to UL150 and occurs by endocytosis and low-pH fusion. *J Virol,* 80**,** 710-22.

RYCKMAN, B. J., RAINISH, B. L., CHASE, M. C., BORTON, J. A., NELSON, J. A., JARVIS, M. A. & JOHNSON, D. C. 2008b. Characterization of the human cytomegalovirus gH/gL/UL128-131 complex that mediates entry into epithelial and endothelial cells. *J Virol,* 82**,** 60-70.

SACCOCCIO, F. M., SAUER, A. L., CUI, X., ARMSTRONG, A. E., HABIB EL, S. E., JOHNSON, D. C., RYCKMAN, B. J., KLINGELHUTZ, A. J., ADLER, S. P. & MCVOY, M. A. 2011. Peptides from cytomegalovirus UL130 and UL131 proteins induce high titer antibodies that block viral entry into mucosal epithelial cells. *Vaccine,* 29**,** 2705-11.

SAFFERT, R. T. & KALEJTA, R. F. 2006. Inactivating a cellular intrinsic immune defense mediated by Daxx is the mechanism through which the human cytomegalovirus pp71 protein stimulates viral immediate-early gene expression. *J Virol,* 80**,** 3863-71.

SAFFERT, R. T. & KALEJTA, R. F. 2007. Human cytomegalovirus gene expression is silenced by Daxx-mediated intrinsic immune defense in model latent infections established in vitro. *J Virol,* 81**,** 9109-20.

SAFFERT, R. T., PENKERT, R. R. & KALEJTA, R. F. 2010. Cellular and viral control over the initial events of human cytomegalovirus experimental latency in CD34+ cells. *J Virol,* 84**,** 5594-604.

SANGER, F. & COULSON, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol,* 94**,** 441-8.

SANZ-RAMOS, M., MANNO, D., KAPAMBWE, M., NDUMBA, I., MUSONDA, K. G., BATES, M., CHIBUMBYA, J., SIAME, J., MONZE, M., FILTEAU, S., GOMPELS, U. A. & TEAM, C. S. 2013. Reduced Poliovirus vaccine neutralising-antibody titres in infants with maternal HIV-exposure. *Vaccine,* 31**,** 2042-9.

SARCINELLA, L., MAZZULLI, T., WILLEY, B. & HUMAR, A. 2002. Cytomegalovirus glycoprotein B genotype does not correlate with outcomes in liver transplant patients. *J Clin Virol,* 24**,** 99-105.

SCHLEISS, M. R. 2006a. Acquisition of human cytomegalovirus infection in infants via breast milk: natural immunization or cause for concern? *Rev Med Virol,* 16**,** 73-82.

SCHLEISS, M. R. 2006b. Role of breast milk in acquisition of cytomegalovirus infection: recent advances. *Curr Opin Pediatr,* 18**,** 48-52.

SCHLEISS, M. R. 2013. Developing a Vaccine against Congenital Cytomegalovirus (CMV) Infection: What Have We Learned from Animal Models? Where Should We Go Next? *Future Virol,* 8**,** 1161-1182.

SCHLEISS, M. R., CHOI, K. Y., ANDERSON, J., MASH, J. G., WETTENDORFF, M., MOSSMAN, S. & VAN DAMME, M. 2014. Glycoprotein B (gB) vaccines adjuvanted with AS01 or AS02 protect female guinea pigs against cytomegalovirus (CMV) viremia and offspring mortality in a CMV-challenge model. *Vaccine,* 32**,** 2756-2762.

SCHLEISS, M. R. & HEINEMAN, T. C. 2005. Progress toward an elusive goal: current status of cytomegalovirus vaccines. *Expert Rev Vaccines,* 4**,** 381-406.

SCHNEPF, N., CORVO, J., PORS, M. J. & MAZERON, M. C. 2011. Antiviral activity of ganciclovir and artesunate towards human cytomegalovirus in astrocytoma cells. *Antiviral Res,* 89**,** 186-8.

SCRIVANO, L., SINZGER, C., NITSCHKO, H., KOSZINOWSKI, U. H. & ADLER, B. 2011. HCMV Spread and Cell Tropism are Determined by Distinct Virus Populations. *PLOS Pathogens,* 7**,** e1001256.

SEKULIN, K., GÖRZER, I., HEISS-CZEDIK, D. & PUCHHAMMER-STÖCKL, E. 2007. Analysis of the variability of CMV strains in the RL11D domain of the RL11 multigene family. *Virus Genes,* 35**,** 577-583.

SHAPIRA, M. Y., RESNICK, I. B., CHOU, S., NEUMANN, A. U., LURAIN, N. S., STAMMINGER, T., CAPLAN, O., SALEH, N., EFFERTH, T., MARSCHALL, M. & WOLF, D. G. 2008. Artesunate as a potent antiviral agent in a patient with late drug-resistant cytomegalovirus infection after hematopoietic stem cell transplantation. *Clin Infect Dis,* 46**,** 1455-7.

SHEN, Z. Z., PAN, X., MIAO, L. F., YE, H. Q., CHAVANAS, S., DAVRINCHE, C., MCVOY, M. & LUO, M. H. 2014. Comprehensive analysis of human cytomegalovirus microRNA expression during lytic and quiescent infection. *PLoS One,* 9**,** e88531.

SINCLAIR, J. 2008. Human cytomegalovirus: Latency and reactivation in the myeloid lineage. *J Clin Virol,* 41**,** 180-5.

SINCLAIR, J. & SISSONS, P. 2006. Latency and reactivation of human cytomegalovirus. *J Gen Virol,* 87**,** 1763-79.

SLYKER, J. A., FARQUHAR, C., ATKINSON, C., ASBJORNSDOTTIR, K., ROXBY, A., DRAKE, A., KIARIE, J., WALD, A., BOECKH, M., RICHARDSON, B., ODEM-DAVIS, K., JOHN-STEWART, G. & EMERY, V. 2014. Compartmentalized cytomegalovirus replication and transmission in the setting of maternal HIV-1 infection. *Clin Infect Dis,* 58**,** 564-72.

SLYKER, J. A., LOHMAN-PAYNE, B. L., JOHN-STEWART, G. C., MALECHE-OBIMBO, E., EMERY, S., RICHARDSON, B., DONG, T., IVERSEN, A. K., MBORI-NGACHA, D., OVERBAUGH, J., EMERY, V. C. & ROWLAND-JONES, S. L. 2009a. Acute cytomegalovirus infection in Kenyan HIV-infected infants. *AIDS,* 23**,** 2173-81.

SLYKER, J. A., LOHMAN-PAYNE, B. L., ROWLAND-JONES, S. L., OTIENO, P., MALECHE-OBIMBO, E., RICHARDSON, B., FARQUHAR, C., MBORI-NGACHA, D., EMERY, V. C. & JOHN-STEWART, G. C. 2009b. The detection of cytomegalovirus DNA in maternal plasma is associated with mortality in HIV-1-infected women and their infants. *AIDS,* 23**,** 117-24.

SÖDERBERG-NAUCLÉR, C. 2008. HCMV microinfections in inflammatory diseases and cancer. *Journal of Clinical Virology,* 41**,** 218-223.

SOHN, Y. M., OH, M. K., BALCAREK, K. B., CLOUD, G. A. & PASS, R. F. 1991. Cytomegalovirus infection in sexually active adolescents. *J Infect Dis,* 163**,** 460-3.

SOROCEANU, L., AKHAVAN, A. & COBBS, C. S. 2008. Platelet-derived growth factor-alpha receptor activation is required for human cytomegalovirus infection. *Nature,* 455**,** 391-5.

SOROCEANU, L. & COBBS, C. S. 2011. Is HCMV a tumor promoter? *Virus Research,* 157**,** 193-203.

SOWMYA, P. & MADHAVAN, H. N. 2009. Analysis of mixed infections by multiple genotypes of human cytomegalovirus in immunocompromised patients. *J Med Virol,* 81**,** 861-9.

SPANO, L. C., FERREIRA, M. S. R., ALMEIDA, M. S., DO NASCIMENTO, J. P. & LEITE, J. P. G. 2007. HCMV gB genotypes in cervical secretion and placenta tissues in the state of Espirito Santo, Southeastearn Brazil. *Brazilian Journal of Microbiology,* 38**,** 424-429.

STAGNO, S., REYNOLDS, D. W., PASS, R. F. & ALFORD, C. A. 1980. Breast milk and the risk of cytomegalovirus infection. *N Engl J Med,* 302**,** 1073-6.

STANTON, R. J., BALUCHOVA, K., DARGAN, D. J., CUNNINGHAM, C., SHEEHY, O., SEIRAFIAN, S., MCSHARRY, B. P., NEALE, M. L., DAVIES, J. A., TOMASEC, P., DAVISON, A. J. & WILKINSON, G. W. 2010. Reconstruction of the complete human

cytomegalovirus genome in a BAC reveals RL13 to be a potent inhibitor of replication. *J Clin Invest,* 120**,** 3191-208.

STANTON, R. J., WESTMORELAND, D., FOX, J. D., DAVISON, A. J. & WILKINSON, G. W. 2005. Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. *J Med Virol,* 75**,** 42-6.

STARAS, S. A., FLANDERS, W. D., DOLLARD, S. C., PASS, R. F., MCGOWAN, J. E. & CANNON, M. J. 2008a. Cytomegalovirus seroprevalence and childhood sources of infection: A population-based study among pre-adolescents in the United States. *J Clin Virol,* 43**,** 266-71.

STARAS, S. A., FLANDERS, W. D., DOLLARD, S. C., PASS, R. F., MCGOWAN, J. E. & CANNON, M. J. 2008b. Influence of sexual activity on cytomegalovirus seroprevalence in the United States, 1988-1994. *Sex Transm Dis,* 35**,** 472-9.

STARK, T. J., ARNOLD, J. D., SPECTOR, D. H. & YEO, G. W. 2012. High-resolution profiling and analysis of viral and host small RNAs during human cytomegalovirus infection. *J Virol,* 86**,** 226-35.

STASSEN, F. R., VEGA-CORDOVA, X., VLIEGEN, I. & BRUGGEMAN, C. A. 2006. Immune activation following cytomegalovirus infection: More important than direct viral effects in cardiovascular disease? *Journal of Clinical Virology,* 35**,** 349-353.

STERN-GINOSSAR, N., ELEFANT, N., ZIMMERMANN, A., WOLF, D. G., SALEH, N., BITON, M., HORWITZ, E., PROKOCIMER, Z., PRICHARD, M., HAHN, G., GOLDMAN-WOHL, D., GREENFIELD, C., YAGEL, S., HENGEL, H., ALTUVIA, Y., MARGALIT, H. & MANDELBOIM, O. 2007. Host immune system gene targeting by a viral miRNA. *Science,* 317**,** 376-81.

STERN-GINOSSAR, N., SALEH, N., GOLDBERG, M. D., PRICHARD, M., WOLF, D. G. & MANDELBOIM, O. 2009. Analysis of human cytomegalovirus-encoded microRNA activity during infection. *J Virol,* 83**,** 10684-93.

STERN-GINOSSAR, N., WEISBURD, B., MICHALSKI, A., LE, V. T., HEIN, M. Y., HUANG, S. X., MA, M., SHEN, B., QIAN, S. B., HENGEL, H., MANN, M., INGOLIA, N. T. & WEISSMAN, J. S. 2012. Decoding human cytomegalovirus. *Science,* 338**,** 1088-93.

STOCK, K., GRIESMAIER, E., BRUNNER, B., NEUBAUER, V., KIECHL-KOHLENDORFER, U. & TRAWOGER, R. 2015. Pasteurization of breastmilk decreases the rate of postnatally acquired cytomegalovirus infections, but shows a nonsignificant trend to an increased rate of necrotizing enterocolitis in very preterm infants--a preliminary study. *Breastfeed Med,* 10**,** 113-7.

STOWELL, J. D., FORLIN-PASSONI, D., RADFORD, K., BATE, S. L., DOLLARD, S. C., BIALEK, S. R., CANNON, M. J. & SCHMID, D. S. 2014. Cytomegalovirus survival and transferability and the effectiveness of common hand-washing agents against cytomegalovirus on live human hands. *Appl Environ Microbiol,* 80**,** 455-61.

STRATTON, D. J., DURCH, J. S. & LAWRENCE, R. S. 2000. *Vaccines for the 21st Century: A Tool for Decision Making,* Washington DC, USA, National Academies Press.

STRINDHALL, J., NILSSON, B. O., LOFGREN, S., ERNERUDH, J., PAWELEC, G., JOHANSSON, B. & WIKBY, A. 2007. No Immune Risk Profile among individuals who reach 100 years of age: Findings from the Swedish NONA immune longitudinal study. *Experimental Gerontology,* 42**,** 753-761.

SUNG, H. & SCHLEISS, M. R. 2010. Update on the current status of cytomegalovirus vaccines. *Expert Rev Vaccines,* 9**,** 1303-14.

SWAIN, M. T., TSAI, I. J., ASSEFA, S. A., NEWBOLD, C., BERRIMAN, M. & OTTO, T. D. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc,* 7**,** 1260-84.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol,* 30**,** 2725-9.

TAN, B. H. 2014. Cytomegalovirus Treatment. *Current Treatment Options in Infectious Diseases,* 6**,** 256-270.

TORRES-MADRIZ, G. & BOUCHER, H. W. 2008. Immunocompromised hosts: perspectives in the treatment and prophylaxis of cytomegalovirus disease in solid-organ transplant recipients. *Clin Infect Dis,* 47**,** 702-11.

TWEEDY, J., SPYROU, M. A., DONALDSON, C. D., DEPLEDGE, D., BREUER, J. & GOMPELS, U. A. 2015a. Complete Genome Sequence of the Human Herpesvirus 6A Strain AJ from Africa Resembles Strain GS from North America. *Genome Announc,* 3.

TWEEDY, J., SPYROU, M. A., HUBACEK, P., KUHL, U., LASSNER, D. & GOMPELS, U. A. 2015b. Analyses of germline, chromosomally integrated human herpesvirus 6A and B genomes indicate emergent infection and new inflammatory mediators. *J Gen Virol,* 96**,** 370-89.

TWEEDY, J., SPYROU, M. A., PEARSON, M., LASSNER, D., KUHL, U. & GOMPELS, U. A. 2016. Complete Genome Sequence of Germline Chromosomally Integrated Human Herpesvirus 6A and Analyses Integration Sites Define a New Human Endogenous Virus with Potential to Reactivate as an Emerging Infection. *Viruses,* 8.

VAN DER SANDE, M. A., KAYE, S., MILES, D. J., WAIGHT, P., JEFFRIES, D. J., OJUOLA, O. O., PALMERO, M., PINDER, M., ISMAILI, J., FLANAGAN, K. L., AVEIKA, A. A., ZAMAN, A., ROWLAND-JONES, S., MCCONKEY, S. J., WHITTLE, H. C. & MARCHANT, A. 2007. Risk factors for and clinical outcome of congenital cytomegalovirus infection in a peri-urban West-African birth cohort. *PLoS One,* 2**,** e492.

VAN DER STRATE, B. W., HARMSEN, M. C., SCHAFER, P., SWART, P. J., THE, T. H., JAHN, G., SPEER, C. P., MEIJER, D. K. & HAMPRECHT, K. 2001. Viral load in breast milk correlates with transmission of human cytomegalovirus to preterm neonates, but lactoferrin concentrations do not. *Clin Diagn Lab Immunol,* 8**,** 818-21.

VAN ZEIJL, M., FAIRHURST, J., BAUM, E. Z., SUN, L. & JONES, T. R. 1997. The human cytomegalovirus UL97 protein is phosphorylated and a component of virions. *Virology,* 231**,** 72-80.

VANARSDALL, A. L., RYCKMAN, B. J., CHASE, M. C. & JOHNSON, D. C. 2008. Human cytomegalovirus glycoproteins gB and gH/gL mediate epithelial cell-cell fusion when expressed either in cis or in trans. *J Virol,* 82**,** 11837-50.

VANARSDALL, A. L., WISNER, T. W., LEI, H., KAZLAUSKAS, A. & JOHNSON, D. C. 2012. PDGF receptor-alpha does not promote HCMV entry into epithelial and endothelial cells but increased quantities stimulate entry by an abnormal pathway. *PLoS Pathog,* 8**,** e1002905.

VARGHESE, V., SHAHRIAR, R., RHEE, S. Y., LIU, T., SIMEN, B. B., EGHOLM, M., HANCZARUK, B., BLAKE, L. A., GHARIZADEH, B., BABRZADEH, F., BACHMANN, M. H., FESSEL, W. J. & SHAFER, R. W. 2009. Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. *J Acquir Immune Defic Syndr,* 52**,** 309-15.

VERGHESE, P. S. & SCHLEISS, M. R. 2013. Letermovir Treatment of Human Cytomegalovirus Infection Antiinfective Agent. *Drugs Future,* 38**,** 291-298.

VISENTIN, S., MANARA, R., MILANESE, L., DA ROIT, A., FORNER, G., SALVIATO, E., CITTON, V., MAGNO, F. M., ORZAN, E., MORANDO, C., CUSINATO, R., MENGOLI, C., PALU, G., ERMANI, M., RINALDI, R., COSMI, E. & GUSSETTI, N. 2012. Early primary cytomegalovirus infection in pregnancy: maternal hyperimmunoglobulin therapy improves outcomes among infants at 1 year of age. *Clin Infect Dis,* 55**,** 497-503.

VYSE, A. J., HESKETH, L. M. & PEBODY, R. G. 2009. The burden of infection with cytomegalovirus in England and Wales: how many women are infected in pregnancy? *Epidemiol Infect,* 137**,** 526-33.

WANG, D. & FU, T. M. 2014. Progress on human cytomegalovirus vaccines for prevention of congenital infection and disease. *Curr Opin Virol,* 6**,** 13-23.

WANG, D. & SHENK, T. 2005a. Human cytomegalovirus UL131 open reading frame is required for epithelial cell tropism. *J Virol,* 79**,** 10330-8.

WANG, D. & SHENK, T. 2005b. Human cytomegalovirus virion protein complex required for epithelial and endothelial cell tropism. *Proc Natl Acad Sci U S A,* 102**,** 18153-8.

WANG, W., TAYLOR, S. L., LEISENFELDER, S. A., MORTON, R., MOFFAT, J. F., SMIRNOV, S. & ZHU, H. 2005. Human cytomegalovirus genes in the 15-kilobase region are required for viral replication in implanted human tissues in SCID mice. *J Virol,* 79**,** 2115-23.

WANG, X., HUONG, S. M., CHIU, M. L., RAAB-TRAUB, N. & HUANG, E. S. 2003. Epidermal growth factor receptor is a cellular receptor for human cytomegalovirus. *Nature,* 424**,** 456-61.

WEBSTER, A., LEE, C. A., COOK, D. G., GRUNDY, J. E., EMERY, V. C., KERNOFF, P. B. & GRIFFITHS, P. D. 1989. Cytomegalovirus infection and progression towards AIDS in haemophiliacs with human immunodeficiency virus infection. *Lancet,* 2**,** 63-6.

WEINBERG, A., JABS, D. A., CHOU, S., MARTIN, B. K., LURAIN, N. S., FORMAN, M. S., CRUMPACKER, C., CYTOMEGALOVIRUS, R., VIRAL RESISTANCE STUDY, G. & ADULT, A. C. T. G. C. L. 2003. Mutations conferring foscarnet resistance in a cohort of patients with acquired immunodeficiency syndrome and cytomegalovirus retinitis. *J Infect Dis,* 187**,** 777-84.

WIKBY, A., JOHANSSON, B., OLSSON, J., LOFGREN, S., NILSSON, B. O. & FERGUSON, F. 2002. Expansions of peripheral blood CD8 T-lymphocyte subpopulations and an association with cytomegalovirus seropositivity in the elderly: the Swedish NONA immune study. *Experimental Gerontology,* 37**,** 445-453.

WILHELM, J. A., MATTER, L. & SCHOPFER, K. 1986. The risk of transmitting cytomegalovirus to patients receiving blood transfusions. *J Infect Dis,* 154**,** 169-71.

WILLE, P. T., WISNER, T. W., RYCKMAN, B. & JOHNSON, D. C. 2013. Human cytomegalovirus (HCMV) glycoprotein gB promotes virus entry in trans acting as the viral fusion protein rather than as a receptor-binding protein. *MBio,* 4**,** e00332-13.

WILLIAMS-AZIZ, S. L., HARTLINE, C. B., HARDEN, E. A., DAILY, S. L., PRICHARD, M. N., KUSHNER, N. L., BEADLE, J. R., WAN, W. B., HOSTETLER, K. Y. & KERN, E. R. 2005. Comparative activities of lipid esters of cidofovir and cyclic cidofovir against replication of herpesviruses in vitro. *Antimicrob Agents Chemother,* 49**,** 3724-33.

WLOCH, M. K., SMITH, L. R., BOUTSABOUALOY, S., REYES, L., HAN, C., KEHLER, J., SMITH, H. D., SELK, L., NAKAMURA, R., BROWN, J. M., MARBURY, T., WALD, A., ROLLAND, A., KASLOW, D., EVANS, T. & BOECKH, M. 2008. Safety and immunogenicity of a bivalent cytomegalovirus DNA vaccine in healthy adult subjects. *J Infect Dis,* 197**,** 1634-42.

WOLF, D. G., SHIMONI, A., RESNICK, I. B., STAMMINGER, T., NEUMANN, A. U., CHOU, S., EFFERTH, T., CAPLAN, O., ROSE, J., NAGLER, A. & MARSCHALL, M. 2011. Human cytomegalovirus kinetics following institution of artesunate after hematopoietic stem cell transplantation. *Antiviral Res,* 90**,** 183-6.

WOODHALL, D. L., GROVES, I. J., REEVES, M. B., WILKINSON, G. & SINCLAIR, J. H. 2006. Human Daxx-mediated repression of human cytomegalovirus gene expression correlates with a repressive chromatin structure around the major immediate early promoter. *J Biol Chem,* 281**,** 37652-60.

WRIGHT, E., BAIN, M., TEAGUE, L., MURPHY, J. & SINCLAIR, J. 2005. Ets-2 repressor factor recruits histone deacetylase to silence human cytomegalovirus immediate-early gene expression in non-permissive cells. *Journal of General Virology,* 86**,** 535-544.

WUSSOW, F., CHIUPPESI, F., MARTINEZ, J., CAMPO, J., JOHNSON, E., FLECHSIG, C., NEWELL, M., TRAN, E., ORTIZ, J., LA ROSA, C., HERRMANN, A., LONGMATE, J., CHAKRABORTY, R., BARRY, P. A. & DIAMOND, D. J. 2014. Human cytomegalovirus vaccine based on the envelope gH/gL pentamer complex. *PLoS Pathog,* 10**,** e1004524.

YAMAMOTO, A. Y., CASTELLUCCI, R. A., ARAGON, D. C. & MUSSI-PINHATA, M. M. 2013. Early high CMV seroprevalence in pregnant women from a population with a high rate of congenital infection. *Epidemiol Infect,* 141**,** 2187-91.

YAMAMOTO, A. Y., MUSSI-PINHATA, M. M., BOPPANA, S. B., NOVAK, Z., WAGATSUMA, V. M., OLIVEIRA PDE, F., DUARTE, G. & BRITT, W. J. 2010. Human cytomegalovirus reinfection is associated with intrauterine transmission in a highly cytomegalovirus-immune maternal population. *Am J Obstet Gynecol,* 202**,** 297 e1-8.

YASUDA, A., KIMURA, H., HAYAKAWA, M., OHSHIRO, M., KATO, Y., MATSUURA, O., SUZUKI, C. & MORISHIMA, T. 2003. Evaluation of cytomegalovirus infections transmitted via breast milk in preterm infants with a real-time polymerase chain reaction assay. *Pediatrics,* 111**,** 1333-6.

ZENG, A. H., OU, Y. Y., GUO, M. M., DAI, X., ZHOU, D. Z. & CHEN, R. 2015. Human embryonic lung fibroblasts treated with artesunate exhibit reduced rates of proliferation and human cytomegalovirus infection in vitro. *J Thorac Dis,* 7**,** 1151-7.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res,* 18**,** 821-9.

ZHANG, G., RAGHAVAN, B., KOTUR, M., CHEATHAM, J., SEDMAK, D., COOK, C., WALDMAN, J. & TRGOVCICH, J. 2007. Antisense transcription in the human cytomegalovirus transcriptome. *J Virol,* 81**,** 11267-81.

ZHONG, J. & KHANNA, R. 2007. Vaccine strategies against human cytomegalovirus infection. *Expert Rev Anti Infect Ther,* 5**,** 449-59.

APPENDIX 1: ETHICS APPROVAL LETTER – UNZABREC.

THE UNIVERSITY OF ZAMBIA
BIOMEDICAL RESEARCH ETHICS COMMITTEE

Telephone: 260-1-256067
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: 260-1-250753
E-mail: unzarec@unza.zm
Assurance No. FWA00000338
IRB00001131 of IORG0000774

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia

23rd August, 2013

Your Ref: 010-06-13

Dr. Kunda Musonda
University Teaching Hospital
Department of Pathology and Microbiology
P/Bag RW 1X
Lusaka

Dear Dr Musonda,

RE: RE-SUBMITTED RESEARCH PROPOSAL: **"MOLECULAR CHARACTERIZATION, TRANSMISSION ROUTES AND DEVELOPMENTAL EFFECTS OF INFANT INFECTIONS WITH HUMAN BATAHERPESVIRUSES (HCMV, HHV-6A AND HHV-6B) IN ZAMBIA: A PILOT STUDY" (REF. NO.:010-06-13)**

The above mentioned research proposal was re-submitted to the Biomedical Research Ethics Committee for ethical review was approved after discussion.

CONDITIONS:

- This approval is based strictly on your submitted proposal. Should there be need for you to modify or change the study design or methodology, you will need to seek clearance from the Research Ethics Committee.
- If you have need for further clarification please consult this office. Please note that it is mandatory that you submit a detailed progress report of your study to this Committee every six months and a final copy of your report at the end of the study.
- This waiver does not release you from the obligation of ensuring confidentiality.
- Please note that when your approval expires you may need to request for renewal. The request should be accompanied by a Progress Report (Progress Report Forms can be obtained from the Secretariat).
- **Ensure that a final copy of the results is submitted to this Committee.**

Yours sincerely,

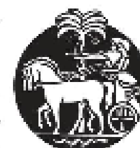
Dr. J.C Munthali
CHAIRPERSON

Date of approval: 23 August, 2013 **Date of expiry:** 22 August, 2014

APPENDIX 2: ETHICS APPROVAL LETTER – LSHTM REC.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE 

Observational / Interventions Research Ethics Committee

Ursula Gompels
Reader in Molecular Virology
PMB / ITD
LSHTM

31 July 2013

Dear Dr. Gompels,

Study Title: **Molecular characterisation, transmission routes, and developmental effects of infant infections with Human betaherpesviruses (HCMV, HHV-6A and HHV-6B) in Zambia: a pilot study.**
LSHTM ethics ref: **6456**

Thank you for your letter of 26 July 2013, responding to the Observational Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

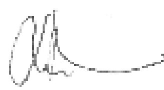
The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
LSHTM ethics application	n/a	3/06/2013
Protocol including Information Sheets & Consent forms	1	29.05.2013

After ethical review

Any subsequent changes to the application must be submitted to the Committee via an E2 amendment form. All studies are also required to notify the ethics committee of any serious adverse events which occur during the project via form E4. At the end of the study, please notify the committee via form E5.

Yours sincerely,



Professor John DH Porter
Chair
ethics@lshtm.ac.uk
http://intra.lshtm.ac.uk/management/committees/ethics/

Improving health worldwide

Page 1 of 1

APPENDIX 3: UTH RESEARCH APPROVAL LETTER.

REPUBLIC OF ZAMBIA

MINISTRY OF HEALTH

University Teaching Hospital

P/Bag Rw 1X
Lusaka - Zambia

Fax: +260 211 250305
e-mail: mduth@yahoo.com

Tel: +260 211 253947 (Switch Board)
+260 211 251451

OFFICE OF THE SENIOR MEDICAL SUPERINTENDENT

Our Ref: UTH/SMS/09/5
Your Ref:

8th May 2013

Dr. Kunda G. Musonda
London School of Hygiene and Tropical Medicine
Department of Pathogen Molecular Biology
LSHTM' University of London
Keppel Street, London
WC1E, UNITED KINGDOM

Dear Dr. Musonda,

PERMISSION TO CONDUCT RESEARCH.

The University Teaching Hospital (UTH) is following your academic progress with keen interest.

We have reviewed your proposed study for your PHD thesis entitled " **Pilot study of Molecular characterization and routes of transmission of Human Cytomegalovirus (HCMV), Human Herpesvirus 6A(HIV-6A) and Human Herpesvirus 6B (HIV-6B) Infant Infection in Zambia".** This topic seems appropriate to our Zambian setting and needs..

University Teaching Hospital therefore supports your research efforts and gives approval for you to conduct this research in the University Teaching Hospital.

Yours sincerely

Dr. L. Kasonka
Managing Director and
Senior Medical Superintendent
UNIVERSITY TEACHING HOSPITAL

c.c. University of Zambia,
Research Ethics Committee

APPENDIX 4: BFPH SAMPLE RESUSE APPROVAL – UNZABREC.

THE UNIVERSITY OF ZAMBIA

RESEARCH ETHICS COMMITTEE

Telephone: 260-1-256067
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: + 260-1-250753
E-mail: unzarec@zamtel.zm

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia

Assurance No. FWA00000338
IRB00001131 of IORG0000774

25 April, 2006
Ref.: 002-05-04

Dr Lackson Kasonka
Breast-Feeding & Postpartum Health Project (BFPHP)
ICH, LUDMHB, UTH
Department of Obstetric & Gynaecology
University Teaching Hospital
P/Bag RW1X
LUSAKA

Dear Dr Kasonka,

RE: **BREASTFEEDING AND POSTPARTUM HEALTH PROJECT**

We acknowledge receipt of your letter dated 4 April, 2006 concerning the above-mentioned study in which you were requesting permission to use the redundant samples for new but related biochemical assays.

We note that the study was successfully completed.

The use of the remaining plasma and breast milk samples for new but related biochemical assays is approved.

With best wishes.

Yours sincerely,

E. M. Nkandu
SECRETARY

APPENDIX 5: PARTICIPANT INFORMATION SHEET – NEONATAL STUDY

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT United Kingdom.

PARTICIPANT INFORMATION SHEET

8th May 2013, version 1.0 Study protocol number: **UNZABREC 010-06-13** / **LSHTM REC 6456**

Study Title: **Molecular characterisation, transmission routes, and developmental effects of infant infections with Human betaherpesviruses (HCMV, HHV-6A and HHV-6B) in Zambia: a pilot study.**

Principal investigator: Dr Ursula Gompels, London School of Hygiene & Tropical Medicine

1. Invitation
You and your child are invited to take part in this study. The following information describes the study to help you decide whether you would like to take part. If you need more information, a member of the research time will be happy to explain further.

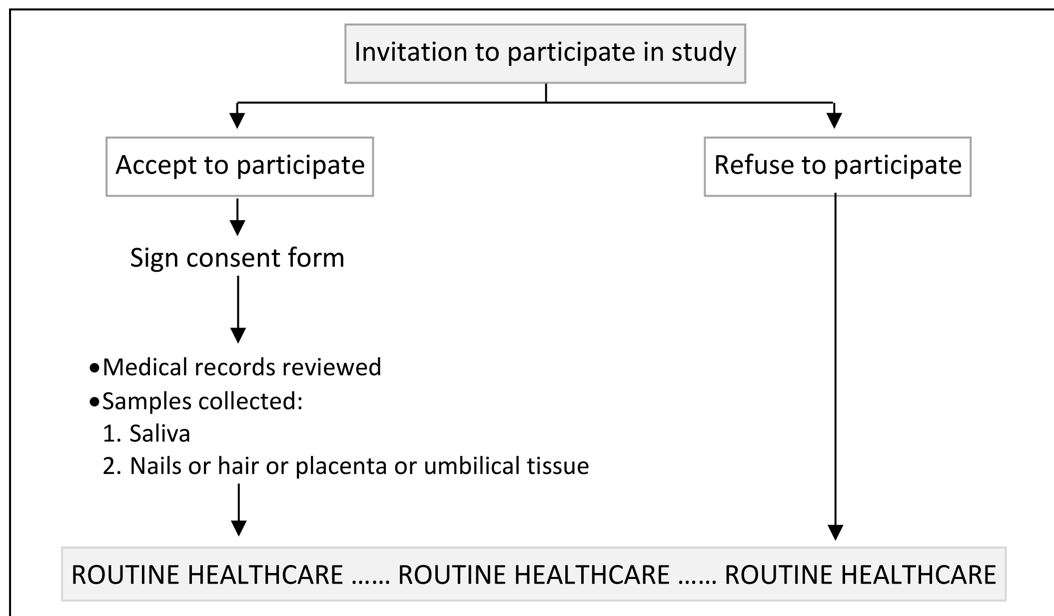
2. What is the purpose of the study?
The purpose of this research study is to obtain information on three related viruses that can be contracted in early childhood. These viruses are Human Cytomegalovirus (HCMV), Human Herpesvirus 6A (HHV-6A) and Human Herpesvirus 6B (HHV-6B). These viruses are part of a small group that can infect your developing baby, called 'congenital' infection. They can cause disease then or later in childhood. This is because they infect a person for life; they remain in a dormant state within the body, but can periodically 'awaken' to cause disease symptoms. During these periods of awakening, they are shed through bodily fluids such as saliva, urine, breastmilk, genital fluids, sweat, and tears. Early infant infection with HCMV can lead to a wide range of conditions including deafness, blindness, mental retardation and developmental difficulties. HHV-6A and HHV-6B are associated with a childhood fever illness and rare severe infections with brain disorders including encephalitis, certain types of epilepsy, and multiple sclerosis. Our immune system protects us from infection. These viruses cause disease if our immune system is not yet developed as in babies, or not functioning well, as in individuals with HIV/AIDS. These viruses can be passed on from a pregnant woman to her child during pregnancy and around the time of delivery. HHV-6A and HHV-6B can be passed to the baby in another very unique way; in a few cases, they can combine with our genes (the way we inherit our characteristics from our parents). Such integrated virus can be detected in hair and nails, making it possible to find this form of infection without requiring a blood sample. Although these viruses can cause serious disease in babies and infants, they are not diagnosed here, so it is not possible yet to stop their infection. In this study, if you take part you will help put in place future screening and possible ways to stop infection. To do this saliva, hair, and nails will be utilized to screen for the presence of HCMV, HHV-6A and HHV-6B, so that we can determine in newborns:
- how common infections with these viruses are,
- how common the integrated form of HHV-6A and HHV-6B is, and,
- what features can help doctors and nurses recognize early those who are infected.

This will provide very important information that can be used to plan for medications or vaccines against these infections.

3. Why have I been chosen?
You have been requested to participate in this study because you have recently had a baby. Approximately 200 mothers and their newborns will take part in this research.

4. Do I have to take part?
No. Your participation is absolutely voluntary. If you decide to participate, you will be asked to sign a consent form to record this voluntary decision.

5. What will happen to me if I take part?
We will ask you a few general questions, such as where you live, and then about your child's birth and current condition. We will need to get some of this information from your medical records or those of your child. Next, we will collect some saliva from your child's mouth using a swab, and also a few finger / toe nail clippings or hair follicles. In place of the nails or hair, a small piece of discarded placenta or umbilical tissue can be collected. These samples can be used for new virus diagnoses instead of blood; therefore no blood will be collected in this study. All this will be done by a trained healthcare professional and should take only a few minutes. The following diagram displays what will happen.



6. Payment for participation
There will be no payment for participating in this study.

7. What are the possible disadvantages and risks of taking part?
There may be some mild discomfort when removing hair strands and/or clipping the nails.

8. What are the possible benefits of taking part?
We cannot promise that the study will help you immediately, but it is hoped that the information we get will in the future help improve the diagnosis and management of children with HCMV, HHV-6A or HHV-6B. There will be consultation with clinicians, therefore a level of additional clinical care.

9. Will my taking part in the study be kept confidential?
Yes. All information about your participation in this study is confidential and will be kept securely in a locked cabinet; electronic records will be stored in a password-protected database. Neither you nor your child will be named in any reports of the study. These will only be looked at by

authorised medical staff within the hospital establishment for the purpose of providing follow-up medical care or advice to you or your child, if necessary.

10. What will happen to any samples I give?
Samples will be examined at the UTH Virology Laboratory by highly trained and experienced scientists. If any material will be left over, it will be stored securely, unless you indicate that you would like it destroyed. For research records and samples that may require to be sent for further tests to co-workers outside UTH, neither you nor your child will be linked to them or identified by name, or address. The records and specimens will be coded.

11. What if I don't want to carry on with the study?
You are free to withdraw from participating in the study at any time, without having to give a reason. You can also request that no information is used.

12. Contact Details
If you have any concerns or require further information about your participation in this study, you may contact: – Dr. Kunda Musonda on 0211-253236 or 0974-136520
– Dr. Mwaka Monze on 0211-256078 or 0977-793654
You may also contact:
The Chairperson – Biomedical Research Ethics Committee, University of Zambia School of Medicine, Ridgeway campus, Lusaka. Telephone 0211-256067.

13. What will happen to the results of the study?
The findings of this study will be published in scientific journals. You/your child will not be identified in any publication unless you have agreed to release such information.
If you would like to be told the laboratory findings from examination of your samples or those of your child, you may make this request to the researchers at the end of the study, using the contact details given above.

14. Who is organising and funding the research?
This research is part of Dr. Kunda Musonda's PhD studies. He is a Medical Doctor based at UTH, and currently studying under a Commonwealth Scholarship at the London School of Hygiene and Tropical Medicine, University of London, in London, UK. Part of this research is supported by the HHV-6 Foundation. The funders have no part in the study.

15. Who has reviewed the study?
This study has been approved by the Research Ethics Committee of the London School of Hygiene and Tropical Medicine, as well as the University of Zambia Biomedical Research Ethics Committee.

Thank you for taking the time to read this information sheet.

APPENDIX 6: INFORMED CONSENT FORM –NEONATAL STUDY

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT United Kingdom.

INFORMED CONSENT FORM

8th May 2013, version 1.0 Study protocol number: **UNZABREC 010-06-13** / **LSHTM REC 6456**

Study: **Molecular characterisation, transmission routes, and developmental effects of infant infections with Human betaherpesviruses (HCMV, HHV-6A and HHV-6B) in Zambia: a pilot study.**

Principal investigator: Dr Ursula Gompels, London School of Hygiene & Tropical Medicine

1. I confirm that I have read the information sheet concerning this study [or have understood the verbal explanation] and I understand what will be required of my child and what will happen to her/him if she/he takes part in it.
2. My questions concerning this study have been fully answered by

   ……..(Name of person obtaining consent)

3. I understand that my child's participation is voluntary and I may withdraw her/him from this study at any time without giving a reason and without affecting her/his normal medical care and management.
4. I understand that where it is relevant to this research, sections of any of my child's medical records may be looked at by the researchers. I give permission for the researchers to access my child's records.
5. I understand that my child will not be named in any reports arising from this research, and that confidentiality shall be maintained at all times.
6. I agree for my child to take part in this study.

Name of child:____ Date of Birth :______________

Name of child's parent/guardian:____

Signature or thumb print of parent/guardian:____

Relationship to child______________________________________ Date__________________

Name of witness:_____________________________________ Signature: __________________

Date___________________

Person obtaining consent:______________________________ Signature: __________________

Date___________________

APPENDIX 7: DATA CAPTURE FORM – NEONATAL STUDY

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT United Kingdom.

8th May 2013, version 1.0 Study protocol number: **UNZABREC 010-06-13** / **LSHTM REC 6456**

Study: **Molecular characterisation, transmission routes, and developmental effects of infant infections with Human betaherpesviruses (HCMV, HHV-6A and HHV-6B) in Zambia: a pilot study.**

Date:.. Ward:...

Hospital No:.. Study No:**UTH**..

NEONATE

General Information *1min* *5min* *10min*

1. Sex: M / F APGAR SCORE: _____ _____ ____
2. Date of Birth:............................
3. Time of Birth:............................
4. Postnatal Age at recruitment:............................days / hours
5. Gestational age at delivery:............................weeks => Term / Pre-Term
6. Mode of delivery: SVD / Instrumental / Caesarean: Indication..
7. Birth weight:............•..............Kg
8. Reason for Admission / Attendance:...
9. Diagnosis:...

Clinical assessment

10. Jaundice	Y / N	13. Hepatomegaly	Y / N
11. Petechiae	Y / N	14. Splenomegaly	Y / N
12. Purpura	Y / N	15. Microcephaly	Y / N
16. Convulsions	Y / N		

17. Congenital malformation / Other findings:...

Laboratory Results

18. Hemoglobin level :..................... g/dL = Anaemia Y / N
19. Platelet count :.....................X10^9 /L = Thrombocytopenia Y / N
20. ALT :...................../L = Elevated Y / N
21. AST :...................../L = Elevated Y / N
22. Bilirubin: a) Conjugated :...................../L = Elevated Y / N
b) Unconjugated :...................../L = Elevated Y / N
23. HIV Exposed: Y / N / U 24. HIV PCR Result: P / N / U

Samples Collected

Saliva ☐ *Time* ☐ Nails ☐ *Time* ☐ Hair ☐ *Time* ☐ Umbilical ts ☐ *Time* ☐ Initials ☐

Breastfed: Y / N / U

MOTHER

1. Date of Birth:.. 2. Age: ....years
3. Residential area:.. 4. Occupation:..
5. Marital Status: Married / Single 6. Children: *Alive*..........*Dead*..........*Miscarried*..........
7. Education: Primary / Secondary / Tertiary / Never 8. HIV: P / N / U 9. HAART/PMTCT: Y / N

APPENDIX 8: HCMV UL55 OLIGONUCLEOTIDE PRIMER AND PROBE LOCATION

gB1 (boxed, positions 16–40: GAGGACAACGAAATCCTGTTGGGCA)

```
HCMV_AD169_c82678-82499  1 TACGGTCAACTGGGCGAGGACAACGAAATCCTGTTGGGCAACCACCGCACTGAGGAATGTCAGCTTCCCAGCCTCAAGATCTTCATCGCC 90
HCMV_BE/11/2010          1 .......................................................................................... 90
HCMV_BE/10/2010          1 .......................................................................................... 90
HCMV_BE/9/2010           1 .......................................................................................... 90
HCMV_HAN31               1 .......................................................................................... 90
HCMV_HAN28               1 .......................................................................................... 90
HCMV_HAN19               1 .......................................................................................... 90
HCMV_HAN16               1 .......................................................................................... 90
HCMV_HAN8                1 .......................................................................................... 90
HCMV_HAN2                1 .......................................................................................... 90
HCMV_6397                1 .......................................................................................... 90
HCMV_TB40E               1 .......................................................................................... 90
HCMV_AF1                 1 .......................................................................................... 90
HCMV_U11                 1 .......................................................................................... 90
HCMV_JP                  1 .......................................................................................... 90
HCMV_Merlin              1 .......................................................................................... 90
HCMV_HAN22               1 .......................................................................................... 90
HCMV_Davis               1 ..............................T........................................................... 90
HCMV_TR                  1 .......................................................................T.................. 90
HCMV_U8                  1 ..............................T........................................................... 90
HCMV_3157                1 ..............................T........................................................... 90
HCMV_BE/27/2010          1 ..............................T........................................................... 90
HCMV_3301                1 ..............................T........................................................... 90
HCMV_HAN12               1 ....................T..................................................................... 90
HCMV_Towne               1 ....................T..................................................................... 90
HCMV_BE/21/2010          1 ..............T.....T..................................................................... 90
HCMV_HAN38               1 ..............T................................................T.......................... 90
HCMV_Toledo              1 ..............T................................................T.......................... 90
HCMV_JHC                 1 ..............T................................................T.......................... 90
HCMV_VR1814              1 ..............T................................................T.......................... 90
HCMV_HAN1                1 ..............T................................................T.......................... 90
HCMV_HAN20               1 ..............T................................................T.......................... 90
HCMV_HAN3                1 ..............T................................................T.......................... 90
HCMV_HAN13               1 ..............T................................................T.......................... 90
```

```
                                                    gB-P3                            gB2
HCMV_AD169_c82678-82499 91 GGGAACTCGGCCTACGAGTA[CGTGGACTACCTCTTCAAACGCATGATTG]A[CCTCAGCAGTATCTCCACCGTCGA]CAGCATGATCGCCCTG 180
HCMV_BE/11/2010         91 .......................................................................................... 180
HCMV_BE/10/2010         91 .......................................................................................... 180
HCMV_BE/9/2010          91 .......................................................................................... 180
HCMV_HAN31              91 .......................................................................................... 180
HCMV_HAN28              91 .......................................................................................... 180
HCMV_HAN19              91 .......................................................................................... 180
HCMV_HAN16              91 .......................................................................................... 180
HCMV_HAN8               91 .......................................................................................... 180
HCMV_HAN2               91 .......................................................................................... 180
HCMV_6397               91 .......................................................................................... 180
HCMV_TB40E              91 .......................................................................................... 180
HCMV_AF1                91 .......................................................................................... 180
HCMV_U11                91 .......................................................................................... 180
HCMV_JP                 91 .......................................................................................... 180
HCMV_Merlin             91 .......................................................................................... 180
HCMV_HAN22              91 ..A....................................................................................... 180
HCMV_Davis              91 .......................................................................................... 180
HCMV_TR                 91 .......................................................................................... 180
HCMV_U8                 91 .......................................................................................... 180
HCMV_3157               91 .......................................................................................... 180
HCMV_BE/27/2010         91 .......................................................................................... 180
HCMV_3301               91 .......................................................................................... 180
HCMV_HAN12              91 ..C........................................................C.............................A 180
HCMV_Towne              91 ..C........................................................C.............................A 180
HCMV_BE/21/2010         91 ..C........................................................C........T....................A 180
HCMV_HAN38              91 ..............T..A..T......................................C.............................. 180
HCMV_Toledo             91 ..............T..A..T......................................C.......................T...... 180
HCMV_JHC                91 ..............T..A..T......................................C.....T........................ 180
HCMV_VR1814             91 ..............T..A..T......................................C.....T..A..................... 180
HCMV_HAN1               91 ..............T..A..T......................................C.....T..A..................... 180
HCMV_HAN20              91 ..............T..A..T......................................C.....T..A..................... 180
HCMV_HAN3               91 ..............T..A..T......................................C.....T..A..................... 180
HCMV_HAN13              91 ..............T..A..T......................................C.....T..A..................... 180
```

Appendix 8 is an alignment of representative complete HCMV genome sequences. Highlighted are the locations of oligonucleotide primers (gB1 and gB2) and labelled probe (gB-P3) used for HCMV UL55 (gB) detection and quantification in this thesis.

APPENDIX 9: HCMV UL74 OLIGONUCLEOTIDE PRIMER LOCATION

```
                                                                   gO-up
HCMV_AD169_c108590-107780/1-811   1 CCGATGCAACAACGGTAGATGAGCAGCAAAACGACCAGAATCAGCAGTGAGTACACGCAGGCAAGCCAAACCACAAGGCAGACGGACGGTGCGGGGTCTCCTCCTCTGTCATGGGGAGAA 120
HCMV_BE/9/2010/1-811              1 ........................................................................................................................ 120
HCMV_BE/27/2010/1-811             1 ........................................................................................................................ 120
HCMV_U11/1-811                    1 ........................................................................................................................ 120
HCMV_JHC/1-811                    1 ...........G............................................................................................................ 120
HCMV_TR/1-811                     1 ...........G............................................................................................................ 120
HCMV_HAN28/1-811                  1 ...........G.................................................................................................T.......... 120
HCMV_HAN3/1-811                   1 ........................................................................................................................ 120
HCMV_HAN20/1-811                  1 ........................................................................................................................ 120
HCMV_VR1814/1-811                 1 ........................................................................................................................ 120
HCMV_TB40E/1-811                  1 .................................................................................................T...................... 120
HCMV_Toledo/1-811                 1 ..............................................................................................A......................... 120
HCMV_BE/21/2010/1-811             1 ......................................G...................................T................T..AA........................ 120
HCMV_HAN1/1-811                   1 ......................................G...................................T................T..AA........................ 120
HCMV_HAN8/1-811                   1 ...........G...............................................................................TA.A..................T..G... 120
HCMV_Davis/1-811                  1 ......................................G...................................T................T..AA........................ 120
HCMV_HAN13/1-811                  1 ......................................G...................................T................T..AA........................ 120
HCMV_HAN38/1-811                  1 ...........G...............................................................................TA.A..................T..G... 120

HCMV_AD169_c108590-107780/1-811 121 AAGAGATGATGGTGAGAGACGTCCCTAAGATGGTGTTTCTAATATCTATATCTTTCTTGCTTGTTTCTTTCATAAACTGTAAAGTTATGTCAAAAGCGCTTTATAATCGTCCTTGGAGGG 240
HCMV_BE/9/2010/1-811            121 ...AGAT....GTGA.A.ACG..CCTAA.A.GG.GT..C.AA..T....A.CT..C..GC..G..TCTC.C.TA..CT..AAA.....GT..AA...G..T...AA..G.CCTTGGA..G 240
HCMV_BE/27/2010/1-811           121 ...AGAT....GTGA.A.ACG..CCTAA.A.GG.GT..C.AA..T....A.CT..C..GC..G..TCTC.C.TA..CT..AAA.....GT..AA...G..T...AA..G.CCTTGGA..G 240
HCMV_U11/1-811                  121 ...AGAT....GTGA.A.ACG..CCTAA.A.GG.GT..C.AA..T....A.CT..C..GC..G..TCTC.C.TA..CT..AAA.....GT..AA...G..T...AA..G.CCTTGGA..G 240
HCMV_JHC/1-811                  121 ..............G.......T..T..A.....CC....G.C..TA.C....C..C....C.C.T.C......TG...G..T.CGG...G.TT.T.CG.AGGA...TACT.GT....TA 240
HCMV_TR/1-811                   121 ..............G.......T..T..A.....CC....G.CG.TA.C....C..C....C.C.T.C......TG...G..T.CGG...G.TT.T.CG.AGGA...TACT.GT....TA 240
HCMV_HAN28/1-811                121 ..............GA......T..T..A.....CC....G.CG.TA.C....C..C....C.C.T.C......TG...G..T.CGG..AG.TT.T.CG.AGGA....ACT.GT....TA 240
HCMV_HAN3/1-811                 121 ...A...C......A.CA...CT.....A...........A.C..TG.C.G..C.........C..........TG..AG.TC....G.....CG.G..GT..CT.GTTAG.T....TA. 240
HCMV_HAN20/1-811                121 ...A...C......A.CA...CT.....A...........A.C..TG.C.G..C.........C..........TG..AG.TC....G.....CG.G..GT..CT.GTTAG.T....TA. 240
HCMV_VR1814/1-811               121 ...A...C......A.CA...CT.....A...........A.C..TG.C.G..C.........C..........TG..AG.TC....G.....CG.G..GT..CT.GTTAG.T....TA. 240
HCMV_TB40E/1-811                121 ...A...C......A.CA...CT.....A...........A.C..TG.C.G..C.........C..........TG..AG.TC....G.....CG.G..GT..CT.GTTAG.T....TA. 240
HCMV_Toledo/1-811               121 .......C......A.CA...CT.....A...........A.C..TG.C.G..C.........C..........TG..AG.TC..C.G.....CG.G..GT..CT.GTTAG.T....TA. 240
HCMV_BE/21/2010/1-811           121 .G............G.....AA.TT......TC....ATGGC..A..T.CC.G.....C.....C....ACTGT..G......CGATC..A........CA.T.........G.T.T... 240
HCMV_HAN1/1-811                 121 .G............G.....AA.TT......TC....ATGGC..A..T.CC.G.....C.....C....ACTGT..G......CGATC..A........CA.T.........G.T.T... 240
HCMV_HAN8/1-811                 121 ..............G.....AA.TT......TC....ATGG...A..T.CC.G.....C.....C....ACTGT..GG.....CGATC..A........CA.T.........G.T.T... 240
HCMV_Davis/1-811                121 .G............G.....AA.TT......TC....ATGGC..A..T.CC.G.....C.....C....A.TGT..G......CGATC..A........CA.T.........G.T.T... 240
HCMV_HAN13/1-811                121 .G............G.....AA.TT......TC....ATGGC..A..T.CC.G.....C.....C....A.TGT..G......CGATC..A........CA.T.........G.T.T... 240
HCMV_HAN38/1-811                121 ..............G.....AA.TT......TC....ATGGC..A..T.CC.G.....C.....C....ACTGT..GG.....CGATC..A........CA.T.........G.T.T... 240

HCMV_AD169_c108590-107780/1-811 241 GCTTGGTACTGTCTAAGATAGGCAAATATAAATTAGATCAGCTTAAGTTAGAAATTTTGAGACAACTAGAAACGACTATTTCTACAAAATACAATGTAAGTAAACAACCGGTTAAAAATC 360
HCMV_BE/9/2010/1-811            241 GC.TGGTAC.G..T..G.T.GGC...T.T.A..T.G.TC.GC..AAG...G..A.T.TGAG.CA..T.G.A.CG..TAT...T.C...AT.C.A.G.AA.T......CCGG....A...C 360
HCMV_BE/27/2010/1-811           241 GC.TGGTAC.G..T..G.T.GGC...T.T.A..T.G.TC.GC..AAG...G..A.T.TGAG.CA..T.G.A.CG...AT...T.C...AT.C.A.G.AA.T......CCGG....A...C 360
HCMV_U11/1-811                  241 GC.TGGTAC.G..T..G.T.GGC...T.T.A..T.G.TC.GC..AAG...G..A.T.TGAG.CA..T.G.A.CG...AT...T.C...AT.C.A.G.AA.T......CCAG....A...C 360
HCMV_JHC/1-811                  241 .AGTACT.TCTA.G.TAGGC..GC..AGACT.GAC....TC.AAT..GAG.TTT.GAA....C..G..A.GG...TCT..A.....T..TT..A....ACC.G....C.C.....A...C 360
HCMV_TR/1-811                   241 .AGTACT.TCTA.G.TAGGC..GC..AGACT.GAC....TC.AAT..GAG.TTT.GAA....C..G..A.GG...TCT..A.....T..TT..A....ACC.G....C.C.....A...C 360
HCMV_HAN28/1-811                241 .AGTACT.TCTA.G.TAGGC..GC..AGACT.GAC....TC.AAT..GAG.TTT.GAA....C..G..A.GG...TCT..A.....T..TT..A....ACC.G....C.C.....A...C 360
HCMV_HAN3/1-811                 241 T...T...A..GGC..GC.....CT.G.T.A..TC....T.GAGA.T.......A..T.G...GGG..CCTT.....A.A.ACTTC..T.TG....G..A.C.CGTT..A.A.CT..C.. 360
HCMV_HAN20/1-811                241 T...T...A..GGC..GC.....CT.G.T.A..TC....T.GAGA.T.......A..T.G...GGG..CCTT.....A.A.ACTTC..T.TG....G..A.C.CGTT..A.A.CT..C.. 360
HCMV_VR1814/1-811               241 T...T...A..GGC..GC.....CT.G.T.A..TC....T.GAGA.T.......A..T.G...GGG..CCTT.....A.A.ACTTC..T.TG....G..A.C.CGTT..A.A.CT..C.. 360
HCMV_TB40E/1-811                241 T...T...A..GGC..GC.....CT.G.T.A..TC....T.GAGA.T.......A..T.G...GGG..CCTT.....A.A.ACTTC..T.TG....G..A.C.CGTT..A.A.CT..C.. 360
HCMV_Toledo/1-811               241 T...T...A..GGC..GC.....CT.G.T.A..TC....T.GAGA.T.......A..T.G...GGG..CCTT.....A.A.ACTTC..T.TG....GA.A.C.CGTT..A.A.CT..C.. 360
HCMV_BE/21/2010/1-811           241 ..C.........T...A..GCT...G..ATT.AA.TTAG.....T..AG.C.....GA..CT..T.CT....TA.AGT.CC..C...T......G.CT...GTT....GTC...C.GT.. 360
HCMV_HAN1/1-811                 241 ..C.........T...A..GCT...G..ATT.AA.TTAG.....T..AG.C.....GA..CT..T.CT....TA.AGT.CC..C...T......G.CT...GTT....GTC...C.GT.. 360
HCMV_HAN8/1-811                 241 .CC.........T...A..GCT...G..ATT.AA.TTAG.....T..AG.C.....GA..CT..T.CT....TA.AGT.CC..C...T......G.CT...GTT....GTC...C.GT.. 360
HCMV_Davis/1-811                241 ..C.........T...A..GCT...G..ATT.AA.TTAG.....T..AG.C.....GA..CT..T.CT....TA.AGT.CC..C...T......G.CT...GTT....GTC...C.GT.. 360
HCMV_HAN13/1-811                241 ..C.........T...A..GCT...G..ATT.AA.TTAG.....T..AG.C.....GA..CT..T.CT....TA.AGT.CC..C...T......G.CT...GTT....GTC...C.GT.. 360
HCMV_HAN38/1-811                241 .CC.........T...A...CT...G..ATT.AA.TTAG.....T..AG.C.....GA..CT..T.CT....TA.AGT.CC..C...T......G.CT...GTT....GTC...C.GT.. 360
```

```
HCMV_AD169_c108590-107780/1-811 481 ACGTTTACTCACAGTACAATCATACGGCTAAAACGATAACATTCAGACCCCCACCTTGTGGTACTGTGCCTTCCATGACTTGTCTTTCCGAAATGCTAAACGTTTCCAAACGTAATGATA 600
HCMV_BE/9/2010/1-811            481 ....T....C...G....A.....CGG......G......TT..G.C.C.C.C....T.........C....C..G.........C..................C..A.........A 600
HCMV_BE/27/2010/1-811           481 ....C....C...G....A.....CGG......G......TT..G.C.C.C.C....T.........C....C..G.........C..................C..A.........A 600
HCMV_U11/1-811                  481 ....C....C...G....A.....CGG......G......TT..G.C.C.C.C....T.........C....C..G.........C..................C..A.........A 600
HCMV_JHC/1-811                  481 ....T.T..C.G.A....A.....CGG.............TT..G.C.C.C.T....C.....G...C....C..G.........C..................C..A.........A 600
HCMV_TR/1-811                   481 ....T.T..C.G.A....A.....CGG.............TT..G.C.C.C.T....C.....G...C....C..G.........C..................C..A.........A 600
HCMV_HAN28/1-811                481 ....T.T..C.G.A....A.....CGG.............TT..G.C.C.C.T....C.....G...C....C..G.........C..................C..A.........A 600
HCMV_HAN3/1-811                 481 ..TC.C.G..CA.TC.T..GGC......G.T....T.C.G....C.....T.T.G.ACA.TGC..TCA.TGA..TG.CT..CCGAAA.GTT..ACG.TTC.AAACG....GA..C..GC. 600
HCMV_HAN20/1-811                481 ..TC.C.G..CA.TC.T..GGC......G.T....T.C.G....C.....T.T.G.ACA.TGC..TCA.TGA..TG.CT..CCGAAA.GTT..ACG.TTC.AAACG....GA..C..GC. 600
HCMV_VR1814/1-811               481 ..TC.C.G..CA.TC.T..GGC......G.T....T.C.G....C.....T.T.G.ACA.TGC..TCA.TGA..TG.CT..CCGAAA.GTT..ACG.TTC.AAACG....GA..C..GC. 600
HCMV_TB40E/1-811                481 ..TC.C.G..CA.TC.T..GGC......G.T....T.C.G....C.....T.T.G.ACA.TGC..TCA.TGA..TG.CT..CCGAAA.GTT..ACG.TTC.AAACG....GA..C..GC. 600
HCMV_Toledo/1-811               481 ..TC.C.G..CA.TC.T..GGC......G.T....T.C.G....C.....T.T.G.ACA.TGC..TCA.TGA..TG.CT..CCGAAA.GTT..ACG.TTC.AAACG....GA..C..GC. 600
HCMV_BE/21/2010/1-811           481 C.A....TA.T...ACGG.C.GA....TA.CGTTT.A.C.C..AT.G.G.G.C..AGT.CC.T.CA......G.C..T..GAAA.GC.AA.CG.TTCC..GCG.AA.G..ACCGGC..A. 600
HCMV_HAN1/1-811                 481 C.A....TA.T...ACGG.C.GA....TA.CGTTT.A.C.C..AT.G.G.G.C..AGT.CC.T.CA......G.C..T..GAAA.GC.AA.CG.TTCC..GCG.AA.G..ACCGGC..A. 600
HCMV_HAN8/1-811                 481 CTA.G..TA.T...ACGG.C.GA....TA.CGTTC.A.C.C..GT.G.G.G.C..AGT.CC.T.CA......G.C..T..GAAA.GC.AA.CG.TTCC..GCG.AA.G..ACCGGC..A. 600
HCMV_Davis/1-811                481 C.A....TA.T...ACGG.C.GA....TA.CGTTT.A.C.C..AT.G.G.G.C..AGT.CC.T.CA......G.C..T..GAAA.GC.AA.CG.TTCC..GCG.AA.G..ACCGGC..A. 600
HCMV_HAN13/1-811                481 C.A....TA.T...ACGG.C.GA....TA.CGTTT.A.C.C..AT.G.G.G.C..AGT.CC.T.CA......G.C..T..GAAA.GC.AA.CG.TTCC..GCG.AA.G..ACCGGC..A. 600
HCMV_HAN38/1-811                481 CTA.G..TA.T...ACGG.C.GA....TA.CGTTC.A.C.C..GT.G.G.G.C..AGT.CC.T.CA...G..G.C..T..GAAA.GC.AA.CG.TTCC..GCG.AA.G..ACCGGC..A. 600

HCMV_AD169_c108590-107780/1-811 601 CTGGCGAACAAGGTTGCGGTAATTTCACCACGTTCAACCCCATGTTTTTCAATGTACCGCGTTGGAACACCAAATTGTACGTGGGTCCGACTAAGGTTAACGTAGATAGTCAAACGATTT 720
HCMV_BE/9/2010/1-811            601 CT....AA.A...T.....T.A..............A.CC..........A...TA...C..T.....A....AT....C.T.G..C.G.C...G.....C...........AA.GA... 720
HCMV_BE/27/2010/1-811           601 CT....AA.A...T.....T.A..............A.CC..........A...TA...C..T.....A....AT....C.T.G..C.G.C...G.....C...........AA.GA... 720
HCMV_U11/1-811                  601 CT....AA.A...T.....T.A..............A.CC..........A...TA...C..T.....A....AT....C.T.G..C.G.C...G.....C...........AA.GA... 720
HCMV_JHC/1-811                  601 CT....AA.A...C.....C.A..............A.CC..........A...T...AC..T.....A.G...T.A....TAG..T.C..G.....A..C..G........AA.GA.C. 720
HCMV_TR/1-811                   601 CT....AA.A...C.....C.A..............A.CC..........A...T...AC..T.....A.G...T.A....TAG..T.C..G.....A..C..G........AA.GA.C. 720
HCMV_HAN28/1-811                601 CT....AA.A...C.....C.A..............A.CC..........A...T...AC..T.....A.G...T.A....TAG..T.C..G.....A..C..G........AA.GA.C. 720
HCMV_HAN3/1-811                 601 ..CAA...T.C..TAATTTT..CACGTT..ACCC...G..TT.CAA.G.A.CAC.TTG..A...CC...A.TGT.C...GG.C....C.AAAGT...C..AG..AGTC.A.CGAT..A.T... 720
HCMV_HAN20/1-811                601 ..CAA...T.C..TAATTTT..CACGTT..ACCC...G..TT.CAA.G.A.CAC.TTG..A...CC...A.TGT.C...GG.C....C.AAAGT...C..AG..AGTC.A.CGAT..A.T... 720
HCMV_VR1814/1-811               601 ..CAA...T.C..TAATTTT..CACGTT..ACCC...G..TT.CAA.G.A.CAC.TTG..A...CC...A.TGT.C...GG.C....C.AAAGT...C..AG..AGTC.A.CGAT..A.T... 720
HCMV_TB40E/1-811                601 ..CAA...T.C..TAATTTT..CACGTT..ACCC...G..TT.CAA.G.A.CAC.TTG..A...CC...A.TGT.C...GG.C....C.AAAGT...C..AG..AGTC.A.CGAT..A.T... 720
HCMV_Toledo/1-811               601 ..CAA...T.C..TAATTTT..CACGTT..ACCC...G..TT.CAA.G.A.CAC.TTG..A...CC...A.TGT.C...GG.C....C.AAAGT...C..AG..AGTC.A.CGAT..A.T... 720
HCMV_BE/21/2010/1-811           601 .....T..G.TAAC.T.ACC..G....ATC.TA.GT.T....ATG.ACCG.G.T..AAC.C..AATTG.A.GT..GTCCGA.T.AA.T...CGT..A..G.CA.ACG.T.T...T.C.AG 720
HCMV_HAN1/1-811                 601 .....T..G.TAAC.T.ACC..G....ATC.TA.GT.T....ATG.ACCG.G.T..AAC.CC.AATTG.A.GT..GTCCGA.T.AA.T...CGT..A..G.CA.ACG.T.T...T.C.AG 720
HCMV_HAN8/1-811                 601 .....T..G.TAAC.T.ACC..G....ATC.TA.GT.T....ATG.ACCG.G.T..AAC.CC.AATTG.A.GT..GTCCGA.T.AA.T...CGT..A..G.CA.ACG.T.T...T.T.AG 720
HCMV_Davis/1-811                601 .....T..G.TAAC.T.ACC..G....ATC.TA.GT.T....ATG.ACCG.G.T..AAC.CC.AATTG.A.GT..GTCCGA.T.AA.T...CGT..A..G.CA.ACG.T.T...T.C.AG 720
HCMV_HAN13/1-811                601 .....T..G.TAAC.T.ACC..G....ATC.TA.GT.T....ATG.ACCG.G.T..AAC.CC.AATTG.A.GT..GTCCGA.T.AA.T...CGT..A..G.CA.ACG.T.T...T.C.AG 720
HCMV_HAN38/1-811                601 .....T..G.TAACCTTACC..G....ATC.TA.GT.T....ATG.ACCG.G.T..AAC.CC.AATTG.A.GT..GTCCGA.T.AA.T...CGT..A..G.CA.ACG.T.T...T.T.AG 720
                                                                      gO-lw
HCMV_AD169_c108590-107780/1-811 721 ATTTTCTAGGTTTAACCGCCCTGCTTTTACGTTACGCACAACGCAACTGTACACACAGTTTCTACCTGGTTAACGCCATGAGCCGGAATCT 811
HCMV_BE/9/2010/1-811            721 .....C...GTT..AC.GC...........TT....A..A...........AC.............G.......C........GG..T.. 811
HCMV_BE/27/2010/1-811           721 .....C...GTT..AC.GC...........TT....A..A...........AC.............G.......C........GG..T.. 811
HCMV_U11/1-811                  721 .....C...GTT..AC.GC...........TT....A..A...........AC.............G.......C........GG..T.. 811
HCMV_JHC/1-811                  721 .C.......G....AC.GC.T.....C....TT....G..A...........GC.............G.......C........G...TT. 811
HCMV_TR/1-811                   721 .C.......G....AC.GC.T.....C....TT....G..A...........GC.............G.......C........G...TT. 811
HCMV_HAN28/1-811                721 .C.......G....AC.GC.T.....C....TT....G..A...........GC.............G.......C........G...TT. 811
HCMV_HAN3/1-811                 721 TGGG....A..GCC..A..T..A.G..ACGC...ACG.A..T...CACAC.GT.T.TACC.GGTTAA..CC.TGA..CG..AT...TT..G 811
HCMV_HAN20/1-811                721 TGGG....A..GCC..A..T..A.G..ACGC...ACG.A..T...CACAC.GT.T.TACC.GGTTAA..CC.TGA..CG..AT...TT..G 811
HCMV_VR1814/1-811               721 TGGG....A..GCC..A..T..A.G..ACGC...ACG.A..T...CACAC.GT.T.TACC.GGTTAA..CC.TGA..CG..AT...TT..G 811
HCMV_TB40E/1-811                721 TGGG....A..GCC..A..T..A.G..ACGC...ACG.A..T...CACAC.GT.T.TACC.GGTTAA..CC.TGA..CG..AT...TT..G 811
HCMV_Toledo/1-811               721 TGGG....A..GCC..A..T..A.G..ACGC...ACG.A..T...CACAC.GT.T.TACC.GGTTAA..CC.TGA..CG..AT...TT..G 811
HCMV_BE/21/2010/1-811           721 G...AACC.....G..T..A.GTTACGCG.A..G.AA.TG.ACAC..A..TT....CTGG.TA..GC.A.G.G.C.G.ATTTAT.TCG.G. 811
HCMV_HAN1/1-811                 721 G...AACC.....G..T..A.GTTACGCG.A..G.AA.TG.ACAC..A..TT....CTGG.TA..GC.A.G.G.C.G.ATTTAT.TCG.G. 811
HCMV_HAN8/1-811                 721 G...AACC.....G..T..A.GTTACGCG.A..G.AA.TG.ACAC..A..TT....CTGG.TA..GC.A.G.G.C.G.ATTTAT.TCG.G. 811
HCMV_Davis/1-811                721 G...AACC.....G..T..A.GTTACGCG.A..G.AA.TG.ACAC..A..TT....CTGG.TA..GC.A.G.G.C.G.ATTTAT.TCG.G. 811
HCMV_HAN13/1-811                721 G...AACC.....G..T..A.GTTACGCG.A..G.AA.TG.ACAC..A..TT....CTGG.TA..GC.A.G.G.C.G.ATTTAT.TCG.G. 811
HCMV_HAN38/1-811                721 G...AACC.....G..T..A.GTTACGCG.A..G.AA.TG.ACAC..A..TT....CTGG.TA..GC.A.G.G.C.G.ATTTAT.TCG.G. 811
```

Appendix 9 is an alignment of representative complete HCMV genome sequences. Highlighted are the locations of oligonucleotide primers (gO-up and gO-lw) used for amplification and Sanger sequencing of HCMV UL74 (gO) in this thesis.

APPENDIX 10: CHARACTERISTICS OF COMPLETE HCMV GENOMES USED IN THIS STUDY.

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KT634296.1_UKNEQAS2	234,873	Amniotic fluid	Fetus	20 weeks	2013	Australia	Sequenced directly from clinical specimen; sequenced from enriched viral DNA (SureSelect); 2 genes are mutated (RL6, RL9)
KC519319.1_BE/9/2010	235,631	Urine	-	-	09-Apr-10	Belgium	Passaged two times in human fibroblasts;
KC519320.1_BE/10/2010	235,215	Urine	-	-	22-Apr-10	Belgium	Passaged two times in human fibroblasts; gene UL9 is mutated;
KC519321.1_BE/11/2010	235,061	Urine	-	-	27-Apr-10	Belgium	Passaged two times in human fibroblasts; two genes are mutated (RL5A and UL1);
KC519322.1_BE/21/2010	235,722	Urine	-	-	26-Jul-10	Belgium	Sequenced directly from clinical material, no cell culture passage; three genes are mutated (RL5A, UL9 and UL150);
KC519323.1_BE/27/2010	234,810	Urine	-	-	21-Oct-10	Belgium	Passaged four times in human fibroblasts; four genes are mutated (RL5A, UL1, UL9 and UL111A);
KP745633.1_BE/45/2011	235,352	NPA	-	-	27-Jan-11	Belgium	Passaged one time in human fibroblasts
KP745634.1_BE/32/2010	235,205	Amniotic fluid	-	-	29-Dec-10	Belgium	Passaged one time in human fibroblasts; genes RL5A and US9 are mutated
KP745635.1_BE/5/2012	235,184	Urine	-	-	22-Jan-12	Belgium	Passaged two times in human fibroblasts
KP745636.1_BE/7/2011	237,117	Urine	-	-	17-Mar-11	Belgium	Passaged two times in human fibroblasts; genes RL12 and UL148 are mutated
KP745637.1_BE/9/2011	235,865	Urine	-	-	14-Apr-11	Belgium	Passaged two times in human fibroblasts; gene UL9 is mutated
KP745638.1_BE/15/2010	235,514	Urine	-	-	31-May-10	Belgium	Passaged three times in human fibroblasts; genes RL1, UL9 and UL111A are mutated; mutations were confirmed in original clinical material
KP745639.1_BE/10/2011	235,054	Urine	-	-	15-Apr-11	Belgium	Passaged two times in human fibroblasts; gene UL1 is mutated, mutation confirmed in original clinical material
KP745640.1_BE/22/2010	235,632	Urine	-	-	31-Aug-10	Belgium	Passaged four times in human fibroblasts
KP745641.1_BE/31/2011	235,844	Urine	-	-	27-Oct-11	Belgium	Passaged four times in human fibroblasts
KP745644.1_BE/31/2010	236,028	Urine	-	-	16-Dec-10	Belgium	Passaged four times in human fibroblasts
KP745645.1_BE/13/2010	236,032	Urine	-	-	06-May-10	Belgium	Passaged three times in human fibroblasts; gene US9 is mutated, mutation confirmed in original clinical material
KP745646.1_BE/8/2012	235,889	Urine	-	-	14-Feb-12	Belgium	Passaged three times in human fibroblasts; gene RL13 is mutated
KP745647.1_BE/18/2010	235,871	Urine	-	-	20-Jun-10	Belgium	Passaged five times in human fibroblasts
KP745648.1_BE/8/2011	235,111	Urine	-	-	13-Apr-11	Belgium	Passaged two times in human fibroblasts; gene US9 is mutated

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KP745649.1_BE/10/2012	234,754	Urine	-	-	02-Apr-12	Belgium	Passaged two times in human fibroblasts
KP745650.1_BE/1/2011	235,833	Urine	-	-	14-Jan-11	Belgium	Passaged three times in human fibroblasts; genes UL1 and UL9 are mutated
KP745651.1_BE/9/2012	235,836	Urine	-	-	26-Mar-12	Belgium	Passaged two times in human fibroblasts; genes UL1 and UL9 are mutated
KP745652.1_BE/2/2011	235,810	Urine	-	-	06-Jan-11	Belgium	Passaged four times in human fibroblasts
KP745653.1_BE/22/2011	235,612	Urine	-	-	18-Jul-11	Belgium	Passaged two times in human fibroblasts; genes RL6 and UL9 are mutated
KP745654.1_BE/19/2011	235,446	Urine	-	-	05-Jul-11	Belgium	Passaged two times in human fibroblasts; genes RL6 and UL9 are mutated
KP745655.1_BE/3/2010	236,597	Urine	-	-	18-Mar-10	Belgium	Passaged two times in human fibroblasts; genes RL5A and UL9 are mutated
KP745656.1_BE/2/2013	235,156	Urine	-	-	17-Jan-13	Belgium	Passaged three times in human fibroblasts; genes RL5A, UL128 and US9 are mutated
KP745657.1_BE/13/2011	235,713	Urine	-	-	10-May-11	Belgium	Passaged two times in human fibroblasts; genes RL5A, UL1, UL9 and US7 are mutated, mutations confirmed in original clinical material
KP745658.1_BE/14/2012	234,931	Urine	-	-	14-May-12	Belgium	Passaged one time in human fibroblasts; genes RL6, UL9, UL40 and US7 are mutated, RL6, UL9 and US7 mutations confirmed in original clinical material
KP745659.1_BE/3/2011	235,726	Urine	-	-	14-Jan-11	Belgium	Passaged four times in human fibroblasts
KP745660.1_BE/6/2011	235,101	Urine	-	-	10-Mar-11	Belgium	Passaged two times in human fibroblasts; gene UL9 is mutated, mutation confirmed in original clinical material
KP745661.1_BE/33/2010	235,605	NPA	-	-	27-Dec-10	Belgium	Passaged one time in human fibroblasts; genes RL6, UL9 and US9 are mutated
KP745662.1_BE/20/2010	235,516	Urine	-	-	20-Jul-10	Belgium	Passaged four times in human fibroblasts; genes RL5A and UL9 are mutated, UL9 mutation was confirmed in original clinical material
KP745663.1_BE/5/2010	236,345	Urine	-	-	25-Mar-10	Belgium	Passaged two times in human fibroblasts; genes RL6 and US9 are mutated
KP745665.1_BE/16/2012	235,910	Urine	-	-	24-May-12	Belgium	Passaged one time in human fibroblasts; gene UL150 is mutated, mutation confirmed in original clinical material
KP745666.1_BE/7/2012	236,053	Urine	-	-	14-Feb-12	Belgium	Passaged three times in human fibroblasts; genes RL13 and UL150 are mutated, UL150 mutation confirmed in original clinical material
KP745667.1_BE/5/2011	235,621	Urine	-	-	07-Mar-11	Belgium	Passaged seven times in human fibroblasts
KP745668.1_BE/18/2011	235,416	Urine	-	-	22-Jun-11	Belgium	Passaged two times in human fibroblasts
KP745669.1_BE/28/2011	235,732	NPS	-	-	20-Oct-11	Belgium	Passaged two times in human fibroblasts; gene UL40 is mutated

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KP745670.1_BE/30/2011	235,350	Urine	-	-	26-Oct-11	Belgium	Passaged two times in human fibroblasts; gene UL150 is mutated
KP745671.1_BE/14/2011	235,498	Urine	-	-	12-May-11	Belgium	Passaged nine times in human fibroblasts; genes UL9, UL11 and US6 are mutated
KP745672.1_BE/29/2011	236,364	Urine	-	-	20-Oct-11	Belgium	Passaged two times in human fibroblasts; genes RL5A and UL9 are mutated
KP745673.1_BE/42/2011	235,462	NPA	-	-	09-Jan-11	Belgium	Passaged one time in human fibroblasts; gene RL5A is mutated
KP745674.1_BE/33/2011	235,276	Urine	-	-	22-Nov-11	Belgium	Passaged two times in human fibroblasts
KP745675.1_BE/23/2011	235,425	NPS	-	-	19-Jul-11	Belgium	Passaged two times in human fibroblasts; gene UL9 is mutated
KP745676.1_BE/28/2010	235,974	Urine	-	-	15-Oct-10	Belgium	Passaged four times in human fibroblasts; genes RL5A and UL9 are mutated
KP745677.1_BE/1/2010	235,705	Urine	-	-	18-Jan-10	Belgium	Passaged two times in human fibroblasts; genes RL5A and US9 are mutated
KP745678.1_BE/25/2010	235,904	Urine	-	-	09-Sep-10	Belgium	Passaged two times in human fibroblasts; gene UL111A is mutated
KP745679.1_BE/24/2010	235,744	Urine	-	-	07-Sep-10	Belgium	Passaged two times in human fibroblasts; gene UL9 is mutated
KP745680.1_BE/11/2012	235,893	Urine	-	-	10-Apr-12	Belgium	Passaged two times in human fibroblasts; gene UL150 is mutated, mutation confirmed in original clinical material
KP745681.1_BE/43/2011	235,100	NPA	-	-	19-Jan-11	Belgium	Passaged one time in human fibroblasts; genes RL5A, RL6 and UL133 are mutated
KP745682.1_BE/46/2011	236,239	NPA	-	-	07-Mar-11	Belgium	Passaged one time in human fibroblasts; gene US9 is mutated
KP745683.1_BE/12/2011	235,258	Urine	-	-	09-May-11	Belgium	Passaged two times in human fibroblasts; genes RL6 and US9 are mutated, mutations confirmed in original clinical material
KP745684.1_BE/11/2011	234,806	Urine	-	-	18-Apr-11	Belgium	Passaged four times in human fibroblasts; gene RL6 is mutated
KP745686.1_BE/39/2011	235,982	NPA	-	-	01-Jan-11	Belgium	Passaged one time in human fibroblasts; genes UL9, UL111A, US9 are mutated
KP745687.1_BE/36/2011	234,373	Urine	-	-	02-Dec-11	Belgium	Passaged two times in human fibroblasts; genes RL6, UL9 and UL40 are mutated, mutations confirmed in original clinical material
KP745688.1_BE/12/2012	235,362	NPS	-	-	21-Apr-12	Belgium	Passaged two times in human fibroblasts; gene RL6 is mutated
KP745689.1_BE/17/2011	235,827	Urine	-	-	20-May-11	Belgium	Passaged two times in human fibroblasts; genes UL1 and UL9 are mutated
KP745690.1_BE/34/2011	235,290	Urine	-	-	28-Nov-11	Belgium	Passaged two times in human fibroblasts
KP745692.1_BE/3/2012	236,051	Urine	-	-	09-Jan-12	Belgium	Passaged two times in human fibroblasts; gene US13 is mutated

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KP745693.1_BE/15/2012	235,508	Urine	-	-	24-May-12	Belgium	Passaged one time in human fibroblasts; genes UL150 and US7 are mutated
KP745694.1_BE/12/2010	235,195	Urine	-	-	29-Apr-10	Belgium	Passaged eight times in human fibroblasts; gene RL5A is mutated
KP745695.1_BE/6/2012	235,164	Urine	-	-	31-Jan-12	Belgium	Passaged five times in human fibroblasts; genes RL6, US9 and US27 are mutated
KP745696.1_BE/27/2011	235,392	Urine	-	-	13-Sep-11	Belgium	Passaged five times in human fibroblasts; gene UL9 is mutated
KP745697.1_BE/23/2010	236,066	Urine	-	-	09-Sep-10	Belgium	Passaged four times in human fibroblasts; genes UL1 and UL9 are mutated
KP745698.1_BE/20/2011	235,272	Urine	-	-	05-Jul-11	Belgium	Passaged two times in human fibroblasts; gene US7 is mutated
KP745699.1_BE/1/2012	235,150	Urine	-	-	02-Jan-12	Belgium	Passaged two times in human fibroblasts; genes UL1 and UL136 are mutated
KP745700.1_BE/4/2011	235,808	Urine	-	-	03-Mar-11	Belgium	Passaged two times in human fibroblasts
KP745701.1_BE/6/2010	235,329	Urine	-	-	23-Mar-10	Belgium	Passaged two times in human fibroblasts; genes RL5A, UL9 and US7 are mutated
KP745702.1_BE/21/2011	235,849	Urine	-	-	15-Jul-11	Belgium	Passaged five times in human fibroblasts
KP745703.1_BE/26/2011	234,902	Urine	-	-	23-Aug-11	Belgium	Passaged two times in human fibroblasts
KP745704.1_BE/32/2011	235,633	Urine	-	-	17-Nov-11	Belgium	Passaged two times in human fibroblasts; gene UL150 is mutated
KP745705.1_BE/38/2011	235,775	NPS	-	-	12-Dec-11	Belgium	Passaged two times in human fibroblasts; genes RL6 and UL9 are mutated
KP745706.1_BE/41/2011	235,332	BAL	-	-	07-Jan-11	Belgium	Passaged one time in human fibroblasts; genes UL9 and UL111A are mutated
KP745707.1_BE/13/2012	235,015	Urine	-	-	23-Apr-12	Belgium	Passaged two times in human fibroblasts; gene UL9 is mutated
KP745708.1_BE/8/2010	235,964	Urine	-	-	12-Apr-10	Belgium	Passaged one time in human fibroblasts; gene UL9 is mutated, mutation confirmed in original clinical material
KP745709.1_BE/48/2011	235,747	NPA	-	-	28-Mar-11	Belgium	Passaged one time in human fibroblasts; genes UL40 and US9 are mutated
KP745710.1_BE/2/2012	236,100	Urine	-	-	09-Jan-12	Belgium	Passaged two times in human fibroblasts; genes RL5A and UL133 are mutated
KP745711.1_BE/24/2011	235,745	Urine	-	-	04-Aug-11	Belgium	Passaged two times in human fibroblasts
KP745712.1_BE/19/2010	235,365	Urine	-	-	06-Jul-10	Belgium	Passaged five times in human fibroblasts; genes UL9 and UL136 are mutated
KP745713.1_BE/35/2011	235,941	Urine	-	-	30-Nov-11	Belgium	Passaged two times in human fibroblasts; genes UL142 and US9 are mutated, UL142 mutation confirmed in original clinical material

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KP745714.1_BE/29/2010	234,922	Urine	-	-	07-Dec-10	Belgium	Passaged seven times in human fibroblasts
KP745715.1_BE/44/2011	235,301	NPA	-	-	27-Jan-11	Belgium	Passaged one time in human fibroblasts; genes RL5A, UL150A, IRS1 and US9 are mutated
KP745716.1_BE/16/2010	235,366	NPS	-	-	06-Jun-10	Belgium	Passaged five times in human fibroblasts; genes UL1, UL111A and US9 are mutated
KP745717.1_BE/2/2010	235,138	NPS	-	-	16-Mar-10	Belgium	Passaged two times in human fibroblasts; genes UL9 and UL111A are mutated
KP745719.1_BE/26/2010	235,908	Urine	-	-	15-Sep-10	Belgium	Passaged two times in human fibroblasts; gene UL150 is mutated
KP745720.1_BE/15/2011	235,905	Urine	-	-	13-May-11	Belgium	Passaged five times in human fibroblasts; genes RL5A and UL9 are mutated, mutations confirmed in original clinical material
KP745721.1_BE/14/2010	234,537	NPS	-	-	11-May-10	Belgium	Passaged two times in human fibroblasts; genes RL6, UL9 and UL40 are mutated
KP745722.1_BE/40/2011	235,716	NPA	-	-	04-Jan-11	Belgium	Passaged one time in human fibroblasts; gene RL6 is mutated
KP745723.1_BE/37/2011	234,858	NPS	-	-	05-Dec-11	Belgium	Passaged five times in human fibroblasts; gene RL6 is mutated
KP745724.1_BE/4/2012	234,950	Urine	-	-	16-Jan-12	Belgium	Passaged two times in human fibroblasts; gene RL5A is mutated, mutation confirmed in original clinical material
KP745725.1_BE/49/2011	235,317	NPA	-	-	26-Apr-11	Belgium	Passaged one time in human fibroblasts; gene RL5A is mutated
KP745726.1_BE/30/2010	235,642	Urine	-	-	08-Dec-10	Belgium	Passaged two times in human fibroblasts; genes UL9, US9 and US27 are mutated, US27 mutation confirmed in original clinical material
KP745727.1_BE/17/2010	235,836	Urine	-	-	14-Jun-10	Belgium	Passaged four times in human fibroblasts; genes RL5A, RL12 and UL111A are mutated
KP745728.1_BE/4/2010	236,428	Urine	-	-	22-Mar-10	Belgium	Passaged two times in human fibroblasts; genes RL5A and UL111A are mutated
KJ426589.1_isolate HAN	236,144	-	-	-	10-Jan-07	China	-
KP745642.1_CZ/1/2012	235,030	Urine	-	-	23-May-12	Czech Republic	Passaged two times in human fibroblasts; gene UL9 is mutated
KP745643.1_CZ/2/2012	235,226	Urine	-	-	15-Oct-12	Czech Republic	Passaged two times in human fibroblasts; gene US13 is mutated
KP745664.1_CZ/2/2013	235,191	Blood	-	-	01-Mar-13	Czech Republic	Passaged two times in human fibroblasts; genes RL6, UL9 and UL40 are mutated
KP745685.1_CZ/3/2012	234,598	Urine	-	-	05-Nov-12	Czech Republic	Passaged two times in human fibroblasts; gene UL145 is mutated
KP745691.1_CZ/1/2013	235,139	Blood	-	-	22-Feb-13	Czech Republic	Passaged two times in human fibroblasts; gene UL111A is mutated
KP745718.1_CZ/1/2011	234,758	Urine	-	-	20-May-11	Czech Republic	Passaged two times in human fibroblasts; gene UL9 is mutated

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KT959235.1_DB	235,512	Cervical swab	Pregnant woman	30 years	2009	France	Passaged 3 times in monocyte-derived human macrophages; 2 genes are mutated (RL13, UL9); 1 gene has disabling SNPS (UL130); 1 gene has an SNP (UL100); US9 is the short form
GQ221973.1_HAN13	236,219	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts; 1 gene is mutated (RL5A)
GQ396662.1_HAN38	236,112	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts; 2 genes are mutated (RL6 and US9)
GQ396663.1_HAN20	235,728	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts
JX512199.1_HAN1	235,006	BAL	-	-	2007	Germany	Acronym:
JX512200.1_HAN2	232,940	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts; 7 genes are mutated (UL1, UL6, UL7, UL8, UL9, UL30 and US7); except for the UL6-UL9 deletion, these mutations are also present in the original clinical specimen
JX512201.1_HAN3	235,703	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts
JX512202.1_HAN8	234,951	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts; 1 gene is mutated (UL111A); this mutation is also present in the original clinical specimen
JX512203.1_HAN12	236,006	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts; 1 gene is lacking (RL6) and 1 gene is mutated (UL9); this mutation is also present in the original clinical specimen; most of the population contains an unannotated deletion extending from TRL into the left end of UL
JX512204.1_HAN16	235,112	Urine	Infant	-	2007	Germany	Passaged twice in human fibroblasts; 2 genes are mutated (US7 and US12)
JX512205.1_HAN19	235,810	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts; 1 gene is mutated (RL5A)
JX512206.1_HAN22	236,379	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts
JX512207.1_HAN28	236,017	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts
JX512208.1_HAN31	235,720	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts; 1 gene is mutated (UL1)
KF297339.1_TB40/E Lisa	237,683	Throat wash	BMT recipient	-	1999	Germany	Generated on human fibroblasts by passaging strain TB40/E once, plaque purifying three times, and passaging once more; 5 genes are mutated (RL13, UL128, IRS1, US1 and US2); US33A, US34, US34A and TRS1 are duplicated in place of IRS1, US1 and part of US2
KJ361946.1_2CEN2	235,360	BAL	-	-	2009	Germany	Passaged once in human fibroblasts; 2 genes are mutated (UL1 and UL9)
KJ361947.1_2CEN5	235,567	BAL	-	-	2009	Germany	Passaged once in human fibroblasts; 1 gene is mutated (RL5A)
KJ361948.1_2CEN15	234,949	BAL	-	-	-	Germany	Passaged once on human fibroblasts; genome sequence derived directly from clinical specimen is identical; 1 gene is mutated (UL97)

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KJ361949.1_2CEN30	236,168	BAL	-	-	-	Germany	Passaged once on human fibroblasts; genome sequence derived directly from clinical specimen is identical; 3 genes are mutated (UL1, UL9 and UL14)
KJ361950.1_HAN11	235,276	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts; 3 genes are mutated (RL5A, UL150 and US9)
KJ361951.1_HAN21	235,834	BAL	-	-	2006	Germany	Passaged 3 times in human fibroblasts
KJ361952.1_HAN27	235,861	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts; 2 genes are mutated (RL12 and UL133)
KJ361953.1_HAN30	235,483	BAL	-	-	2006	Germany	Passaged twice in human fibroblasts
KJ361954.1_HAN32	235,458	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts; 1 gene is mutated (UL9)
KJ361955.1_HAN33	235,512	BAL	-	-	2007	Germany	Passaged 3 times in human fibroblasts; 1 gene is mutated (UL9)
KJ361956.1_HAN36	234,844	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts; 1 gene is mutated (UL20)
KJ361957.1_HAN39	235,056	BAL	-	-	2007	Germany	Passaged once in human fibroblasts; 1 gene is mutated (UL1)
KJ361958.1_HAN40	235,763	BAL	-	-	2007	Germany	Passaged twice in human fibroblasts; 2 genes are mutated (RL1 and US13)
GU179288.1_U8	235,709	Urine	Congenitally infected infant	-	2003	Italy	Unpassaged strain in the original clinical sample
GU179289.1_VR1814	235,233	Cervical secretions	Pregnant woman with primary HCMV	-	1996	Italy	Unpassaged strain in the original clinical sample
GU179291.1_AF1	235,937	Amniotic fluid	-	-	2003	Italy	Unpassaged strain in the original clinical sample; 2 genes are mutated (RL6 and UL9)
KJ361959.1_PAV1	235,815	Amniotic fluid	-	-	2005	Italy	Sequenced directly from clinical specimen; 1 gene is mutated (RL5A)
KJ361960.1_PAV4	235,272	Amniotic fluid	-	-	2006	Italy	Sequenced directly from clinical specimen; 3 genes are mutated (RL5A, UL150 and US7)
KJ361961.1_PAV5	235,485	Amniotic fluid	-	-	2006	Italy	Sequenced directly from clinical specimen; 1 gene is mutated (US8)
KJ361962.1_PAV6	235,432	Amniotic fluid	-	-	2007	Italy	Sequenced directly from clinical specimen; 2 genes are mutated (RL1 and US7)
KJ361963.1_PAV7	235,142	Amniotic fluid	-	-	2007	Italy	Sequenced directly from clinical specimen; 1 gene is mutated (US7)
KJ361964.1_PAV8	235,432	Amniotic fluid	-	-	2007	Italy	Sequenced directly from clinical specimen; 1 gene is mutated (UL9)
KJ361965.1_PAV11	236,310	Amniotic fluid	-	-	-	Italy	Sequenced directly from clinical specimen
KJ361966.1_PAV12	235,616	Amniotic fluid	-	-	-	Italy	Sequenced directly from clinical specimen; 1 gene is mutated (RL5A)

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
KJ361967.1_PAV23	235,700	Amniotic fluid	-	-	2012	Italy	Sequenced directly from clinical specimen; 2 genes are mutated (RL1 and RL5A)
KJ361968.1_PAV24	235,361	Amniotic fluid	-	-	2012	Italy	Sequenced directly from clinical specimen; 1 gene is mutated (UL30)
KJ361969.1_PAV25	235,902	Amniotic fluid	-	-	2013	Italy	Sequenced directly from clinical specimen; 2 genes are mutated (UL9 and UL150)
KJ361970.1_PAV26	236,180	Amniotic fluid	-	-	2013	Italy	Sequenced directly from clinical specimen; 2 genes are mutated (UL1 and UL9)
KJ872539.1_PAV16	236,240	Amniotic fluid	-	-	2009	Italy	Sequenced directly from clinical specimen; sequenced from enriched viral DNA (SureSelect); 3 genes are mutated (RL5A, UL9 and UL111A)
KJ872540.1_PAV18	234,739	Amniotic fluid	-	-	2008	Italy	Sequenced directly from clinical specimen; sequenced from enriched viral DNA (SureSelect)
KJ872541.1_PAV20	236,293	Amniotic fluid	-	-	2013	Italy	Sequenced directly from clinical specimen; sequenced from enriched viral DNA (SureSelect); 1 gene is mutated (RL5A)
KJ872542.1_PAV21	235,901	Amniotic fluid	-	-	2013	Italy	Sequenced directly from clinical specimen; sequenced from enriched viral DNA (SureSelect); 2 genes are mutated (UL1 and UL9); 1 gene is mutated in a proportion of genomes (RNA4.9)
HQ380895.1_JHC	235,476	Blood	BMT recipient	-	2003	South Korea	-
GQ221974.1_3157	235,154	Urine	Congenitally infected infant	-	2001	United Kingdom	Sequenced at passage 3 in human fibroblasts; three genes are mutated (RL13, UL40 and UL128)
GQ221975.1_JP	236,375	Prostate tissue (post mortem)	AIDS patient	-	2001	United Kingdom	Sequenced directly from patient material, not passaged in cell culture; two genes are mutated (RL5A and UL111A)
GQ466044.1_3301	235,703	Urine	Congenitally infected infant	-	2001	United Kingdom	Sequenced directly from patient material; not passaged in cell culture
GU179290.1_U11	234,732	Urine	Congenitally infected infant	-	2003	United Kingdom	Unpassaged strain in the original clinical sample; 1 gene is mutated (UL9)
JX512197.1_6397	235,870	Urine	Congenitally infected infant	-	2001	United Kingdom	Passaged 3 times in human fibroblasts; 12 genes are mutated (RL13, UL9, UL130, UL131A, IRS1, US1, US2, US3, US6, US7, US8 and US9); region normally at left end of US replaced by inverted copy of region from right end of US, resulting in longer TRS/IRS; TRL/IRL is very short probably owing to deletion at UL junction
KJ361971.1_UKNEQAS1	235,190	Urine	Congenitally infected infant	-	2012	United Kingdom	Passaged twice in human fibroblasts; sequenced from enriched viral DNA (SureSelect); 1 gene is mutated (RL5A)
AY446894.2_Merlin	235,646	Urine	Congenitally infected child	-	1999	United Kingdom: Cardiff	Originally named strain 742; passaged 3 times in human fibroblasts; gene UL128 is mutated; populations of mutants in RL13 predominate, but the consensus sequence of this gene is not mutated;
FJ527563.1_AD169	231,781	Adenoids	Girl	7 years	1956	USA	Passaged in human fibroblasts; contains most of UL/b' region, which is absent from other AD169 substrains, and instead contains smaller deletion that affects genes UL144, UL142, UL141 and UL140; contains

Accession num._Strain	Size (bp)	Isolation Tissue	Source Description	Source Age	Date collected	Country	Notes
							mutations in genes RL5A, RL13, and UL131A; majority population contains deletion that affects IRS1, US1 and US2
FJ616285.1_Towne	235,147	Urine	Infant with microcephaly and hepatosplenomegaly	2 months	1970	USA	Passaged many times in human fibroblasts; contains mutations in genes RL13, UL1, UL40, UL130 and US1; majority population contains large deletion in UL/b' region associated with inverted duplication of sequence from near left genome end
GU937742.1_Toledo	235,398	Urine	Congenitally infected infant	-	1984	USA	Passaged several times in human fibroblasts; contains inversion of 14336 bp region (genes UL128 to UL133); 3 genes are mutated (RL13, UL9 and UL128)
JX512198.1_Davis	229,768	Liver biopsy	Congenitally infected infant	-	1957	USA	Passaged many times in human fibroblasts; 10 genes are mutated (RL5A, RL12, RL13, UL1, UL2, UL4, UL5, UL6, UL99 and UL130); a 472 bp region is deleted from the region between genes US34A and TRS1; the population possibly contains various unannotated deletions extending from TRL into the left end of UL
KF021605.1_TR	235,681	Vitreous humor	HIV-positive male	-	1996	USA	Passaged several times in human fibroblasts

Appendix 10. This table outlines key features of the 163 complete HCMV genome sequences in GenBank database release 211 that were used for reference in this thesis. BAL, Bronchoalveolar lavage; BMT, Bone marrow transplant; NPA, Nasopharyngeal aspirate; NPS, Nasopharyngeal swab.

APPENDIX 11: PARTIAL AND COMPLETE HCMV gO CDSs USED IN THIS STUDY.

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
1	AF531315.1_1176	1,389	-	-	-	-
2	AF531316.1_122	1,392	-	-	-	-
3	AF531317.1_1276	1,398	-	-	-	-
4	AF531318.1_1960	1,395	-	-	-	-
5	AF531319.1_279	1,392	-	-	-	-
6	AF531320.1_298	1,401	-	-	-	-
7	AF531321.1_3052	1,419	-	-	-	-
8	AF531322.1_4088	1,401	-	-	-	-
9	AF531323.1_452	1,374	-	-	-	-
10	AF531324.1_540	1,389	-	-	-	-
11	AF531325.1_6444	1,401	-	-	-	-
12	AF531326.1_650	1,392	-	-	-	-
13	AF531327.1_7614	1,395	-	-	-	-
14	AF531328.1_7868	1,398	-	-	-	-
15	AF531329.1_791	1,398	-	-	-	-
16	AF531330.1_851	1,401	-	-	-	-
17	AF531332.1_DM13	1,374	-	-	-	-
18	AF531333.1_DM2	1,389	-	-	-	-
19	AF531334.1_DM7	1,392	-	-	-	-
20	AF531335.1_DM8	1,398	-	-	-	-
21	AF531336.1_SW1758	1,389	-	-	-	-
22	AF531337.1_SW1	1,401	-	-	-	-
23	AF531338.1_SW1020	1,395	-	-	-	-
24	AF531339.1_SW1102	1,398	-	-	-	-
25	AF531340.1_SW1324	1,395	-	-	-	-
26	AF531341.1_SW1701	1,398	-	-	-	-

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
27	AF531342.1_SW1715	1,398	-	-	-	-
28	AF531343.1_1736	1,389	-	-	-	-
29	AF531344.1_SW1762	1,401	-	-	-	-
30	AF531345.1_SW2	1,392	-	-	-	-
31	AF531346.1_SW3	1,401	-	-	-	-
32	AF531347.1_SW4	1,401	-	-	-	-
33	AF531348.1_SW475	1,395	-	-	-	-
34	AF531349.1_SW490	1,389	-	-	-	-
35	AF531350.1_SW5	1,395	-	-	-	-
36	AF531351.1_SW6	1,395	-	-	-	-
37	AF531352.1_SW7	1,395	-	-	-	-
38	AF531353.1_SW8	1,398	-	-	-	-
39	AF531354.1_SW990	1,392	-	-	-	-
40	AY326961.1_Air	421	Mouth rinse	-	Malawi	-
41	AY326962.1_Aiu	430	Urine	-	Malawi	-
42	AY326963.1_Bir	433	Mouth rinse	-	Malawi	-
43	AY326964.1_Biu	433	Urine	-	Malawi	-
44	AY326965.1_B1u	433	Urine	-	Malawi	-
45	AY326966.1_B4r	433	Mouth rinse	-	Malawi	-
46	AY326967.1_B4u	421	Urine	-	Malawi	-
47	AY326968.1_Cir	433	Mouth rinse	-	Malawi	-
48	AY326969.1_Eir	421	Mouth rinse	-	Malawi	-
49	AY326970.1_E4u	421	Urine	-	Malawi	-
50	AY326971.1_Fiu	430	Urine	-	Malawi	-
51	AY326972.1_F1u	430	Urine	-	Malawi	-
52	AY326973.1_F2u	433	Urine	-	Malawi	-
53	AY326974.1_G3u	433	Urine	-	Malawi	-
54	AY326975.1_G3r	433	Mouth rinse	-	Malawi	-
55	AY326976.1_H3u	425	Urine	-	Malawi	-

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
56	AY326977.1_L1r	433	Mouth rinse	-	Malawi	-
57	AY326978.1_Mir	430	Mouth rinse	-	Malawi	-
58	AY326979.1_M1r	427	Mouth rinse	-	Malawi	-
59	AY326980.1_Riu	433	Urine	-	Malawi	-
60	AY326981.1_R2r	424	Mouth rinse	-	Malawi	-
61	AY326982.1_T3u	433	Urine	-	Malawi	-
62	AY326983.1_T4r	433	Mouth rinse	-	Malawi	-
63	AY326984.1_T4u	433	Urine	-	Malawi	-
64	AY326985.1_Xir	433	Mouth rinse	-	Malawi	-
65	AY326986.1_Zir	433	Mouth rinse	-	Malawi	-
66	AY582436.1_K1	460	-	-	United Kingdom	-
67	AY582437.1_A1	469	-	-	United Kingdom	-
68	AY582438.1_A2	469	-	-	United Kingdom	-
69	AY582439.1_B1	466	-	-	United Kingdom	-
70	AY582440.1_B2	466	-	-	United Kingdom	-
71	AY582441.1_B3	466	-	-	United Kingdom	-
72	AY582442.1_B4	466	-	-	United Kingdom	-
73	AY582443.1_B5	466	-	-	United Kingdom	-
74	AY582444.1_B6	466	-	-	United Kingdom	-
75	AY582445.1_B7a	466	-	-	United Kingdom	-
76	AY582447.1_C1a	463	-	-	United Kingdom	-
77	AY582448.1_C1b	460	-	-	United Kingdom	-
78	AY582449.1_C2	460	-	-	United Kingdom	-
79	AY582450.1_C3	460	-	-	United Kingdom	-
80	AY582451.1_C4a	460	-	-	United Kingdom	-
81	AY582453.1_D1	484	-	-	United Kingdom	-
82	AY582454.1_D2a	484	-	-	United Kingdom	-
83	AY582456.1_D3	484	-	-	United Kingdom	-
84	AY582457.1_D4	484	-	-	United Kingdom	-

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
85	AY582458.1_D5	484	-	-	United Kingdom	-
86	AY582459.1_D6	484	-	-	United Kingdom	-
87	AY582460.1_D7	484	-	-	United Kingdom	-
88	AY582461.1_D8a	469	-	-	United Kingdom	-
89	AY582463.1_D9a	484	-	-	United Kingdom	-
90	AY582464.1_D9b	469	-	-	United Kingdom	-
91	AY582465.1_E1	469	-	-	United Kingdom	-
92	AY582466.1_E2	469	-	-	United Kingdom	-
93	AY582467.1_E3	469	-	-	United Kingdom	-
94	AY582468.1_E4	469	-	-	United Kingdom	-
95	AY582469.1_F1	466	-	-	United Kingdom	-
96	AY582470.1_F2	466	-	-	United Kingdom	-
97	AY582471.1_F3	466	-	-	United Kingdom	-
98	AY582472.1_F4	466	-	-	United Kingdom	-
99	AY582473.1_G1a	466	-	-	United Kingdom	-
100	AY582475.1_G2a	484	-	-	United Kingdom	-
101	AY582477.1_G3	484	-	-	United Kingdom	-
102	AY582479.1_H3	484	-	-	United Kingdom	-
103	AY582480.1_I1	463	-	-	United Kingdom	-
104	AY582481.1_I5	463	-	-	United Kingdom	-
105	AY582482.1_J1	469	-	-	United Kingdom	-
106	EU348351.1_C177	1,392	-	-	Japan	-
107	EU348352.1_C83	1,392	-	-	Japan	-
108	EU348353.1_Y01	1,392	-	-	Japan	-
109	EU348354.1_FUK19U	1,389	-	-	Japan	-
110	EU348355.1_C164	1,398	-	-	Japan	-
111	EU348356.1_ASA1	1,419	-	-	Japan	-
112	EU348357.1_C95	1,419	-	-	Japan	-
113	EU348358.1_C149	1,419	-	-	Japan	-

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
114	EU348359.1_FUK28	1,419	-	-	Japan	-
115	EU348360.1_N62	1,419	-	-	Japan	-
116	EU348361.1_C154	1,401	-	-	Japan	-
117	EU348362.1_FUK16	1,401	-	-	Japan	-
118	EU348363.1_S01	1,401	-	-	Japan	-
119	EU348364.1_C49	1,395	-	-	Japan	-
120	EU377488.1_M5509/05	326	Clinical specimen	-	India	-
121	EU377490.1_VRF3053/06	326	Clinical specimen	-	India	-
122	EU686463.1_N5bb	445	-	-	Zambia	Genotype: gO1a
123	EU686464.1_N6bb	445	-	-	Zambia	Genotype: gO1a
124	EU686465.1_N14a	445	-	-	Zambia	Genotype: gO1a
125	EU686466.1_N17bb	445	-	-	Zambia	Genotype: gO1a
126	EU686467.1_N33bb	445	-	-	Zambia	Genotype: gO1a
127	EU686468.1_N2a	445	-	-	Zambia	Genotype: gO1b
128	EU686469.1_N9a	445	-	-	Zambia	Genotype: gO1b
129	EU686470.1_N15cc	445	-	-	Zambia	Genotype: gO1b
130	EU686471.1_N18bb	445	-	-	Zambia	Genotype: gO1b
131	EU686472.1_N19a	445	-	-	Zambia	Genotype: gO1b
132	EU686473.1_N23a	445	-	-	Zambia	Genotype: gO1b
133	EU686474.1_N24bb	445	-	-	Zambia	Genotype: gO1b
134	EU686475.1_N25a	445	-	-	Zambia	Genotype: gO1b
135	EU686476.1_N27bb	445	-	-	Zambia	Genotype: gO1b
136	EU686477.1_N34a	445	-	-	Zambia	Genotype: gO1b
137	EU686478.1_N36bb	445	-	-	Zambia	Genotype: gO1b
138	EU686479.1_N31a	439	-	-	Zambia	Genotype: gO1c
139	EU686480.1_N32a	439	-	-	Zambia	Genotype: gO1c
140	EU686481.1_N8bb	436	-	-	Zambia	Genotype: gO2a
141	EU686482.1_N29cc	436	-	-	Zambia	Genotype: gO2a
142	EU686483.1_N7bb	442	-	-	Zambia	Genotype: gO2b

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
143	EU686484.1_N10a	442	-	-	Zambia	Genotype: gO2b
144	EU686485.1_N11bb	442	-	-	Zambia	Genotype: gO2b
145	EU686486.1_N21a	442	-	-	Zambia	Genotype: gO2b
146	EU686487.1_N6a	448	-	-	Zambia	Genotype: gO3
147	EU686488.1_N7a	448	-	-	Zambia	Genotype: gO3
148	EU686489.1_N8a	448	-	-	Zambia	Genotype: gO3
149	EU686490.1_N11a	448	-	-	Zambia	Genotype: gO3
150	EU686491.1_N12a	448	-	-	Zambia	Genotype: gO3
151	EU686492.1_N17a	448	-	-	Zambia	Genotype: gO3
152	EU686493.1_N20bb	448	-	-	Zambia	Genotype: gO3
153	EU686494.1_N26a	448	-	-	Zambia	Genotype: gO3
154	EU686495.1_N27b	448	-	-	Zambia	Genotype: gO3
155	EU686496.1_N35bb	448	-	-	Zambia	Genotype: gO3
156	EU686497.1_N12cc	433	-	-	Zambia	Genotype: gO4
157	EU686498.1_N26bb	433	-	-	Zambia	Genotype: gO4
158	EU686499.1_N28a	433	-	-	Zambia	Genotype: gO4
159	EU686500.1_N33a	433	-	-	Zambia	Genotype: gO4
160	EU686501.1_N3aa	461	-	-	Zambia	Genotype: gO5
161	EU686502.1_N13bb	461	-	-	Zambia	Genotype: gO5
162	EU686503.1_N29b	461	-	-	Zambia	Genotype: gO5
163	EU686504.1_N30bb	461	-	-	Zambia	Genotype: gO5
164	EU686505.1_N35aa	461	-	-	Zambia	Genotype: gO5
165	EU686506.1_N36a	461	-	-	Zambia	Genotype: gO5
166	EU686507.1_37M18	445	-	14-Nov-06	Zambia	Genotype: gO1b
167	EU686508.1_40M6	445	-	21-Nov-05	Zambia	Genotype: gO1a
168	EU686509.1_263M6	445	-	30-Aug-06	Zambia	Genotype: gO1b
169	EU686510.1_178M18	436	-	05-Jul-07	Zambia	Genotype: gO2a
170	EU686511.1_K60	445	-	-	Zambia	Genotype: gO1a
171	EU686512.1_K33	445	-	-	Zambia	Genotype: gO1b

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
172	EU686513.1_K57	445	-	-	Zambia	Genotype: gO1b
173	EU686514.1_K67	445	-	-	Zambia	Genotype: gO1b
174	EU686515.1_K86	445	-	-	Zambia	Genotype: gO1b
175	EU686516.1_K142	445	-	-	Zambia	Genotype: gO1b
176	EU686517.1_K61	442	-	-	Zambia	Genotype: gO2b
177	EU686518.1_K141	461	-	-	Zambia	Genotype: gO5
178	GQ227753.1_HAN15	2,843	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts; acronym: HCMV; acronym: HHV-5
179	GQ227755.1_HAN17	2,813	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts; acronym: HCMV; acronym: HHV-5
180	GQ227760.1_HAN23	2,813	Bronchoalveolar lavage	2007	Germany	Passaged 3 times in human fibroblasts; acronym: HCMV; acronym: HHV-5
181	GQ227761.1_HAN24	2,841	Bronchoalveolar lavage	2007	Germany	Passaged 3 times in human fibroblasts; acronym: HCMV; acronym: HHV-5
182	GQ227762.1_HAN25	2,837	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts; acronym: HCMV; acronym: HHV-5
183	GQ227765.1_HAN29	2,841	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts; acronym: HCMV; acronym: HHV-5
184	GQ227770.1_HAN34	2,841	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts; acronym: HCMV; acronym: HHV-5
185	GQ227780.1_U4	2,859	Urine from a congenitally infected infant	2003	United Kingdom	Passaged 10 times in human fibroblasts; acronym: HCMV; acronym: HHV-5
186	GQ227785.1_W1	2,852	Post mortem lung tissue from an AIDS patient	2001	United Kingdom	Sequenced directly from patient material, not passaged in cell culture; two strains (W1 and W2) were detected in the same material; acronym: HCMV; acronym: HHV-5
187	GQ227786.1_W2	2,843	Post mortem lung tissue from an AIDS patient	2001	United Kingdom	Sequenced directly from patient material, not passaged in cell culture; two strains (W1 and W2) were detected in the same material; acronym: HCMV; acronym: HHV-5
188	KT987996.1_human-wt/ITA/PR184/2010	267	Urine	2010	Italy	Genotype: gO1a
189	KT987997.1_human-wt/ITA/PR1343/1995	267	Urine	1995	Italy	Genotype: gO1c
190	KT987998.1_human-wt/ITA/PR817/1998	267	Urine	1998	Italy	Genotype: gO2b
191	KT987999.1_human-wt/ITA/PR217/1997	267	Urine	1997	Italy	Genotype: gO2b
192	KT988000.1_human-wt/ITA/PR206/1995	267	Urine	1995	Italy	Genotype: gO3
193	KT988001.1_human-wt/ITA/PR2197/2012	267	Urine	2012	Italy	Genotype: gO3

	Accession num._Strain/Isolate	Size (bp)	Isolation source	Date collected	Country	Notes
194	KT988002.1_human-wt/ITA/PR2115/2001	267	Urine	2001	Italy	Genotype: gO4
195	KT988003.1_human-wt/ITA/PR2528/1998	267	Urine	1998	Italy	Genotype: gO4

Appendix 11. This table depicts key features of the 195 partial and complete HCMV gO CDSs sequences available in GenBank release 211 that were analysed in this study.

APPENDIX 12: PARTIAL AND COMPLETE HCMV gN CDSs USED IN THIS STUDY.

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
1	AF224679.1_E1	-	-	-	-
2	AF224680.1_E4	-	-	-	-
3	AF224681.1_E5	-	-	-	-
4	AF224682.1_E8	-	-	-	-
5	AF224683.1_E10	-	-	-	-
6	AF224684.1_E11	-	-	-	-
7	AF224685.1_E13	-	-	-	-
8	AF224686.1_E14	-	-	-	-
9	AF224687.1_E15	-	-	-	-
10	AF224688.1_E16	-	-	-	-
11	AF309969.1_LN	-	-	-	-
12	AF309970.1_GR	-	-	-	-
13	AF309971.1_DB	-	-	-	-
14	AF309972.1_SB	-	-	-	-
15	AF309973.1_HDs	-	-	-	-
16	AF309974.1_HR	-	-	-	-
17	AF309975.1_DL	-	-	-	-
18	AF309976.1_Can2	-	-	-	-
19	AF309977.1_Can4	-	-	-	-
20	AF309978.1_Can7	-	-	-	-
21	AF309979.1_Can8	-	-	-	-
22	AF309980.1_BD	-	-	-	-
23	AF309981.1_ML	-	-	-	-
24	AF309982.1_TS	-	-	-	-
25	AF309983.1_FL	-	-	-	-
26	AF309984.1_B-AL	-	-	-	-

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
27	AF309985.1_B-AS	-	-	-	-
28	AF309986.1_PS	-	-	-	-
29	AF309987.1_MS	-	-	-	-
30	AF309988.1_Can10	-	-	-	-
31	AF309989.1_BO	-	-	-	-
32	AF309990.1_LC	-	-	-	-
33	AF309991.1_BC	-	-	-	-
34	AF309992.1_ZV	-	-	-	-
35	AF309993.1_HDu	-	-	-	-
36	AF309994.1_SR	-	-	-	-
37	AF309995.1_RL	-	-	-	-
38	AF309996.1_SE	-	-	-	-
39	AF309997.1_BR	-	-	-	-
40	AF309998.1_LV	-	-	-	-
41	AF309999.1_CR	-	-	-	-
42	AF310000.1_GJ	-	-	-	-
43	AF310001.1_RC	-	-	-	-
44	AF310002.1_PN	-	-	-	-
45	AF310003.1_VL	-	-	-	-
46	AF310004.1_MN	-	-	-	-
47	AF310005.1_BN	-	-	-	-
48	AF310006.1_PM	-	-	-	-
49	AF390748.1_CER	-	-	-	Clinical isolate
50	AF390749.1_FR	-	-	-	Clinical isolate
51	AF390750.1_LA10	-	-	-	Clinical isolate
52	AF390751.1_LA12	-	-	-	Clinical isolate
53	AF390752.1_LA14	-	-	-	Clinical isolate
54	AF390753.1_LA15	-	-	-	Clinical isolate
55	AF390754.1_LA16	-	-	-	Clinical isolate

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
56	AF390755.1_LA19	-	-	-	Clinical isolate
57	AF390756.1_LA2	-	-	-	Clinical isolate
58	AF390757.1_LA20	-	-	-	Clinical isolate
59	AF390758.1_LA21	-	-	-	Clinical isolate
60	AF390759.1_LA22	-	-	-	Clinical isolate
61	AF390760.1_LA23	-	-	-	Clinical isolate
62	AF390761.1_LA25	-	-	-	Clinical isolate
63	AF390762.1_LA28	-	-	-	Clinical isolate
64	AF390763.1_LA33	-	-	-	Clinical isolate
65	AF390764.1_LA34	-	-	-	Clinical isolate
66	AF390765.1_LA38	-	-	-	Clinical isolate
67	AF390766.1_LA5	-	-	-	Clinical isolate
68	AF390767.1_LA9	-	-	-	Clinical isolate
69	AF390768.1_LA7	-	-	-	Clinical isolate
70	AF390769.1_LA8	-	-	-	Clinical isolate
71	AF390770.1_LS	-	-	-	Clinical isolate
72	AF390771.1_m23	-	-	-	Clinical isolate
73	AF390772.1_PL	-	-	-	Clinical isolate
74	AF390773.1_RZ	-	-	-	Clinical isolate
75	AF390774.1_SA	-	-	-	Clinical isolate
76	AF390775.1_SH	-	-	-	Clinical isolate
77	AF390776.1_SV	-	-	-	Clinical isolate
78	AF390777.1_RV	-	-	-	Clinical isolate
79	AF390778.1_SF	-	-	-	Clinical isolate
80	AF390779.1_DEL	-	-	-	Clinical isolate
81	AF390780.1_BRT	-	-	-	Clinical isolate
82	AF390781.1_CU	-	-	-	Clinical isolate
83	AF390782.1_FC	-	-	-	Clinical isolate
84	AF390783.1_GL	-	-	-	Clinical isolate

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
85	AF390784.1_KA	-	-	-	Clinical isolate
86	AF390785.1_MD	-	-	-	Clinical isolate
87	AF390786.1_MR	-	-	-	Clinical isolate
88	AF390787.1_SG	-	-	-	Clinical isolate
89	AF390788.1_SM	-	-	-	Clinical isolate
90	AF390789.1_ST	-	-	-	Clinical isolate
91	AF390790.1_VT	-	-	-	Clinical isolate
92	AF390791.1_GS	-	-	-	Clinical isolate
93	AF390792.1_RO	-	-	-	Clinical isolate
94	AF390793.1_RF	-	-	-	Clinical isolate
95	AF390794.1_GAR	-	-	-	Clinical isolate
96	AF390795.1_PR	-	-	-	Clinical isolate
97	AF390796.1_TR	-	-	-	Clinical isolate
98	AF390797.1_A1-36A	-	-	-	-
99	AF390798.1_A3-90B	-	-	-	-
100	AF390799.1_A4-15F	-	-	-	-
101	AF390800.1_A5-14B	-	-	-	-
102	AF390801.1_A7-45E	-	-	-	-
103	AF390802.1_A8-27F	-	-	-	-
104	AF390803.1_A9-11F	-	-	-	-
105	AF390804.1_A11-79A+	-	-	-	-
106	AF390805.1_A12-35A	-	-	-	-
107	AF390806.1_A16-28B	-	-	-	-
108	AF390807.1_A17-30A	-	-	-	-
109	AF390808.1_A18-16B	-	-	-	-
110	AF390809.1_A20-17F	-	-	-	-
111	AF390810.1_A22-19F	-	-	-	-
112	AF390811.1_A23-20F	-	-	-	-
113	AF390812.1_A24-92A	-	-	-	-

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
114	AF390813.1_A25-17A	-	-	-	-
115	AF390814.1_A26-41A	-	-	-	-
116	AF390815.1_A27-38A	-	-	-	-
117	AF390816.1_A28-99A	-	-	-	-
118	AF390817.1_A29-37C	-	-	-	-
119	AF390818.1_A30-48B	-	-	-	-
120	AF390819.1_A32-96A	-	-	-	-
121	AF390820.1_A33-70A	-	-	-	-
122	AF390821.1_A34-49A	-	-	-	-
123	AF390822.1_A35-55A	-	-	-	-
124	AF390823.1_A37-6E	-	-	-	-
125	AF390824.1_A10-11B	-	-	-	-
126	AF390825.1_A13-73D	-	-	-	-
127	AF390826.1_A14-84A+	-	-	-	-
128	AF390827.1_A15-96D	-	-	-	-
129	AF390828.1_A2-28A	-	-	-	-
130	AF390829.1_A21-18F	-	-	-	-
131	AF390830.1_A31-64A	-	-	-	-
132	AF390831.1_A36-3E	-	-	-	-
133	AF390832.1_A39-19E*	-	-	-	-
134	AF390833.1_A6-45B	-	-	-	-
135	AF390834.1_A40-5A	-	-	-	-
136	AF390835.2_P143	-	-	-	-
137	AF390836.2_209	-	-	-	-
138	AF390837.2_33spa	-	-	-	-
139	AF390838.2_36YUEK	-	-	-	-
140	AF390840.2_210	-	-	-	-
141	AF390841.2_236	-	-	-	-
142	AF390842.2_244	-	-	-	-

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
143	AF390843.2_11PAN	-	-	-	-
144	AF390844.2_20FI	-	-	-	-
145	AF390845.2_7RO	-	-	-	-
146	AF390846.2_7330ET	-	-	-	-
147	AF390847.2_5929C	-	-	-	-
148	AF390848.1_C13	-	-	-	-
149	AF390849.1_C16	-	-	-	-
150	AF390850.1_C22	-	-	-	-
151	AF390851.1_C35	-	-	-	-
152	AF390852.1_C7	-	-	-	-
153	AF390853.1_C9	-	-	-	-
154	AF390854.1_C69	-	-	-	-
155	AF390855.1_C14	-	-	-	-
156	AF390856.1_C32	-	-	-	-
157	AF390857.1_C33	-	-	-	-
158	AF395118.1_104M	-	-	-	Clinical isolate
159	AF395119.1_109B	-	-	-	Clinical isolate
160	AF395120.1_13B	-	-	-	Clinical isolate
161	AF395121.1_2B	-	-	-	Clinical isolate
162	AF395122.1_Vp293	-	-	-	Clinical isolate
163	AF396713.1_USA1	-	-	-	Clinical isolate
164	AF396714.1_USA2	-	-	-	Clinical isolate
165	AF396715.1_USA3	-	-	-	Clinical isolate
166	AF396716.1_USA4	-	-	-	Clinical isolate
167	AF396717.1_USA5	-	-	-	Clinical isolate
168	AF396718.1_USA6	-	-	-	Clinical isolate
169	AF396719.1_USA7	-	-	-	Clinical isolate
170	AF396720.1_USA8	-	-	-	Clinical isolate
171	AF396721.1_USA9	-	-	-	Clinical isolate

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
172	AF396722.1_BH	-	-	-	Clinical isolate
173	AF396723.1_BL	-	-	-	Clinical isolate
174	AF396724.1_CL	-	-	-	Clinical isolate
175	AF396725.1_CN	-	-	-	Clinical isolate
176	AF396726.1_CS	-	-	-	Clinical isolate
177	AF396727.1_DEL	-	-	-	Clinical isolate
178	AF396728.1_DM	-	-	-	Clinical isolate
179	AF396729.1_LA22	-	-	-	Clinical isolate
180	AF396730.1_ME	-	-	-	Clinical isolate
181	AF396731.1_MS	-	-	-	Clinical isolate
182	AF396732.1_NT	-	-	-	Clinical isolate
183	AF396733.1_RL	-	-	-	Clinical isolate
184	AF396734.1_SCG	-	-	-	Clinical isolate
185	AF396735.1_SC	-	-	-	Clinical isolate
186	AF396736.1_SN	-	-	-	Clinical isolate
187	AF396737.1_SP	-	-	-	Clinical isolate
188	AF396738.1_TD	-	-	-	Clinical isolate
189	AF396739.1_TM	-	-	-	Clinical isolate
190	AF396740.1_ZN	-	-	-	Clinical isolate
191	AF396741.1_MCV	-	-	-	Clinical isolate
192	AF396742.1_MR	-	-	-	Clinical isolate
193	AF396743.1_CLZ	-	-	-	Clinical isolate
194	AF396744.1_DLG	-	-	-	Clinical isolate
195	AF396745.1_DG	-	-	-	Clinical isolate
196	AY326987.1_A2u	Urine	-	Malawi	-
197	AY326988.1_Bir	Mouth rinse	-	Malawi	-
198	AY326989.1_Biu	Urine	-	Malawi	-
199	AY326990.1_B1r	Mouth rinse	-	Malawi	-
200	AY326991.1_B1u	Urine	-	Malawi	-

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
201	AY326992.1_B4r	Mouth rinse	-	Malawi	-
202	AY326993.1_B4u	Urine	-	Malawi	-
203	AY326994.1_Cir	Mouth rinse	-	Malawi	-
204	AY326995.1_Dir	Mouth rinse	-	Malawi	-
205	AY326996.1_Eir	Mouth rinse	-	Malawi	-
206	AY326997.1_E2r	Mouth rinse	-	Malawi	-
207	AY326998.1_E3u	Urine	-	Malawi	-
208	AY326999.1_E4u	Urine	-	Malawi	-
209	AY327000.1_F2r	Mouth rinse	-	Malawi	-
210	AY327001.1_F1u	Urine	-	Malawi	-
211	AY327002.1_F2u	Urine	-	Malawi	-
212	AY327003.1_Gir	Mouth rinse	-	Malawi	-
213	AY327004.1_G3r	Mouth rinse	-	Malawi	-
214	AY327005.1_G3u	Urine	-	Malawi	-
215	AY327006.1_I2u	Urine	-	Malawi	-
216	AY327007.1_M3u	Urine	-	Malawi	-
217	AY327008.1_Nir	Mouth rinse	-	Malawi	-
218	AY327009.1_Niu	Urine	-	Malawi	-
219	AY327010.1_Piu	Urine	-	Malawi	-
220	AY327011.1_Riu	Urine	-	Malawi	-
221	AY327012.1_R1r	Mouth rinse	-	Malawi	-
222	AY327013.1_R2r	Mouth rinse	-	Malawi	-
223	AY327014.1_R2u	Urine	-	Malawi	-
224	AY327015.1_R3u	Urine	-	Malawi	-
225	AY327016.1_Tir	Mouth rinse	-	Malawi	-
226	AY327017.1_T3u	Urine	-	Malawi	-
227	AY327018.1_T4r	Mouth rinse	-	Malawi	-
228	AY327019.1_T4u	Urine	-	Malawi	-
229	AY327020.1_Qiu	Urine	-	Malawi	-

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
230	AY327021.1_Uir	Mouth rinse	-	Malawi	-
231	AY327022.1_U1r	Mouth rinse	-	Malawi	-
232	AY327023.1_U3r	Mouth rinse	-	Malawi	-
233	AY327024.1_Xir	Mouth rinse	-	Malawi	-
234	AY327025.1_X2r	Mouth rinse	-	Malawi	-
235	AY327026.1_X3r	Mouth rinse	-	Malawi	-
236	AY327027.1_W4r	Mouth rinse	-	Malawi	-
237	AY327028.1_W6r	Mouth rinse	-	Malawi	-
238	AY327029.1_Zir	Mouth rinse	-	Malawi	-
239	AY327030.1_Ziu	Urine	-	Malawi	-
240	AY327031.1_Rir	Mouth rinse	-	Malawi	-
241	AY327032.1_E4r	Mouth rinse	-	Malawi	-
242	EU377503.1_M449/05	Clinical specimen	-	India	-
243	EU377505.1_VRF3754/06	Clinical specimen	-	India	-
244	EU377506.1_VRF2054/07	Clinical specimen	-	India	-
245	EU377509.1_VRF3053/06	Clinical specimen	-	India	-
246	EU377510.1_VRF2758/07	Clinical specimen	-	India	-
247	EU686416.1_N5a	-	-	Zambia	Genotype:gN1
248	EU686417.1_N8c	-	-	Zambia	Genotype:gN1
249	EU686418.1_N17a	-	-	Zambia	Genotype:gN1
250	EU686419.1_N33a	-	-	Zambia	Genotype:gN1
251	EU686420.1_N7c	-	-	Zambia	Genotype:gN2
252	EU686421.1_N10c	-	-	Zambia	Genotype:gN2
253	EU686422.1_N21b	-	-	Zambia	Genotype:gN2
254	EU686423.1_N2b	-	-	Zambia	Genotype:gN3a
255	EU686424.1_N9a	-	-	Zambia	Genotype:gN3a
256	EU686425.1_N10a	-	-	Zambia	Genotype:gN3a
257	EU686426.1_N19a	-	-	Zambia	Genotype:gN3a
258	EU686427.1_N23a	-	-	Zambia	Genotype:gN3a

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
259	EU686428.1_N25a	-	-	Zambia	Genotype:gN3a
260	EU686429.1_N34a	-	-	Zambia	Genotype:gN3a
261	EU686430.1_N8a	-	-	Zambia	Genotype:gN3b
262	EU686431.1_N6a	-	-	Zambia	Genotype:gN4a
263	EU686432.1_N11a	-	-	Zambia	Genotype:gN4a
264	EU686433.1_N21a	-	-	Zambia	Genotype:gN4b
265	EU686434.1_N27b	-	-	Zambia	Genotype:gN4b
266	EU686435.1_N28a	-	-	Zambia	Genotype:gN4b
267	EU686436.1_N31a	-	-	Zambia	Genotype:gN4c
268	EU686437.1_N32a	-	-	Zambia	Genotype:gN4c
269	EU686438.1_N1c	-	-	Zambia	Genotype:gN4d
270	EU686439.1_N12c	-	-	Zambia	Genotype:gN4d
271	EU686440.1_N18a	-	-	Zambia	Genotype:gN4d
272	EU686441.1_N24a	-	-	Zambia	Genotype:gN4d
273	EU686442.1_N29a	-	-	Zambia	Genotype:gN4d
274	EU686443.1_N30a	-	-	Zambia	Genotype:gN4d
275	EU686444.1_N36a	-	-	Zambia	Genotype:gN4d
276	EU686445.1_35M6	-	14-Nov-05	Zambia	Genotype:gN3a
277	EU686446.1_133M18	-	30-Mar-07	Zambia	Genotype:gN3a
278	EU686447.1_26M18	-	07-Nov-06	Zambia	Genotype:gN3a
279	EU686448.1_263M6	-	30-Aug-06	Zambia	Genotype:gN3a
280	EU686449.1_492M6	-	15-Jun-07	Zambia	Genotype:gN3a
281	EU686450.1_503M6	-	21-Jun-07	Zambia	Genotype:gN3a
282	EU686451.1_K68	-	-	Zambia	Genotype:gN1
283	EU686452.1_K33	-	-	Zambia	Genotype:gN3a
284	EU686453.1_K57	-	-	Zambia	Genotype:gN4a
285	EU686454.1_K60	-	-	Zambia	Genotype:gN1
286	EU686455.1_K135	-	-	Zambia	Genotype:gN3a
287	EU686456.1_K141	-	-	Zambia	Genotype:gN4d

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
288	EU686457.1_K152	-	-	Zambia	Genotype:gN3a
289	EU686458.1_K103	-	-	Zambia	Genotype:gN1
290	EU686459.1_K110	-	-	Zambia	Genotype:gN1
291	EU686460.1_K61	-	-	Zambia	Genotype:gN1
292	EU686461.1_K190	-	-	Zambia	Genotype:gN1
293	EU686462.1_K137	-	-	Zambia	Genotype:gN1
294	GQ227753.1_HAN15	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts
295	GQ227755.1_HAN17	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts
296	GQ227760.1_HAN23	Bronchoalveolar lavage	2007	Germany	Passaged 3 times in human fibroblasts
297	GQ227761.1_HAN24	Bronchoalveolar lavage	2007	Germany	Passaged 3 times in human fibroblasts
298	GQ227762.1_HAN25	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts
299	GQ227765.1_HAN29	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts
300	GQ227770.1_HAN34	Bronchoalveolar lavage	2007	Germany	Passaged twice in human fibroblasts
301	GQ227780.1_U4	Urine from a congenitally infected infant	2003	United Kingdom	Passaged 10 times in human fibroblasts
302	GQ227785.1_W1	Post mortem lung tissue from an AIDS patient	2001	United Kingdom	Sequenced directly from patient material, not passaged in cell culture; two strains (W1 and W2) were detected in the same material
303	GQ227786.1_W2	Post mortem lung tissue from an AIDS patient	2001	United Kingdom	Sequenced directly from patient material, not passaged in cell culture; two strains (W1 and W2) were detected in the same material
304	GU376720.1_N3201	Plasma from bone marrow recipient	27-Jul-09	China	Genotype:gN3b
305	GU376721.1_N1802	Plasma from bone marrow recipient	23-Apr-09	China	Genotype:gN4a
306	GU376722.1_N3403	Plasma from bone marrow recipient	02-Jul-09	China	Genotype:gN4a
307	GU376723.1_N2701	Plasma from bone marrow recipient	20-Apr-09	China	Genotype:gN4b
308	GU376724.1_N1503	Plasma from bone marrow recipient	13-Apr-09	China	Genotype:gN4d
309	GU376725.1_N1507	Plasma from bone marrow recipient	13-Apr-09	China	Genotype:gN2
310	GU376726.1_N3405	Plasma from bone marrow recipient	02-Jul-09	China	Genotype:gN3a
311	GU376727.1_N1809	Plasma from bone marrow recipient	23-Apr-09	China	Genotype:gN4b
312	GU441773.1_N3408	Plasma from bone marrow recipient	02-Jul-09	China	Genotype:gN1
313	GU583628.1_102F	-	11-Aug-08	India	-
314	GU583629.1_103F	-	19-May-07	India	Genotype:1
315	GU583630.1_105F	-	12-May-08	India	-

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
316	GU583631.1_110F	-	19-Mar-08	India	Genotype:1
317	GU583632.1_111F	-	30-Oct-07	India	Genotype:4b
318	GU583633.1_113F	-	11-Sep-07	India	Genotype:4c
319	GU583634.1_114F	-	13-Sep-07	India	Genotype:4b
320	GU583635.1_115F	-	19-Mar-07	India	-
321	GU583636.1_223F	-	02-May-08	India	-
322	GU583637.1_261F	-	28-May-08	India	-
323	GU583638.1_293F	-	28-Jul-07	India	Genotype:4b
324	GU583639.1_323F	-	16-Oct-07	India	Genotype:4c
325	GU583640.1_459F	-	06-Aug-07	India	Genotype:4b
326	GU583641.1_1011F	-	28-Aug-06	India	Genotype:4b
327	GU583642.1_1022F	-	04-Mar-08	India	-
328	GU583643.1_1095F	-	27-Dec-08	India	Genotype:4c
329	GU583644.1_1193F	-	12-Dec-07	India	Genotype:4a
330	GU583645.1_1194F	-	28-Dec-07	India	-
331	GU583646.1_1210F	-	13-Jun-07	India	-
332	GU583647.1_1225F	-	03-Sep-07	India	-
333	GU647095.1_N1592	Plasma from bone marrow recipient	30-Jul-09	China	Genotype:gN4c
334	KF875976.1_1	Inflammatory breast cancer tissue	20-Aug-13	Egypt	Genotype:gN1
335	KF875977.1_2	Non-inflammatory breast cancer tissue	20-Aug-13	Egypt	Genotype:gN1
336	KF875978.1_3	Inflammatory breast cancer tissue	24-Oct-13	Egypt	Genotype:gN2
337	KF875979.1_4	Non-inflammatory breast cancer tissue	24-Oct-13	Egypt	Genotype:gN2
338	KF875980.1_5	Inflammatory breast cancer tissue	13-Jun-13	Egypt	Genotype:gN3
339	KF875981.1_6	Non-inflammatory breast cancer tissue	13-Jun-13	Egypt	Genotype:gN3
340	KF875982.1_7	Inflammatory breast cancer tissue	08-Jul-13	Egypt	Genotype:gN4
341	KF875983.1_8	Non-inflammatory breast cancer tissue	08-Jul-13	Egypt	Genotype:gN4
342	KR261654.1_Ah14	HIV1-infected patient	04-Dec-14	China	Genotype:gN4b
343	KR992948.1_1C	Urine	17-Jan-03	Spain	H12VL:CMV2;genotype:1
344	KR992949.1_1cCMV	Urine	29-Jan-03	Spain	H12VL:CMV1;genotype:1

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
345	KR992950.1_2C	Urine	20-Jan-03	Spain	H12VL:CMV4;genotype:1
346	KR992951.1_2cCMV	Urine	25-Feb-03	Spain	H12VL:CMV3;genotype:3a
347	KR992952.1_3C	Urine	10-Apr-03	Spain	H12VL:CMV6;genotype:1
348	KR992953.1_3cCMV	Urine	07-Jan-05	Spain	H12VL:CMV5;genotype:4c
349	KR992954.1_4cCMV	Foetal blood	03-Feb-06	Spain	H12VL:CMV7;genotype:1
350	KR992955.1_5C	Urine	23-Sep-04	Spain	H12VL:CMV8;genotype:4b
351	KR992956.1_6C	Urine	20-Sep-05	Spain	H12VL:CMV10;genotype:1
352	KR992957.1_6cCMV	Amniotic fluid	20-Oct-06	Spain	H12VL:CMV9;genotype:4a
353	KR992958.1_7C	Urine	05-Jan-06	Spain	H12VL:CMV12;genotype:2
354	KR992959.1_7cCMV	Urine	27-Mar-07	Spain	H12VL:CMV11;genotype:4a
355	KR992960.1_8cCMV	Urine	04-May-07	Spain	H12VL:CMV13;genotype:3a
356	KR992961.1_9C	Urine	04-Apr-06	Spain	H12VL:CMV15;genotype:4c
357	KR992962.1_9cCMV	Foetal blood	05-Jun-07	Spain	H12VL:CMV14;genotype:4a
358	KR992963.1_10C	Urine	08-May-06	Spain	H12VL:CMV16;genotype:4b
359	KR992964.1_11C	Urine	09-Jan-07	Spain	H12VL:CMV17;genotype:4b
360	KR992965.1_12C	Urine	11-Jan-07	Spain	H12VL:CMV18;genotype:4c
361	KR992966.1_13C	Urine	31-Jan-07	Spain	H12VL:CMV20;genotype:3a
362	KR992967.1_13cCMV	Amniotic fluid	05-Nov-07	Spain	H12VL:CMV19;genotype:1
363	KR992968.1_14cCMV	Amniotic fluid	12-Feb-08	Spain	H12VL:CMV21;genotype:1
364	KR992969.1_15C	Urine	23-Mar-07	Spain	H12VL:CMV23;genotype:4c
365	KR992970.1_15cCMV	Dried blood spot	23-Feb-10	Spain	H12VL:CMV22;genotype:1
366	KR992971.1_16C	Urine	08-Oct-08	Spain	H12VL:CMV25;genotype:3b
367	KR992972.1_16cCMV	Urine	12-Apr-10	Spain	H12VL:CMV24;genotype:2
368	KR992973.1_17cCMV	Foetal blood	13-Apr-10	Spain	H12VL:CMV26;genotype:4c
369	KR992974.1_19C	Urine	09-Mar-09	Spain	H12VL:CMV28;genotype:3b
370	KR992975.1_19cCMV	Urine	22-Feb-11	Spain	H12VL:CMV27;genotype:3a
371	KR992976.1_20cCMV	Urine	15-Apr-11	Spain	H12VL:CMV29;genotype:2
372	KR992977.1_21C	Urine	27-Oct-09	Spain	H12VL:CMV31;genotype:3a
373	KR992978.1_21cCMV	Urine	01-Jul-11	Spain	H12VL:CMV30;genotype:4c

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
374	KR992979.1_22C	Urine	25-Nov-09	Spain	H12VL:CMV32;genotype:3a
375	KR992980.1_23C	Urine	20-Mar-10	Spain	H12VL:CMV34;genotype:3a
376	KR992981.1_23cCMV	Urine	06-Sep-11	Spain	H12VL:CMV33;genotype:4a
377	KR992982.1_24C	Urine	23-Mar-10	Spain	H12VL:CMV36;genotype:4a
378	KR992983.1_24cCMV	Foetal blood	13-Dec-11	Spain	H12VL:CMV35;genotype:1
379	KR992984.1_25C	Urine	24-May-10	Spain	H12VL:CMV38;genotype:3b
380	KR992985.1_25cCMV	Foetal blood	17-Jan-12	Spain	H12VL:CMV37;genotype:4a
381	KR992986.1_26cCMV	Urine	31-Jan-12	Spain	H12VL:CMV39;genotype:4c
382	KR992987.1_27C	Urine	22-Jul-10	Spain	H12VL:CMV41;genotype:3a
383	KR992988.1_27cCMV	Urine	30-Mar-05	Spain	H12VL:CMV40;genotype:4a
384	KR992989.1_28C	Urine	10-Aug-10	Spain	H12VL:CMV42;genotype:4b
385	KR992990.1_29C	Urine	15-Sep-10	Spain	H12VL:CMV43;genotype:1
386	KR992991.1_30C	Urine	25-Oct-10	Spain	H12VL:CMV44;genotype:3a
387	KR992992.1_31C	Urine	05-Nov-10	Spain	H12VL:CMV46;genotype:4c
388	KR992993.1_31cCMV	Urine	18-Oct-07	Spain	H12VL:CMV45;genotype:1
389	KR992994.1_32cCMV	Amniotic fluid	01-Dec-09	Spain	H12VL:CMV47;genotype:1
390	KR992995.1_33C	Urine	11-Jan-11	Spain	H12VL:CMV49;genotype:2
391	KR992996.1_33cCMV	Dried blood spot	17-Dec-09	Spain	H12VL:CMV48;genotype:1
392	KR992997.1_34C	Urine	01-Mar-11	Spain	H12VL:CMV51;genotype:3b
393	KR992998.1_34cCMV	Urine	19-Jan-11	Spain	H12VL:CMV50;genotype:2
394	KR992999.1_35cCMV	Foetal blood	07-Mar-11	Spain	H12VL:CMV52;genotype:3b
395	KR993000.1_35C	Urine	28-Mar-11	Spain	H12VL:CMV53;genotype:1
396	KR993001.1_36C	Urine	24-Jun-11	Spain	H12VL:CMV55;genotype:4c
397	KR993002.1_36cCMV	Dried blood spot	10-Mar-11	Spain	H12VL:CMV54;genotype:2
398	KR993003.1_37C	Urine	24-Jun-11	Spain	H12VL:CMV56;genotype:4a
399	KR993004.1_38C	Urine	28-Nov-11	Spain	H12VL:CMV58;genotype:4c
400	KR993005.1_38cCMV	Saliva	13-Jan-12	Spain	H12VL:CMV57;genotype:3a
401	KR993006.1_39cCMV	Dried blood spot	22-Jun-10	Spain	H12VL:CMV59;genotype:2
402	KR993007.1_39C	Urine	25-Jan-12	Spain	H12VL:CMV60;genotype:1

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
403	KR993008.1_40C	Urine	31-Aug-12	Spain	H12VL:CMV61;genotype:1
404	KR993009.1_41C	Urine	01-Oct-12	Spain	H12VL:CMV62;genotype:4c
405	KR993010.1_42C	Urine	11-Nov-13	Spain	H12VL:CMV63;genotype:3b
406	KR993011.1_43C	Urine	28-Nov-13	Spain	H12VL:CMV64;genotype:3b
407	KR993012.1_44C	Urine	31-Jan-14	Spain	H12VL:CMV65;genotype:2
408	KR993013.1_45C	Urine	14-Apr-14	Spain	H12VL:CMV66;genotype:4a
409	KR993014.1_46C	Urine	09-Jul-14	Spain	H12VL:CMV67;genotype:4b
410	KR993015.1_47C	Urine	30-Jul-14	Spain	H12VL:CMV68;genotype:4a
411	KR993016.1_48C	Urine	31-Jul-12	Spain	H12VL:CMV69;genotype:4a
412	KR993017.1_49C	Urine	07-Apr-12	Spain	H12VL:CMV70;genotype:4c
413	KR993018.1_50cCMV	Urine	18-Jul-11	Spain	H12VL:CMV71;genotype:4a
414	KR993019.1_56C	Urine	10-Sep-14	Spain	H12VL:CMV72;genotype:1
415	KR993020.1_57C	Urine	28-Apr-09	Spain	H12VL:CMV73;genotype:3b
416	KR993021.1_58cCMV	Foetal blood	28-Aug-12	Spain	H12VL:CMV74;genotype:3a
417	KR993022.1_59cCMV	Urine	28-Nov-12	Spain	H12VL:CMV75;genotype:4b
418	KR993023.1_60cCMV	Urine	13-Jun-13	Spain	H12VL:CMV76;genotype:3b
419	KR993024.1_61C	Urine	03-Dec-13	Spain	H12VL:CMV78;genotype:4a
420	KR993025.1_61cCMV	Urine	26-Jun-13	Spain	H12VL:CMV77;genotype:1
421	KR993026.1_62C	Urine	29-Oct-14	Spain	H12VL:CMV79;genotype:4b
422	KR993027.1_64C	Urine	11-Jan-13	Spain	H12VL:CMV80;genotype:1
423	KR993028.1_65C	Urine	22-Sep-05	Spain	H12VL:CMV81;genotype:4b
424	KR993029.1_66C	Urine	03-Sep-12	Spain	H12VL:CMV83;genotype:3b
425	KR993030.1_66cCMV	Urine	23-Apr-03	Spain	H12VL:CMV82;genotype:3b
426	KR993031.1_68C	Urine	09-Feb-06	Spain	H12VL:CMV84;genotype:3a
427	KR993032.1_69C	Urine	04-Mar-03	Spain	H12VL:CMV86;genotype:4a
428	KR993033.1_69cCMV	Urine	31-Oct-13	Spain	H12VL:CMV85;genotype:3a
429	KR993034.1_70C	Urine	12-May-03	Spain	H12VL:CMV88;genotype:4c
430	KR993035.1_70cCMV	Urine	02-Dec-13	Spain	H12VL:CMV87;genotype:1
431	KR993036.1_71C	Urine	02-Apr-12	Spain	H12VL:CMV90;genotype:2

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
432	KR993037.1_71cCMV	Urine	07-Oct-13	Spain	H12VL:CMV89;genotype:4c
433	KR993038.1_73C	Urine	09-May-12	Spain	H12VL:CMV91;genotype:3a
434	KR993039.1_74C	Urine	08-Mar-07	Spain	H12VL:CMV92;genotype:4b
435	KR993040.1_76C	Urine	14-Nov-11	Spain	H12VL:CMV94;genotype:3b
436	KR993041.1_76cCMV	Foetal blood	04-Jun-14	Spain	H12VL:CMV93;genotype:1
437	KR993042.1_77C	Urine	24-Dec-12	Spain	H12VL:CMV95;genotype:2
438	KR993043.1_79C	Urine	26-Dec-11	Spain	H12VL:CMV97;genotype:2
439	KR993044.1_79cCMV	Urine	23-Jul-14	Spain	H12VL:CMV96;genotype:4c
440	KR993045.1_80C	Urine	16-Oct-12	Spain	H12VL:CMV99;genotype:3a
441	KR993046.1_80cCMV	Urine	13-Nov-09	Spain	H12VL:CMV98;genotype:2
442	KR993047.1_81cCMV	Urine	14-Oct-09	Spain	H12VL:CMV100;genotype:3b
443	KR993048.1_82C	Urine	12-Apr-04	Spain	H12VL:CMV102;genotype:3b
444	KR993049.1_82cCMV	Amniotic fluid	16-Oct-14	Spain	H12VL:CMV101;genotype:3b
445	KR993050.1_83cCMV	Urine	19-Aug-08	Spain	H12VL:CMV103;genotype:2
446	KR993051.1_83C	Urine	24-Feb-10	Spain	H12VL:CMV104;genotype:4a
447	KR993052.1_84cCMV	Urine	27-Nov-09	Spain	H12VL:CMV105;genotype:2
448	KR993053.1_85C	Urine	06-Jul-06	Spain	H12VL:CMV106;genotype:3a
449	KR993054.1_86cCMV	Urine	17-May-10	Spain	H12VL:CMV107;genotype:4a
450	KR993055.1_87cCMV	Foetal blood	25-Nov-14	Spain	H12VL:CMV108;genotype:4a
451	KR993056.1_88C	Urine	20-Jun-07	Spain	H12VL:CMV110;genotype:2
452	KR993057.1_88cCMV	Urine	10-Jun-04	Spain	H12VL:CMV109;genotype:2
453	KR993058.1_89C	Urine	11-Oct-07	Spain	H12VL:CMV111;genotype:3a
454	KR993059.1_90C	Urine	14-Sep-08	Spain	H12VL:CMV112;genotype:1
455	KR993060.1_92C	Urine	17-Jul-14	Spain	H12VL:CMV113;genotype:3a
456	KT987979.1_human-wt/ITA/PR219/1998	Urine	1998	Italy	Genotype:gN1
457	KT987980.1_human-wt/ITA/PR421/2001	Urine	2001	Italy	Genotype:gN1
458	KT987981.1_human-wt/ITA/PR217/1997	Urine	1997	Italy	Genotype:gN3a
459	KT987982.1_human-wt/ITA/PR691/1997	Urine	1997	Italy	Genotype:gN3a
460	KT987983.1_human-wt/ITA/PR2121/1995	Urine	1995	Italy	Genotype:gN3b

	Accession num._Strain/Isolate	Isolation source	Date collected	Country	Notes
461	KT987984.1_human-wt/ITA/PR2349/1997	Urine	1997	Italy	Genotype:gN4a
462	KT987985.1_human-wt/ITA/PR2724/1997	Urine	1997	Italy	Genotype:gN4a
463	KT987986.1_human-wt/ITA/PR1663/1996	Urine	1996	Italy	Genotype:gN4b
464	KT987987.1_human-wt/ITA/PR1749/2011	Urine	2011	Italy	Genotype:gN4b
465	KT987988.1_human-wt/ITA/PR1343/1995	Urine	1995	Italy	Genotype:gN4c
466	KT987989.1_human-wt/ITA/PR817/1998	Urine	1998	Italy	Genotype:gN4c

Appendix 12. This table depicts key features of the 466 partial and complete HCMV gO CDSs sequences available in GenBank release 211 that were analysed in this study.

APPENDIX 13: HCMV STRAIN ZMB240 GENE LIST.

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
RL1	Forward	Protein RL1	RL11 family	310	MPDTDSNSTHTTPLHPEYRHTLPLHHSNTQPPIHTSDKHANEEHRTQMELDAADYAACAQARQHLYAQTQPQLHAYPNANPQESAHFSTENQHQLTHLLHNIGEGAALGYPVPRAELRRGGGDWADSASDFDADCWCMWGRFGTMGRQPIVTLLLARQRDGLADWNVVRCRGTGFRAHDSEDGVSVWRQHLVFLLGGHGRRVQLERPSAGEAQARGLLPRIRITPISTSPRPKPPQPTISTASHPHATTHPHHTLFPIPSTPSATVHNPRNYAVRLHAETTRTWRWARRGERGAWMPAETFTCPKDKRPW
RL5A	Reverse	Protein RL5A	RL11 family	90	MTSGKSATLNTKLNSTWTFFNGTGSDVTVCNVVGTNNYTVSDSGLTGICNCTHSSLGFCNVSTKHEGSYDLRRRFDNGPEDVGTWFVVVK
RL8A	Reverse	Protein RL8A	Contains potential transmembrane domain	90	MPHGHLRQALSPTSWTCEGLLLLLGLLVLFFHHHNQSAVERRRRVSFVEVDRLPHESGWYSSDDDGDRDGDEETGESHNRNSVGLSAVFS
RL9A	Reverse	Protein RL9A	Contains potential transmembrane domain	47	MSLDAASHQPAARRLLDSALVRRVLSCMIIVIMIAISIWILTYVLFL
RL10	Forward	Envelope glycoprotein RL10	Type 1 membrane protein; possibly related to RL11 family	170	MYPRVMHVVCFLALGLISYVAVCTENTVTTNCLVKAENTHLTCKCNPNTTSNTNNGSKCHAMCKCRVTEPITMLGAYSAWGAGSFVATLIVLLVVFFVIYAREEEKNNTGTEVDQCLAYRSLTRKKLEQHAAKKQNIYERIPYRPSRQNDNSPLIEPTGTDDEEDEDDDV
RL11	Forward	Membrane glycoprotein RL11	Type 1 membrane protein; binds IgG Fc; involved in immune regulation; RL11 family	234	MQTYSTPLTLVIVTSLFLFTTQGGSSNAVEPTKKPVKLANYRATCEDRTRTLVTRLNTSHHSVVWQRYDIYDRYMRRIPPLCIITDAYKETTRQGGAAFTCTRQNLTLYNLTVKDTGVYLLQDQYTGDVEAFYLIIHPRSFCRALETRRCFYPGPGRVVVTDSQEADRAIISDLKRQWSGLSLHCAWVSGLMIFVGALVICFLRSQRIGEQDAEQLRTNLDTEPLLLTVDGDLE
RL12	Forward	Membrane protein RL12	Type 1 membrane protein; RL11 family	399	MRSIWKPGSNLIYRLIYVLTIWTPLYRSKCENTTNSSSTAESTSIITTTLTTDLSSTYQISSASTSTPISNTATTSTFITSNLNTSYVSPSSNTSIILSTSTETNVGSTEINVTFMNTSATSNVTSHVTDIKVTTVTITPTVVTSTTISNISSVSYEMFNTTINNTQDRDECQLIKVNKTDIEAEEWTNVTIQSNFTIPHCHKVVWMRQYNLTTHGDYFPIRHKRPVKGALYSREICGHTYTHNLLHSYDLCISCDNGTLHLYGVNTTHSGRYTARCHIYEHNDHGTHEDKNFNLIIYPRNNTNNTNGIWICPRPTKDEAQENNQSEEKHITTTDNSVSHKRNHYPRTSHRSAWTVTLLCVACILLFFFRRLFNKKYRMLDDTASESEFIVRYNPEHED
RL13	Forward	Membrane protein RL13	Type 1 membrane protein; RL11 family; consensus is wild type, consisting of several mutants present in different populations	307	MDWRVTVTWTILISTLSESCKPTCPCQCLCNDTTTNYSINTTSTTVINNISTPESSDTTTRTDCSATTAQTTTVSTEPSVTTTRISTLSTKIETTTCMNTTTNVTCDGLNYTVHKTCGRSYEVINVTGHVGSNVTLKKCNHTIWHNVDWIKYEEPYSSYKMCELGNYHQTTPRSNICFECNDTSLTIYNLTAENAGKYTRRDRYNGYEENYYVTVLSEGSTSSTPGTCPVKYKVKSSESTNTEGTVGSNLIETIRNTNIPLGIHAVWAGIVVSVALIALYMGSRRIPRKPRYTRLPKYDPDEFWTKT
UL1	Forward	Membrane protein UL1	Type 1 membrane protein; RL11 family	215	MSVQCNTKLLLLIALIPIAIILTGTLVPVILHEQKKAFYWRLFLQSQQSQHIDAPITVIQGDTVYLNASNNPCNYSSFWYHGNCELCGWNGHLHNFTEYHTNTSCSPKFICINETKGLQLHNVTLNDSGTYTEDVYECDLLCNITNCTYTINSTKYIITVLSPHHSKHTNSHVSTHVGWTAAVVTVIIICVLIYFNVPTTLRHKLQTRNNVNRIT
UL2	Reverse	Protein UL2	Contains potential transmembrane domain	60	MPEDSVSIFIVEDDDDTYPSFGTLPASHAQYGFRLLRGIFLITLVIWTVVWLKLLRDALL

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL4	Forward	Envelope glycoprotein UL4	Contains signal peptide; RL11 family	149	MMLRTWILLQMVLLAAYCHCVFGTCSISTTTAPVEWKSPDRRIPSNITCANYSGTVNGSVTFRGLQNKTEDFLHWLLGWGHKSICSFFPKLQGNKEQHYRYGVTNLTYNCTHDSLTLLNLTTENSGKYYFKREDVNSTFYYSCYNLTVS
UL5	Forward	Protein UL5	Contains potential transmembrane domain; RL11 family	165	MFLGYSDCVDPGFAVYCISRSRLTLVFVVWLVGLRLHDCATFESCCYDITEAESNKAISRDEAAFTSSVSTRTPSLAIAPPPDRSMLLSREEELVPWSRLIITKQFYGGLIFHTTWVTGFVLLGLLTLFASLFRVPQSICRFCIDRLRDIARPLKYRYQRLVATV
UL6	Forward	Membrane protein UL6	Type 1 membrane protein; RL11 family	280	MNGWAGVCLVTHCLNTRSRTYVALNMLAFARTPRGMPSCLFNKVWVSRYALVLILMVCASEGSTNWVVVSHRLPNCSAVSTTVGQNVELCGSASSGCNITQWGRYQNGSTLGPWCTLWGPYTQVSLGHRVAFGCSWTTFFMYNLSQNHSGTYYRKGDNCTDKHITLSCFNLTVHPKAAQSTTTVVTPTVVTNATANVSPITSTLAVNSSAFKHVSYQRQQRVENRTSSKNITNLAFTYGSWGVAMLLFAAVMVLIDLGLPQSAWRRWRIHVDDEERGLLT
UL7	Forward	Membrane glycoprotein UL7	Type 1 membrane protein; modulates cytokine production; involved in immune regulation; RL11 family	222	MVSGVGSHLMTVPRFRWTVHHVYNKLLILALFAPVILESVIYVYAPEGGNVTLVSNFTSNISVRWFRWDGNHSDLICFYKTKEGFSTPYVGLSLSCAASQITIFNLTLNNSGRYGAEGFMKSGENETFLWYNLTVTLKSLKTTSANNVTTIVTTTPTVTGAESNGTRNAILTPQLRAVAGFSNQTPSENNTHLALVGVVVFLILIVVCIMGWWKLLCSKSEL
UL8	Forward	Membrane glycoprotein UL8	Type 1 membrane protein; RL11 family	324	MVSGVGSHLMTVPRFRWTVHHVYNKLLILALFAPVILESVIYVYAPEGGNVTLVSNFTSNISVRWFRWDGNHSDLICFYKTKEGFSTPYVGLSLSCAASQITIFNLTLNNSGRYGAEGFMKSGENETFLWYNLTVTLKSLKTTSANNVTTIVTTTPTVTGAESNGTRNAILTPQLRAVAGFSNQTPSENNTHLALGEGFVPTMTSPGLSTSENYNGNYGLTKTANTTRTNNSDRTTLGASASLLGSTETAVNFDNATTIIPQRVEHPVGEIQYQRTTTHYSWMLIIVIILIIFIIICLRAPRKVYDRWKDSKQYGQVFMTDTEL
UL9	Forward	Membrane glycoprotein UL9	Type 1 membrane protein; RL11 family	231	MMFRYALLLRWITIILCTRKSNYWNYISTPCTSIVGYSGQNISLSPVNKLSAKDDAFQWYIDKPRVTNALCIYQNNECSVQPNENAPNIKWQCVQNHTLILINLTTTYSRNYYFNSFETLGVTIAKYNTLCYNVSVHSAYQTHCCTTTLSMYSPTPVHRSYTLTSTNFTHVAVHYTAGNVEAQHDTATPHTMWIIPLVIVITIIVLICFKFPQKAWNKFTQYRYNSMLAAT
UL10	Forward	Membrane protein UL10	Type 1 membrane protein; RL11 family	254	MRRLINHIVNHDLFRWSVVTAMIFYRYSETCMEVTVRVGDPVTLGSGHGYHPGQKVHWYNQSCVGISNGENTHPICTYDPPKPGRQKTMKTTPLPSPLLYECHNSTLSILHVNVSDPRNYCRRKCPPKGNCEFPTCFTLSLISRTTTRKPEQKTTLLRLKTTPNKHTQHKRSTRRTSPKDYNVTGLPKGFADSFTGNVEAHRAKDAAHSAWILIVIIIIIVVILFFFKIPQRLREKWDTKGYLYKGTDGLPTTD
UL11	Forward	Membrane glycoprotein UL11	Type 1 membrane protein; disrupts T cell function via inhibition of CD45; involved in immune regulation; RL11 family	272	MLLRYITFHREKVLYLTAACIFGVYISLHDACIPVVGKIGTNVTLNAVDVLPPRDQVRWSYGPGGQGYMLCIFTGTSTTTFNNTRFNFSCLSNYSLLLINVTTQYSTTYRTMTSLDHWLHQRHNHGSRWTLDTCYNLTVNENGTFPTTTTKKPTTTTRTTTTTTQRTTTTRTTTTAKKTTISTTHHKHSSPKKSTTPNSHVEHHVGFEATAAETPLQPSPQHQHVATHALWVLAVVIVIIIIIIFYFRIPQKLWLLWQHDKHGIVLIPQTDL
UL13	Forward	Protein UL13	Contains signal peptide	472	MLWAHCGRFLRYHLLPLLLCRLPFLLFFQRPQWAHGLDIVEEDEWLQEIQGATYQLSIVRQAMQHAGFQVRAASVMTRRNAVDLDRPPLWSGSLPHLPVYDVRSPRPLRPPSSQHHAVSPELPSRDGIRWQYQELQYMVEEQRRRNQSRNAIPRPSFPPPDPPSQPAEDARDADAERAESPHSAESTVRHDASENALRQRHERRRYNALTVRSRDSLLLTRIRFSNQRCFGRGRLRHPAGSGPNTGGPRPGGAGLRQLRQQLTVRWQLFRLRCHGWTQQVSSQIRTRWEESNVVSQTATRVRTWFVERTTLWRRTWVPGQNPAAEAQELAVIPPAPTVLRQNEEPRQQLTGEETRNSTHTQREEVEDVSREDAREGNDGSRASGNDERRNNAGRYDDHEVQEPQVTYPAGQGELNRRSQEENEEGGPCESPPMTTNTLTVACPPREPPHRALFRLCLGLWVSSYLVRRPMTI

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL14	Forward	Membrane protein UL14	Type 1 membrane protein; possibly involved in immune regulation; UL14 family	327	MGGGRLPPLWLPLLIAWSEWGNCCLDAPPVVRSPCLQPVRDRNRERNPGSPQLLPYGDRLEVACIFPAHDWPEVSIRVHLCYWPEIVRSLVVDARSGQVLHNDASCYIAGGRWRFEDGGAAQRLSLSFRLITETAGTYTCVLGNETHSLATETTALVADVHDLRHSDRSCDLAFGSRSQTRYLWTPDPSRLRSINCGWEGERHRVVHYIPGTSGLLPSCEEDERELCVPFISHSIADNNCSRRHRVDGARRRYYLRRDYWLTDPKIGLLAAGSVALTSLCHLLCYWCSESYRRLNTEEESEAAEETAAGEASAVAAAAVSEEEQQRE
UL15A	Forward	Protein UL15A	Contains potential transmembrane domain	102	MKRMIRSHGRKTECQMTGAGERRGSAVGALICGSGTRRGSGANERRNSDVGPIAHSSGTRRGSANETSACTRTDHQKADIGLWFMFLFFGLCSWLSMRYRAQ
UL16	Forward	Membrane glycoprotein UL16	Type 1 membrane protein; inhibits NK cell cytotoxicity; sequesters MICB, ULBP1 and ULBP2; acts on activating receptor NKG2D; involved in immune regulation	230	MERRRGTVPLGWVFFVLCLSASSPCAVDLGSKSSNSTCRLNVTELASIHPGETWTLHGMCISICYYENVTEDEIIGVAFTWQHNESVVDLWLYQNDTVIRNFSDITTNILQDGLKMRTVPVTKLYTSRMVTNLTVGRYDCLRCENGTMKIIERLHVRLGSLYPRPPGSGLAKHPSVSADEELSATLARDIVLVSAITLFFFLLALRIPQRLCQRLRIRLPHRYQRLRTED
UL17	Forward	Protein UL17	Possibly related to UL19	104	MDHALLTHFVGRPRHCRLEMLILDEQVSKRSWDTTVYHRRRKHLPRRRAPCGPQRPAEIPKRRKKAAVLLFWHDLCWLFRRLFFPREDSEPLMSDPARSPEEEE
UL18	Forward	Membrane glycoprotein UL18	Type 1 membrane protein; inhibits NK cell cytotoxicity; acts on inhibitory receptor LIR-1 (ILT2); involved in immune regulation; MHC family	367	MMTTWCLTLFVLWMLRVVGMHVLRYGYTGIFGESHMTLTVVGIFDGQHFFTYHVNSSDRTSSRANGTISWMANVSAAYPTYLDGERAKGDLIFNQTEQNLLELEIALGYQSQSVLTWTHECNTTENGSFLAGYEGFGWDGETLMEFKDNLTLWTGPDYKISWLRQNKTYIDGKIKNISEEDTTTQRNYLKGNCTQWSVIYSGFQTPITHPVVKGGVRNHNDNRAEAFCTSYGFFPGEINITFIHYGDKVPEDSEPQCNPLLPTLDGTFHQGCYVAIFCNQNYTCRVTHGNWTVEIPISVTSPDDSSSGEVPDHPTANKRYNTMTISSVLLALLLCALLFAFLHYFTTLKQYLRNLAFAWRYRKVRSS
UL19	Forward	Protein UL19	Possibly related to UL17	98	MTSNALYELFRRRLPRAPVNTVMFLTRRTRDGFCGRLTSIATNSHYTMFVLDHGSVRIERPSQSEVDCASLMETLKRIRLRNSWVASEDELDVSRRDA
UL20	Forward	Membrane protein UL20	Type 1 membrane protein	340	MLGIHAMLVMLNYHWAQVTTNNDARNNNTDTIFVSLLTGPNGVTRTAVGGLYSNYTDLTGTFNFIQGNISANASSGDNWSVANLTKNCINRGESYLTTLWLLNCTQNDTYWYSGNAYNYTNNTCGSTVSGYLLGMCELWKKWVGNDTSHNTTRIELLKNETRCTLPAKQYTLNVTVEWYNKSEGDIPKEFMSYAILSSVAVLTCGLQEAYIFDMTRRNTYLFSVSCIGITSIISIILASLSLLILICYYRCGRLLICPRGFERLPEFTEEEEEKENLLTHKDIEVQVPIRTRRLLVPWIRESKMWTLPPPLPPRPPHLIEFPPSPPSSPEPTHMVICIPS
UL21A	Reverse	Protein UL21A	Facilitates viral DNA synthesis	123	MGGSPVPQLTTVTQGLMPSVRMDFRARRPLRRLAFYAPRARRRLFQNHIHPEQRRVLVGEGDEEMLPGLPMEIDIVIDRPPQQPLPNPLVLLLDDVPPHVPGFAPYRVPRPHPMIPEEHWDQF
UL22A	Forward	Glycoprotein UL22A	Contains signal peptide; secreted protein; binds CC chemokine RANTES; involved in immune regulation	106	MARRLWILSLLAVTLTVALAAPSQKSKRSVTVEQPSTSTNSDGNNTTRNKDVTLSQGGSTTDGDEDYSGGDYDVLITDTDGGNHQQPQKKTNEHKEEHTKENEKTQ
UL23	Reverse	Tegument protein UL23	US22 family	284	MSVIKDCFLNLLDRWRPPKTSRPWKPGQRVALVWPKDRCLVIRRRWRLVRDEGRDAQRLASYLCCPEPLRFVGSICTYNFLKHKGDHNVPSELYLGASGAMYLWTDHIYSDSLTFVAESITEFLNIGLRRCNFITVPEELPHTASLRALAGCMHIHAFAQWRATYRGRIMVVGDYTVIRVSTIRLYDWSEINDWRVMVGSNHVEPLGWLVSPYDVINLFVDDCMRVFAANNQHVCIVADSLMEFVTRGMTRCHENGIYYGTRSMRKLNQPTCPYGVDHQLFDDA
UL24	Reverse	Tegument protein UL24	US22 family	358	MEETRAGRYIVRAGFGGQTEDTASGAESVLATLAAVRTRRRSVVCDGAPGSPTDSARHMSDLASLALTAEFGLGCLEAYVRINAGQVLPVVWPPGWNLVLQEIETDEDFKPEDVKAWSHYLCCETRLAFVGRFVNEAVLSPDQQKKTAVCLISDEGYVFCYVREDTAVYYLARNLMEFARVGLRAVETLHCMRYLTSSLVKRYFRPLLRAWSLGL

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					DTMARFIIRHHGQFMPLTYPPGTELRLCNLRCFENSVEGGHLLRNIKAAFGMRVLGLGTVSLKGENAPFPHL RWPVDLIPIVVAYTGAVYACDVRDDRYIRVGDNLNTFMCLGLNLLFENRRFSGHNGIYDRVPDCPKGRQHR
UL25	Forward	Tegument protein UL25	UL25 family	656	MSSRRRSSSRRSGEPSTVIYIPSSNEDTPADEEAEDSVFTSTRARSATEDLDRMEAGLSPYSVSSDAPSSFELVR ETGGTGAAKKPSEKKRSSSRRQPQIAAGAPRGSPATPKAGKSPKVSRPPSVPSLPENGAGGGGDDNSSSGG SSSRTTSNSSRSTSPVAPGEPSAAEGDEFSFCDSDIEDFERECYRVSVADNLGFEPSVVAPQHVEYLKFVLQDF DVQHLRRLNECIPMPAFALTSLVDPVLNNVAPGERDLTRRIITHAVIINYYYVAQKKARHMVEAIRTTVRGDT VRRVAAQVNNQSRSGRAAALALHFLTSRKGVTDGQYATSLRRLDEELRHRGTPESPRLTEVYQTLRDYNVLF YTAHYTSRGALYLYRQNLQRLNENHRGMLRLLSVEEICEEHTLNDLAFLVGVELMITHFQRTIRVLRCYLQHQ LQSISELCYLIYVQLPSLREDYAQLSDVLYWAVSQNYDYALYASTPALFDFMRVVRQQDAFICTDYVYCALRLL ACPDRPIIGDTGGSSSSQRLVGEFMVRDPLLRDPRAAHLRQKLITRDICVPRLQAQPSSRHIPVEHTGVSSVTL LKIFSQVPPDEREEDTLREMALKAFMEANGNHPEQICRSPPPPLPPRDYPQRDDRDRHRRDRRDSGEYCC
UL26	Reverse	Tegument protein UL26	transcriptional activator of major immediate early promoter; involved in gene regulation; US22 family	188	MTSRRAPDGGLNLDDFMRRQRGRHLDLPYPRGYTLFVCDVEETILTPRDVEYWKLLVVTQGQLRVIGTIGLA NLFSWDRSVAGVAADGSVLCYEISRENFVVRAADSLPQLLERGLLHSYFEDVERAAQGRLRHGNRSGLRRD ADGQVIRESACYVSRALLRHRVTPGKQEITDAMFEAGNVPSALLP
UL27	Reverse	Protein UL27	Low-level resistance to maribavir	608	MNPVDQPPPPLPTQQPEEQAKEDHDDGDERLFRDPLTTYEYLDDCRDDEEFCHQFLRAYLTPIRNRQEAVR AGLLCRTPEDLAAAGGQKKKTPAPKHPKHAMVYIRRSCLVHSACATAHGKYDIRGLTLESDLAVWAALRGV PLPPDPQHFRWLNAGAFRRLVHEAQYLPEISRAAKRIALAVATGQYVVCTLLDYKTFGTRTHYLRQLCSMTE ELYLRLDGTLCLFLEPEERELIGRCLPAALCRGLPVKYRTHRAAVFFHATFMARAEAALKDLYAAFCECGDGRD NGGNHDGNHDGNDHSSLSPSAVASHHSRLEHAELRLERNRHLGAFHLPAIRHLTAGDVARVQDSVSRDLG FADWSQTLVDDYFLLPAGWACANPRRGYAMYLASNAVLALRIIRLLRASIRHEYTACIRMLSGDVQRLIRLFK GEAALLRKGLAQNPVQRRELSRFRKHVHDLKRIRFTEDTFVETFCDFLELVQRIPDYRSVSLRIKRELLCLHVFK LRRGCRAPPTPETARVQRLLWHSLRHGDAPQDRTRLPQFSSALSDAELSNHANRCRRKAPLELGPAVVAAP GPSVRYRAHIQKFERLHVRRFRPHEVGGHAT
UL29	Reverse	Protein UL29	Interacts with components of the NuRD complex, promoting accumulation of immediate early RNA; US22 family	701	MSGRRKGCSAATASSSSSSPPSRLPPLPGHARRPRRKRCLVPEVFCTRDLADLCVRRDYEGLRRYLRRFEGSC VSLGWPSQCIYVVGGEHSPHSLTEIDLEHCQNDFFGEFRALHLIGTVSHATCRYQVFVDAYGAVFAYDAQED CLYELASDLAGFFAKGMIRCDPVHESICARLQPNVPLVHPDHRAELCRRSRASARGRYLRSLLAFRELLACEDT AARCAYVEAHREAQLTLIWPEKHSLVLRTARDLGLSASMLRRFQRSLYTREPVMPLGEIEGAEDKTFFHRVRI LCGDTGTVYAALVGQDKLVRLARDLRGFVRVGLALLIDDFRYESIGPVDRSSLYEANPELRLPFKKRRLVVGYF DSLSSLYLRGQPKFSSIWRGLRDAWTHKRPKPRERASGVHLQRYVRATAGRWLPLCWPPLHGIMLGDTQY FGVVRDHKTYRRFSCLRQAGRLYFIGLVSVYECVPDANTAPEIWVSGHGHAFAYLPGEDKVYVLGLSFGEFFE NGLFAVYSFFERDYVDEIVEGAWFKHTFAGMYELSQILHDRANLLRVCQLHAGSKIRLGGSPACTFTFGSWN VAEADEANNFVIGVLEQAHFVVIGWMEPVNKAVFMDAHGGIHVLLYGTMLVKLAETLRGFIRQGSFWFRC PRRFCFSPLDSSATVAAKPVSSHTSPAYDVSEYVFSGRSVLDSVSGTGAS
UL30	Reverse	Protein UL30	UL30 family	121	MTTSAMTAPDTRRQLQHVETLRRFLRGDSCFVHDLRGMMDYHDGLSRRQQRAFCRAGRVLTDPEPIQSET EGENKQFTEHTHKVVSFFIKSVFVFPCLVLPNCCQVSVDRSRVPETGGRWL
UL30A	Reverse	Protein UL30A	UL30 family; alternative start codon	79	MACRDARWVRFIRLRTLWSVRPRAHVADLEDFAAFYRTLSDSEQQEFEQEAELASRSQRVRDLREARRQLK MDLMCHGG

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL31	Forward	Protein UL31	DURP family	595	MGDKPTLVTLLTVAVSSPPPSSPLPLVSFTELLLPPPSVAAAAVAATATSEVGEKTAEQEVAAAGPETGNERRENREDEGGETRTTGTTAVKRSHDGIPRQLAERLRLCRHMDPEQDYRLPAQDVVTSWIEALRDADRDNYGRCVRHAKIHRSASHLTAYESYLVSITEQYNTASNVTEKASYVQGCIFLSFPVIYNNTQGCGYKYDWSNVVTPKAAYAELFFLLCSTSESSVVLQPLITKGGLCSSMAVYDEETMRQSQAVQIGFLHTQLVMVPFVPHACPHYAVPFTTPGKPGCGGAPSGVAGLEEAAPFGRVSVTRHGATLLCRVDHLTWISKRVTTYGHKKITRYLAQFRGTMDDDEAALPGEDEAWIASKNVQYEFMGLIFTVNVDSLCVDAEQRQLLGTVATSFCHRVSDKITARNMPRAFSFYLLTSAQRGYDLRFSRNPSLFFSGDALNCPLLNEPNVFSLTVHAPYDIHFGVQPRQTVELDLRYVQITDRCFLVANLPHEDAFYTGLSVWRGGEPLKVTLWTRTRSIVIPQGTPIATLYQITEGDGNVYSYNHHTVFRQMHAAGATTFFLGDMQLPADNFLTSPHP
UL32	Reverse	Tegument protein pp150	Major tegument protein; binds to capsids	1048	MSLQFIGLQRRDVVALVNFLRHLTQKPDVDLEAHPKILKKCGEKRLHRRTVLFNELMLWLGYYRELRFHNPDLSSVLEEFEVRCAAVARRGYTYPFGDRGKARDHLAVLDHTEFDTDVRHDAEIVERALVSAVILAKMSVRETLVTAIGQTEPIAFVHLKDTEVQRIEENLEGVRRNMFCVKPLDLNLDRHANTALVNAVNKLVYTGRLIMNVRRSWEELERKCLARIQERCKLLVKELRMCLSFDSNYCRNILKHAVENGDSADTLLELLIEDFDIYVDSFPQSAHTFLGARSPSLEFDDDANLLSLGGGSAFSSVPKKHVPTQPLDGWSWIASPWKGHKPFRFEAHGSLAPAAEAHAARSAAVGYYDEEEKRRERQKRVDDEVVQREKQQLKAWEERQQNLQQRQQQPPPSTRKPGASRRLFGSSADEDDDDDDEKNIFTPIKKPGTSGKGAASGGGVSNIFSGLLSSGSQKPTSGPLNIPQQQQRHAAFSLVSPQVTKASPGRVRRDSAWDVRPLTETRGDLFSGDEDSDSSDGYPPNRQDPRFTDTLVDITDTETNVKPPVTTAYKFEQPTLTFGAGVNVPAGAGAAILTPTPVNPSTAPAPAPTPTFAGTQTPVNGNSPWAPTAPLPGDMNPANWPRERAWALKNPHLAYNPFRMPTTSTASQNTVSTTPRRPSTPRAAVTQTASQNAADEVWALRDQTAESPVEDSEEEDDDSSDTGSVVSLGHTTPSSDYNDVISPPSQTPEQSTPSRIRKAKLSSPMTTTSTSQKPVLGKRVATPHASARAQTVTSTPVQGRLEKQVSGTPSTVPATLLQHQPASSKTTSSRNVTSGAGTSSASSARQPSASASVLSPTEDDVVSPATSPLSMLSSASPSPAKSAPPSPVRGRGSRVGVVPSLKPTLGGKAVVGRPPSVPVSGSAPGRLSGTSRAASTTPTYPAVTTVYPPSSTAKSSVSNAPPVASPSILKPGASAALQSRRSTGTAAVGSPVKSTTGMKTVAFDLSSPQKSGTGPQPGSAGMGGAKTPSDAVQNILQKIEKIKKTEE
UL33	Forward	Envelope glycoprotein UL33	Type 3 membrane protein; 7 transmembrane domains; putative chemokine receptor; involved in intracellular signalling; GPCR family	411	MDTIIHNTTNRSTSTPHVNSTCNMTEPLSAIRTTEAVINTFIIFVGGPLNAIVLVTQLLTNRVLGYSTPTIYMTNLYSTNFLTLTVLPFIVLSNQWLLPASVTSCKFLSVIYYSSCTVGFATVALIAADRYRVLHKRTYARQSYRSTYIILLLTWFAGLIFSMPAAVYTTVVIHNGTDENTNGHATCVLYFIADEVYTVLLSWKVLLTLVWGAAPVIMMTWFYAFFYSTVQRASQKQRSRTLTFVSVLLISFVALQTPYVSIMIFNSYATAAWPMDCEHLTLRRTIGTLSRLVPHLHCLINPILYALLGHDFLQRMRQCFRGQLLDRRAFLRSQQNQRATAETNLAAGNNSQSVATSLDTNSKNCNQHAKRSVSFNFPSGTWKGGQKTASNDTSTKIPHRLSQSHHNLSGV
UL34	Forward	Protein UL34	Represses US3 transcription; involved in gene regulation	407	MNFIITTRDFSNDDSVLRAAEMRDNVAGSISKAYKGTVRAEGKKKLLLKHLPVPPGGCSRRNSNLFVFCTERDYRKFHQGIAQLKRAPAELDPHEIQQVTASIRCRLQPSLREPPTPADELQTAVSRVCALFNQLVFTAQLRHYCEHQDKVVSYARDELTKRCGEKSALGVEVHQLVALLPHERHRELCHVLIGLLHQTPHMWARSIRLIGHLRHYLQNSFLHLLMNSGLDIAQVFDGCYHSEAYRMLFQIGHTDSVSAALELSHSAAAGPPEADENNDEGEEDDDELRHSDPAPLHESKKPRNARRPRTRVPPHEQKPEENEEEEEELFPSCKATAAFLRAEPSVSNDDGNGGERCDTLATALRHRADEEDGPLASQTAVRVAATPSPSVTPAFTPVTSPITPLCI
UL35	Forward	Tegument protein UL35	UL25 family	641	MAQGSRAPSGPPLPVLPVDDWLNFRVDLFGDEHRRLLLEMLTQGCSNFVGLLNFGVPSPVYALEALVDFQVRNAFMKVKPVAQEIIRICILANHYRNSRDVLRDLRTQLDVLYSEPLKTRLLRGLIRLCRAAQTGVKPEDISVHL

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					GADDVTFGVLKRALVRLHRVRDALGLRASPEAEARYPRLTTYNLLFHPPPFTTVEAVDLCAENLSDVTQRRNRPLRCLTSIKRPGSRTLEDALNDMYLLLTLRHLQLRHALELQMMQDWVVERCNRLCDALYFCYTQAPETRQTFVTLVRGLELARQHSSPAFQPMLYNLLQLLTQLHEANVYLCPGYLHFSAYKLLKKIQSVSDARERGEFGDEDEEQENDGEPREAQLDLEADPTAREGELFFFSKNLYGNGEVFRVPEQPSRYLRRRMFVERPETLQIFYNFHEGKITTETYHLQRIYSMMIEGASRQTGLTPKRFMELLDRAPLGQESEPEITEHRDLFADVFRRPVTDAASSSSASSSSSSASPNSVSLPSARSSSTRTTTPASTYTSAGTSSTTGLLLSSSSLSGSHGISSADLEQPPRQRRRMVSVTLFSPYSVAYSHHRRHRRRRSPPPAPRGPAHTRFQGPDSMPSTSYGSDVEDPRDDLAENLRHL
UL36	Reverse	Tegument protein vICA	Inhibitor of caspase-8-induced apoptosis; involved in apoptosis; US22 family	476	MDDLRDTLMAYGCIAIRAGDFNGLNDFLEQECGTRLHVAWPERCFIQLRSRSALGPFVGKMGTVCSQGAYVCCQEYLHPFGFVEGPGFMRYQLIVLIGQRGGIYCYDDLRDCVYELAPTMKDFLRHGFRHCDHFHTMRDYQRPMVQYDDYWNAVMLYRGDVESLSAEVTKRGYASYTIDDPFDECPDTHFAFWTYNTEVMKFKETSFSVVRAGGSIQTMELMIRTVPRITCYHQLLGALGHEVPERKEFLVRQYVLVDTFGVVYGYDPAMDAVYRLAEDVVMFTCVMGKKGHRNHRFSGRREAIVRLEKTPTCQHPKKTPDPMIMFDEDDDDELSLPRNVMTHEEAESRLYDAITENLMHCVKLVTTDSPLATHLWPQELQALCDSPALSLCTDDVEGVRQKLRARTGSLHHFELSYRFHDEDPETYMGFLWDIPSCDRCVRRRRFKVCDVGRRHIIPGAANGMPPLTPPHAYMNN
UL37	Reverse	Envelope glycoprotein UL37	Type 1 membrane protein; involved in apoptosis; involved in gene regulation; major isoform (mitochondrial inhibitor of apoptosis, vMIA) encoded by unspliced mRNA	490	MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETERSVDLVAALLAEAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPIDAEGLSGLAREACEALKKALRRHRFLWQRRRRARLLQHNGPQQSHHAAVFCRVHGLRGFQVSVWLLLTLLWSTGNGVSVRCTYHGTDINITSNTTSMNCQLNCTCNNTQIYNGPCVGTEARLPLNVTFNQSRRQWHSVMLKFGFQYHLEGWFPLRVLNESHETNVTEVHGEVACFKNDTNITVGQLMLNFTGHSYVLRAVAHTSPFESYVHWEETNATSNMTSNVTSLENTTTVMLTLAKYAESDYIFLQDMCPRFLRRTLKITKRKTINATFTGTNVTSLPMWTPECEGWKYWTTLSIMWKNRRSALLRAKSRALGHWALLSICTVAAGSIALLSLFCILLIGLRRDLLEDFRYICRDEGSSSTKNDVHRIV
UL38	Reverse	Protein UL38	Involved in apoptosis	331	MTTTTHSTAAIMSLLDEAEWRQTQMDVGGLIQASALGKVALRYAVRKLMKRGARLRHDSGLYVCICDPSYEFLQMNLSKISWLERHCPPLDQELIMFGVIEAWEEASRRPTRQLVLFMTPKWDVFAYDSGILFFLAPSMAQFWHGAIVLEYWNALFPVEVRSHVRQHAHTMDDLVMVFHQLDYEKQVLEARRDKNTEGPRTFAKSVNSYVRAILESERRIREGKIPMTFVDRDSLRANSLAHIQATGAQPPHAPAQRVLSAPPSLPSPVSEEDPAAAAIPPSSAATTPPSSVVSASVESELSSSPPLPPVVVKDVMYTAGEGDVVQMVVVV
UL40	Reverse	Membrane glycoprotein UL40	Type 1 membrane protein; sequence in signal peptide inhibits NK cell cytotoxicity; upregulates cell surface expression of HLA-E; acts on inhibitory receptor CD94/NKG2A; involved in immune regulation	221	MNKFSNTRIGFTCAVMAPRTLILTLGLLCMRIRSLLCSPAETTVTTAGAMFAHGPRCPLVFQGWVYAVYHQGDMALMTLDVYCCRQTSSNTVVAFSHHPADNTLLIEVGNNTRRHVDGISCQDHFRAQHQDCPAQTVHVRGVNESAFGLTHLQSCCLNEHSQLSERVAYHLKLRPATFGLETWAMYTVGILALGSFSSFYSQIARSLGVLPNDHHYALKKA
UL41A	Reverse	Protein UL41A	Contains potential transmembrane domain	78	MTLFCRTANSTAGYVDMNVCIGMIGVVCFVFGVFIIIFCSTLKAFYVRQKTYHLLGTESDDEETTVWEKRRMESDTDF
UL42	Reverse	Protein UL42	Contains potential transmembrane domain	122	MEPTPMLRDRDHDDAPPTYEQAMGLCPTTVSTPPPPPPDCSPPPYRPPYCLVSSPSPRHTFDMDMMEMPATMHPTTGAYFDNGWKWTFALLVVAILGIIFLAVVFTVVINRDNSTTTGTTSG
UL43	Reverse	Tegument protein UL43	US22 family	423	MEKTPAETTAVSAGNVPRDSIPCITNVSADTRGRTRPSRPATVPQRRPARIGHFRRCSASLSFLDWPDDSVTEGVRTTSASVAASAARFDEIRRRRQSINDEMKERTLEDALAVELVNETFRCSVTADARKDLQKLVRRVSGTVLRLNWPNGWFFTYCDLLRVGYFGHLNIKGLEKTFLCCDKFLLPVGTVSRCEAIGRPPLPILIGEGGRVYVYSPVV

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					ESLYLVSRSGFRGFVQEGLRNYAPLREELGYVRFETGGDVGREFMLARDLLALWRLCMKREGSIFSWRDGN EALTTVVLNGSQTYEDPAHGNWLKETCSLNVLQVFVVRAVPVESQQRLDISILVNESGAVFGVHPETRQAHF LARGLLGFFRVGFLRFCNNYCFARDCFTHPESVAPAYRATGCPRELFCRRLRKKKGLFARR
UL44	Reverse	DNA polymerase processivity subunit	dsDNA-binding protein; involved in DNA replication	433	MDRKTRLSEPPTLALRLKPYKTAIQQLRSVIRALKENTTVTFLPTPSLILQTVRSHCVSKITFNSSCLYITDKSFQP KTINNSTPLLGNFMYLTSSKDLTKFYVQDISDLSAKISMCAPDFNMEFSSACVHGQDIVRESENSAVHVDLDF GVVADLLKWIGPHTRVKRNVKKAPCPTGTVQILVHAGPPAIKFILTNGSELEFTANNRVSFHGVKNMRINVQ LKNFYQTLLNCAVTKLPCTLRIVTEHDTLLYVASRNGLFAVENFLTEEPFQRGDPFDKNYVGNSGKSRGGGG GGGSLSSLANAGGLHDDGPGLDNDLMNEPMGLGGLGGGGGGGGKKHDRGGGGGSGTRKMSSGGGGG DHDHGLSSKEKYEQHKITSYLTSKGGSGGGGGGGGGGLDRNSGNYFNDAKEESDSEDSVTFEFVPNTKKQK CG
UL45	Reverse	Ribonucleotide reductase subunit 1	Enzymatically inactive; tegument protein	906	MNPADADEEQRVSSVPAHRCRPGRIPSRGSETETEESSAEAAADTTGGDDSELEEGPLPGGDKEASAGNTN VSSGVACVAGFTSGGGVVSWRPESPSPDGTPSVLSLTRDSGPAVPSRGGRVSSGLSTFNPAGATRMELDSV EEEDDFGASLCKVSPPIQATRMLMGKKCHCHGYWGKFRFCGVQEPARELPSDRNALWREMDTVSRHSAG LGSFRLFQLIMRHGPCLIRHSPRCDLLLGRFYFKANWARESRTPLCYASELCDESVRRFVLRHMEDLPKLAEET ARFVELAGCWGLYAAILCLDKVCRQLHGQDESPGGVFLRIAVALTAAIENSRHSRIYRFHLDARFEGEVLESVL KRCRDGQLSLSTFTMSTVGFDRVPQYDFLISADPFSRDASWAAMCKWMSTLSCGVSVSVNVTRLNADVNS VIRCLGGYCDLIREKEVHRPVVRVFVDMWDVAAIRVINFILKESTSELTGVCYAFNVPSVLMKRYRAREQRYS LFGRPVSRRLSDLGQESAFEKEYSRCEQSCPKVVVNTDDFLKKMLLCALKGRASVVFVHHVVKYSIMADSVC LPPCLSPDMASCHFGECDMPVQRLTVNVARCVFARSDEQKLHLPDVVLGNTRRYFDLSVLRELVTEAVVW GNARLDALMSASEWWVESALEKLRPLHIGVAGLHTALMRLGFTYFASWDLIERIFEHMYFAAVRASVDLCK SGLPRCEWFERTIYQEGKFIFELYRLPRLSIASARWEALRADMLEFGLRNCQFLAVGPDDEVAHLWGVTPSV WASRGTVFEEETVWSLCPPNRECYFPTVVRRPLCVPVVNYAWLEQHQEEGKATQCLFQAAPAIQNDVEM AAVNLSVFVDQCVALVFYYDSGMTPDVLLARMLKWYHWRFKVGVYKYCAS
UL46	Reverse	Capsid triplex subunit 1	Complexed 1:2 with capsid triplex subunit 2 to connect capsid hexons and pentons; involved in capsid morphogenesis	290	MDARAVAKRPRDPADEDNELVTALKAKREVNTISVRYLYHADHQALTARFFVPEGLVEFEAQPGALLIRMET GCDSPRHLYISLYLLGIRASNVSASTRCLLESVYTASAARAALQWLDLGPHLLHRRLETLGCVKTVSLGITSLLTC VMRGYLYNTLKTEVFALMIPKDMYLTWEETRGRLQYVYLIIVYDYDGPETRPGIYVLTSSIAHWQTLVDVAR GKFARERCSFVNRRITRPRQIPLCTGVIQKLGWCLADDIHTSFLVHKELKLSVVRLDNFSVELGDFREFV
UL47	Forward	Tegument protein UL37	Complexed with large tegument protein; involved in virion morphogenesis	983	MMARRTVDFKKLIEQLRARATDKAEALNTVSQLEIGAVDAQDVTASAVRAFVGALPSSGYHFGFVRQNVVF YLLSHATVQTARDPLYAAEQLHEQLDRFLRHQHDGGGDEDRLPFYHNGATLTAFQKLLQTLREIQTVIAEQS GGAAAAADLIASNNASAERRGKKGGSSSGGQQPLVRRVITQLETAVTEARPYVNCRAVAELLDLTYQRLIYW ACTLMPYVLFRRDTDTELDTVLLMHFFYTHYRSVNGDLAVEFQNYVKNSVRHMSSFVSSDIDGDQKPGAEH MRDVSYKLFVGNLQARDASGLMFPIISTRISTVNLYLSPERMFFHPGLISRLLSEEVSPRANLDAYARVCDRVL EDHLHTPRRVQRLLDLTQMVTRLVELGFNHDTCAAYAQMALIQPASQKSSLFVSEIREKLIQIIYNFYTFFTCLY VYSPTFLFDHRRRLILEQHRSTLIGSKEELQHVWSNVTLNVNTHFAVQYTEEDFEAHTKGATEAEREYLYRDL HSKWGVHLFTLRPSRGAAGAASPLPPLDGVTRSDILRECALVNLNEGRVNYASLLAFSHHPEFPSIFAQLVVV TEFSEIFGIPQGLFQAVGSPRLFALIQLCRVLLPEQVTLYQNLVSIYNLTTFVKHIDAAVFKTVRDCVFDIATTLE HLSGVSVTPNVDLLAELMARSVAHNLYTTVNPLIEDVMRSSAGSLRNYLRHTRLCFGLARGRARLSEDGVTV YVEVQGQYGLRVPTTRFVEQLRELVRRDRLLAENLRGLNERLLSVRVRVRQISSDTEEVSRHAKGHRTVAQM

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					SKALKKTASKIKVLETRVTLALEQAQRSNGAVVTAVQRALAVFDVLSRENLERRGAQLCLTEATSLLHRHRAL APMTWPAGTGVAAAAEADRALREFLEAPWESAPQPPRLRMTPDTDHEESTAGATSVPEVLGARYEPAHLA ASDLLNWYIVPVSQAQQDILSSIDPPAGSTSVSLPPASP
UL48	Forward	Large tegument protein	Complexed with tegument protein UL37; ubiquitin-specific protease (N-terminal region); involved in capsid transport	2240	MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLDARLERELQKK LPAGGRLPVYRLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCPGIFDFLRYAHAKPRPTYVLVTVN SLARAVVFTEDHMLVFDPHSSAECHNAAVYHCEGLHQVLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEA EIAARLVSTYRDRDIDLTGVVRESADTAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPDKRR KDLSGSKHGGKKKPPSTTSKTLATASSSSAIAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKALTLP PLPANTSAWTLHAAGTESGANAATATAPSFDEAFLTDRLQQLIIHAVNQRSCLRRPCGPQSAAQQAVRAYL GLSKKLDAFLLNWLHHGLDLRRMHDYLSHKTTKGTYSTLDRALLEKMQVVFDPYGRQHGPALIAWVEEML RYVESKPTNELSQRLQRFVTKRPMPVSDSFVCLRPVDFQRLTQVIEQRRRVLQRQREEYHGVYEHLAGLITSI DIHDLGASDLNRREILKALQPLDDNAKQELFRLGNAKMLELQMDLDRLSTQLLTRVHNHILNGFLPVEDLKQ MERVVEQVLRLFYDLRDLKLCDGSYEEGFVVIREQLSYLMTGTVRDNVPLLQEILQLRHAYQQATQQNEGRL TQIHDLLHVIETLVRDPGSRGSALTLALVQEQLAQLEALGGLQLPEVQQRLQNAQIALSRLYEEEEETQRFLD GLSYDDPPTEQTIKRHPQLREMLRRDEQTRLRLINAVLSMFHTLVMRLARDESPRPTFFDAVSLLLQQLPPDS HEREDLRAANATYAQMVKKLEQIEKAGTGASEKRFQALQELVYFFRNHEYFFQHMVGRLGVGPQVTELYER YQHEMEEQHLERLEREWQEEAGKLTVTSVEDVQRVLARAPSHRVMHQMQQTLTTKMQDFLDKEKRKQE EQQRQLLDGYQKKVQQDLQRVVDAVKGEMLSTIPHQPLEATLELLLGLDQRAQPLLDKFNQDLLSALQQLS KKLDGRINECLHGVLTGDVERRCHPHREAAMQTQASLNHLDQILGPQLLIHETQQALQHAVHQAQFIEKCQ QGDPTTAITGSEFESDFARYRSSQQKMEGQLQETRQQMTETSERLDRSLRQDPGSSSVTRVPEKPFKGQEL AGRITPPPADFQRPVFKTLLDQQADAARKALSDEADLLNQKVQTQLRQRDEQLSTAQNLWTDLVTRHKMS GGLDVTTPDAKALMEKPLETLRELLGKATQQLPYLSAERTVRWMLAFLEEALAQITADPTHPHHGSRTHYRN LQQQAVESAVTLAHQIEQNAACENFIAQHQEATANGASTPRVDMVQAVEAVWQRLEPGRVAGGAARHQ KVQELLQRLGQTLGDLELQETLATEYFALLHGIQTFSYGLDFRSQLEKIRDLRTRFAELAKRRGTRLSNEGALPN PRKPQATTSLGAFTRGLNALERHVQLGHQYLLNKLNSSSLVYRLEDIPSVLPPTHETDPALIMRDRLRRLCFAR HHDTFLEVVDVFGMRQIVTQAGEPIHLVTDYGNVAFKYLALRDDGRPLAWRRRCSGGGLKNVVTTRYKAIT VAVAVCQTLRTFWPQISQYDLRPYLTQHQSHTHPAETHTLHNLKLFCYLVSTAWHQRIDTQQELTAADRVG SGEGGDVGEQRPGRGTVLRLSLQEFCVLIAALYPEYIYTVLKYPVQMSLPSLTAHLHQDVIHAVVNNTHKMP PDHLPEQVKAFCITPTQWPAMQLNKLFWENKLVQQLCQVGPQKSTPPLGKLWLYAMATLVFPQDMLQCL WLELKPQYAETYASVSELVQTLFQIFTQQCEMVTEGYTQPQLPTGEPVLQMIRVRRQDTTTTDTNTTTEPGL LDVFIQTETALDYALGSWLFGIPVCLGVHVADLLKGQRILVARHLEYTSRDRDFLRIQRSRDLNLSQLLQDTW TETPLEHCWLQAQIRRLRDYLRFPTRLEFIPLVIYNAQDHTVVRVLRPPSTFEQDHSRLVLDEAFPTFPLYDQD DNSSADNVAASGAAPTPPVPFNRVPVNIQFLRENPPPIARVQQPPRRHRHRAAAAADDDGQIDHVQDDTS RTADSALVSTAFGGPVFQENRLGETPLCRDELVAVAPGAASTSFASPPITVLTQNVLSALEILRLVRLDLRQLA QSVQDTIQHMRFLYLL
UL48A	Reverse	Small capsid protein	Located externally on capsid hexons; involved in capsid morphogenesis; possibly involved in capsid transport	75	MSNTAPGPTVANKRDEKHRHVVNVVLELPTEISEATHPVLATMLSKYTRMSSLFNDKCAFKLDLLRMIAVSR TRR

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL49	Reverse	Protein UL49		570	MANRRLRHAPHTATDEFHQALRRLFAPLCVHEDHFHVQLVIGRGALQPEEAAVETSQPPAQFAAQTSAVL QQQLVHHVPRSCVLHLFVTDKRFLNRELGDRLYQRFLREWLVCRQAEREAVTALFQRMVMTKPYFVFLAYV YSMDCLHTVAVRTMAFLRFERYDTDYLLRRLRLYPPERLHALLDGVTASLLGDLHRFLFGVDLRLPVLHPTSSP CLALLRAKRFDARADLAVYHRNQWRHQRQPRSPQLRGLIAALRRHAGQVPCGNPLYVLARQAVQTFCDTC PRYLVPLRALGLHDETRGGGSTAAAAAVGHAGAGQQARHVEPTKIVLFALSAALRGGLIGSVIDLPLWCLCR LKCERHLDARSLVAVVCRQCGHCLNLGKEKLHCQQNFPLNSMFYYRDRQEKSVIFNTHAELVHCSLCGSQR VVRQRVYELVSETLFGQRCVRVGWKAVLGLNAACAVYDHRLAFDVILPCAARTCDSTVVVRGVTVPRLLRLT SHGHGLLCARCQTGEYRDSCLESEDGAPLCRGCALVKQTACHVGGHIVQQARGGLAATSSSSSPQGLPHV
UL50	Reverse	Nuclear egress membrane protein	Type 2 membrane protein; interacts with nuclear egress lamina protein; involved in nuclear egress	397	MEINKVLHQDLVQATRRILKLGPSELRVTDAGLICKNPNYSVCDAMLKTDTVYCVEYLLSYWESRTDHVPCFI FKNTGCAVSLCCFVRAPVKLVSPARHVGEFNVLKVNESLIVTLKDIEEIKPSAYGVLTKCVVRKSNSASVFNIELI AFGPENEGEYENLLRELYAKKAASTSLAVRNHVTVSSHSGSGPSLWRARMSAALTRTAGKRSSRTASPPPPP RHPSCSPTMVAAGGAAAGPRPPPPPMAAGSWRLCRCEACMGRCGCASEGDADEEEEELLALAGEGKAAA AAAGQDVGGSARRPLEEHVSRRRGVSTHHRHPPSPPCAPSLERTGYRWAPSSWWRARSGPSRPQSGPWL PARFAILGPLVLALLLVLALLWRGHGQSSSPTRSAHRD
UL51	Reverse	DNA packaging protein UL33	Interacts with DNA packaging terminase subunit 2; involved in DNA encapsidation	157	MSWAKQRVPFLEDDDGEEENDVQDDVDSPVPTRPLVIDEDAEPAAGTSGGFEGGGGDDEDGEDGHALP DLDDDLLLQFEPMLPRVYDLLLPSLDARLNFVNAGQKYAAFLKYVHGDCATCSHGEILREKTQLLTAIVSKLM DINGILEGKDEPAPGK
UL52	Forward	DNA packaging protein UL32	Involved in DNA encapsidation; possibly involved in capsid transport	668	MNPSTHVNSNGPTTPPHGPHTTFLPPTSPAPSTSSVAAATLCSPQRQAVSRYSGWSTEYTQWHSDLTTELL WHAHPRQVPMDEALAAAAAASYQVNPQHPANRYRHYEFQTLSLGTSEVDELLNCCAEETTCGGTQSTVLT NATNTTSCGGAVAGSSNAGPAGASAACDLDAELAGLETSAADFEQLRRLCAPLAIDTRCNLCAIISICLKQDC DQSWLLEYSLLCFKCSYAPRAALSTLIIMSEFTHLLQQHFSDLRIDDLFRHHVLTVFDFHLHFFINRCFEKQVGD AVDNENVTLNHLAVVRAMVMGEDTVPYNKPRRHPQQKQKNNPYHVEVPQELIDNFLEHSSPSRDRFVQL LFYMWAGTGVMSTTPLTELTHTKFARLDALSTASEREDARMMIEEEEDEEGGEKGGDDPGRHNGGGTSG GFSESTLKKNVGPIYLCPVPAFFTKNQTSTVCLLCELMACSYYDNVVLRELYRRVVSYCQNNVKMVDRIQLVL ADLLRECTSPLGAAHEDVARCGLEAPTSPGGDLDYHGLSGVDGALARPDPVFCHVLRQAGVTGIYKHFFCDP QCAGNIRVTNEAVLFGRLHPHHVQEVKLAICHDNYYISRLPRRVWLCITLFKAFQITKRTYKGKVHLADFMRD FTQLLESCDIKLVDPTYVIDKYV
UL53	Forward	Nuclear egress lamina protein	Interacts with nuclear egress membrane protein; involved in nuclear egress	376	MSSVSGVRTPRERRSALRSLLRKRRQRELASKVASTVNGATSANNHGEPPSPADARPRLTLHDLHDIFREHP ELELKYLNMMKMAITGKESICLPFNFHSHRQHTCLDISPYGNEQVSRIACTSCEDNRILPTASDAMVAFINQT SNIMKNRNFYYGFCKSSELLKLSTNQPPIFQIYYLLHAANHDIVPFMHAEDGRLHMHVIFENSDVHIPCDCIT QMLTAAREDYSVTLNIVRDHVVISVLCHAVSASSVKIDVTILQRKIDEMDIPNDVSESFERYKELIQELCQSSCS NLYEEATSSYAIRSPLTASPLHVVSTNGCGPSSSSQSTPPHLHPPSQATQPHHYSHHQSQSQQHHHHPQSPP PPLFLNSIRAP
UL54	Reverse	DNA polymerase catalytic subunit	Involved in DNA replication	1242	MFFNPYLSGGVTGGAVAGGRRQRSQPGSAQGSGKRPPQKQFLQIVPRGVMFDGQTGLIKHKTGRLPLMF YREIKHLLSHDMVWPCPWRETLVGRVVGPIRFHTYDQTDAVLFFDSPENVSPRYRQHLVPSGNVLRFFGAT EHGYSICVNVFGQRSYFYCEYSDTDRLREVIASVGELVPEPRTPYAVSVTPATKTSIYGYGTRPVPDLQCVSISN WTMARKIGEYLLEQGFPVYEVRVDPLTRLVIDRRITTFGWCSVNRYDWRQQGRASTCDIEVDCDVSDLVAV PDDSSWPRYRCLSFDIECMSGEGGFPCAEKSDDIVIQISCVCYETGGNTAVDQGIPNGNDGRGCTSEGVIFG

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					HSGLHLFTIGTCGQVGPDVDVYEFPSEYELLLGFMLFFQRYAPAFVTGYNINSFDLKYILTRLEYLYKVDSQRFC KLPTAQGGRFFLHSPAVGFKRQYAAAFPSASHNNPASTAATKVYIAGSVVIDMYPVCMAKTNSPNYKLNTM AELYLRQRKDDLSYKDIPRCFVANAEGRAQVGRYCLQDAVLVRDLFNTINFHYEAGAIARLAKIPLRRVIFDG QQIRIYTSLLDECACRDFILPNHYSKGTTVPETNSVAVSPNAAIISTAAVPGDAGSVAAMFQMSPPLQSAPSS QDGVSPGSGSNSSSSVGVFSVGSGSSGGVGVSNDNHGAGGTAAVSYQGATVFEPEVGYYNDPVAVFDFAS LYPSIIMAHNLCYSTLLVPGGEYPVDPADVYSVTLENGVTHRFVRASVRVSVLSELLNKWVSQRRAVRECMR ECQDPVRRMLLDKEQMALKVTCNAFYGFTGVVNGMMPCLPIAASITRIGRDMLERTARFIKDNFSEPCFLH NFFNQEDYVVGTREGDSEESSTLPEGLETSSGGLDERRVEARVIYGDTDSVFVRFRGLTPQALVARGPSLAHY VTACLFVEPVKLEFEKVFVSLMMICKKRYIGKVEGASGLSMKGVDLVRKTACEFVKGVTRDVLSLLFEDREVS EAAVRLSRLSLDEVKKYGVPRGFWRILRRLVQARDDLYLHRVRVEDLVLSSVLSKDISLYRQSNLPHIAVIKRLA ARSEELPSVGDRVFYVLTAPGVRTAPQGSSDNGDSVTAGVVSRSDAIDGTDDDADGGGVEESNRRGGEPA KKRARKPPSAVCNYEVAEDPSYVREHGVPIHADKYFEQVLKAVTNVLSPVFPGGETARKDKFLHMVLPRRLH LEPAFLPYSVKAHECC
UL55	Reverse	Envelope glycoprotein B	Type 1 membrane protein; possible membrane fusogen; binds cell surface heparan sulphate; involved in cell entry; involved in cell-to-cell spread	907	MESRIWCLVVCVNLCIVCLGAAVSSSSTRGTSATHSHHSSHTTSAAHSRSGSVSQRVTSSQTVSHGVNETIYN TTLKYGDVVGVNTTKYPYRVCSMAQGTDLIRFERNIVCTSMKPINEDLDEGIMVVYKRNIVAHTFKVRVYQK VLTFRRSYAYIHTTYLLGSNTEYVAPPMWEIHHINSHSQCYSSYSRVIAGTVFVAYHRDSYENKTMQLMPDD YSNTHSTRYVTVKDQWHSRGSTWLYRETCNLNCMVTITTARSKYPYHFFATSTGDVVDISPFYNGTNRNAS YFGENADKFFIFPNYTIVSDFGRPNSALETHRLVAFLERADSVISWDIQDEKNVTCQLTFWEASERTIRSEAED SYHFSSAKMTATFLSKKQEVNMSDSALDCVRDEAINKLQQIFNTSYNQTYEKYGNVSVFETTGGLVVFWQGI KQKSLVELERLANRSSLNLTHNRTKRSTDGNNATHLSNMESVHNLVYAQLQFTYDTLRGYINRALAQIAEAW CVDQRRTLEVFKELSKINPSAILSAIYNKPIAARFMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVV IFNFANSSYVQYGQLGEDNEILLGNHRTEECQLPSLKIFIAGNSAYEYVDYLFKRMIDLSSISTVDSMIALDIDPL ENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYVEDKVVDPLPPYLKGLDDLMSGLGAAGKAVGV AIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIIIYLIYTRQRRLCMQPLQNLFPYLVSADGTTVTSGNT KDTSLQAPPSYEESVYNSGRKGPGPPSSDASTAAPPYTNEQAYQMLLALVRLDAEQRAQQNGTDSLDGQT GTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV
UL56	Reverse	DNA packaging terminase subunit 2	Involved in DNA encapsidation	850	MEMNLLQKLCVVCSKCNEYAMELECLKYCDPNVLLAESTPFKRNAAAIVYLYRKIYPEVVAQNRTQSSLLTLY LEMLLKALHEDTALLDRALMAYSRQPDRAAFYRTVLRLDRCDRHHTVELQFTDNVRFSVSLATLNDIERFLCK MNYVYGILAPEAGLEVCAQLLELLRRLCGISPVARQEVYVEGTTCAQCYEELTIIPNQGRSLNKRLQGLLCNHI AVHRPSSQSDVNIQTVEQDLLDLTTRIPHLAGVLSALKSLFSSSSAYHSYIQEAEEALREYNLFTDIPERIYSLSDF TYWSRTSEVIVKRVGITIQQLNVYHQLCRALMNGISRHLYGEDVEDIFVLGEKALDGEERMFVGSVFAAPNRI IDLITSLSIQAFEDNPVFNKLHESNEMYTKIKHILEEIRRPLPDGTGGDGPEGEAIHLRGREAMSGTGTTLMTA SNSSNSSTHSQRNNGGGGRARGGGKKAVGGGANGQDGDGSENGLRVRNCDEHEALDLVDARSRIHNVT REVNVRKRAYLQKVSEVGYGKVIRCIKTQERLTSKLIDVNLVGPLCLDFISKLMNGFLYRSQYHQDQDVVDVG DQFTYDEHLYVVNNLIHKSLPVESLPLLGQQIYELCNGPLFTHCTDRYPLSHNVDMAYACDNAGVLPHVKDD LVKCAEGTVYPSEWMVVKYMGFFNFSDCQDLNVLQKEMWMHVRELVLSVALYNETFGKQLSIACLRDELH PDRDVILTYNKEWPLLLRHEGSLYKSKDLYLLLYRHLSRPDESGDVPTAPAAKPSTLTAAAAVSGAFREPDRP WLPSPYPSSSTAGVSRRVRATRKRPRRASSLLDLARDEHGIQDLVPGSLR

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL57	Reverse	Single-stranded DNA-binding protein	Contains zinc-finger; involved in DNA replication; possibly involved in gene regulation	1235	MSHEELTALAPVGPAAFLYFSRLNAETQEILATLSLCDRSSSVVIAPLLAGLTVEADFGVSVRTPVLCYDGGVLT KVTSFCPFALYFHHTQGIVAFTEDHGDVHRLCEDARQKYALEAYTPEADRVPTDLAALCAAVGCQASETTVH VVVGNGLKEFLFAGQLIPCVEEATTVRLHGGEAVRVPLYPPTLFNSLQLDAEADEVSLDARSAFVEARGLYVP AVSETLFYYVYTSWCQSLRFSEPRVLIEAALRQFVHDSQQSVKLAPHKRYLGYMSQRLSSLEKDHLMLSDAVV CELAFSFASVFFDSAYQPAESMLFSEWPLVTNATDHRDLIRALTELKLHLSTHVAALVFSANSVLYQHRLVYLQ SSARHPSAGGTASQETLLKAIQFTNGLSAACEDVYNDARKVLKFQGAPLKDERYGPQHLALVCGTCPQLVSG FVWYLNRVSVYNTGLSGSSTLTNHLVGCAAGLCEACGGTCCHTCYQTAFVRVRTRLPVVPKQPKKEPCVITV QSRFLNDVDILGSFGRRYNVDAKDGGLDGKGDDGVPGGGAGGGGGRDVSGGPSDGLGGGRGGGGGD SGGMMGRGGRMLGASVDRTYRLNRILDYCRKMRLIDPVTGEDTFSAHGKSDFVAVFSALNKFVDDEALGF VSEVRLKSSRDEVAGATQAFNLDLNPYAVAFQPLLAYAYFRSVFYVIQNVALITATSYIVDNPLTTNLVSKWM TQHFQSIHGAFSTTSSRKGFLFTKQIKSSKNSDHDRLLDFRLYAQGTYAVVPMEIKLSRLSVPTLIMVRVKNRPI YRAGKGNAGSVFFRRDHVPRRNPAKGCLGFLLYRHHERLFPECGLPCLQFWQKVCSNALPKNVPIGDMGEF NAFVKFLVAVTADYQEHDLLDVAPDCVLSYVESRFHNKFLCYYGFKDYIGSLHGLTTRLTTQNHAQFPHVLG ASPRFSSPAEFALHVKGLKTAGVPAPMAATVARESLVRSVFEHRSLVTVPVSVEKYAGINNSKEIYQFGQIGYF SGNGVERSLNVSSMSGQDYRFMRQRYLLATRLADVLIKRSRRENVLFDADLIKNRVMLALDAENLDCDPEV MAVYEILSVREEIPASDDVLFFVDGCEALAASLMDKFAALQEQGVEDFSLENLRRVLDADAQRLTDAAGGEV HDLSALFAPSGVGTASGVGGGGLLLGESVAGNSICFGVPGETGGGCFLVNAGEDEAGGVGGSSGGGGGSG LLPAKRSRL
UL69	Reverse	Multifunctional expression regulator	RNA-binding protein; shuttles between nucleus and cytoplasm; inhibits pre-mRNA splicing; exports virus mRNA from nucleus; exerts most effects post-transcriptionally; involved in gene regulation; involved in RNA metabolism and transport	742	MELHSRGRHDAPSLSSLSERERRARRARRFCLDYEPVPRKFRRERSPTSPSTRNGAAASEYHLAEDTVGAASH HHRPCVPARRPRYSKDDDTEGDPDHYPPPLPPSSRHALGGTGGHIIMGTAGFRGGHRASSSFKRRVAASAS VPLNPHYGKSYDNDDGEPHHHGGDSTHLRRRVPSCPTTFGSSHPSSANNHHGSSAGPQQQQMLALIDDEL DAMDEDELQQLSRLIEKKKRARLQRGAASSGTSPSSTSPVYDLQRYTAESLRLAPYPADLKVPTAFPQDHQP RGRILLSHDELMHTDYLLHIRQQFDWLEEPLLRKLVVEKIFAVYNAPNLHTLLAIIDETLSYMKYHHLHGLPVN PHDPYLETVGGMRQLLFNKLNNLDLGCILDHQDGWGDHCSTLKRLVKKPGQMSAWLRDDVCDLQKRPPE TFSQPMHRAMAYVCSFSRVAVSLRRRALQVTGTPQFFDQFDTNNAMGTYRCGAVSDLILGALQCHECQNE MCELRIQRALAPYRFMIAYCPFDEQSLLDLTVFAGTTTTTANNHATAGGQQRGGDQIHPTDEQCANMESR TDPATLTAYDKKDREGSHRHPSPMIAAAPPAQPPSQPQQHYSEGELEEDEDSDDASSQDLVRATDRHGDT VVYKTTAVPPSPPAPLAGVRSHRGELNLMTPSPSHGGSPPQVPHKQPIIPVQSANGNHSTAATQQHQPPPP PPPVPQEDDSVVMRCQTPDYEDMLCYSDDMDD
UL70	Reverse	Helicase-primase primase subunit	Involved in DNA replication	946	MTLVLFATEYDSAHIVANVLSQTPTDHCVFPLLVKHQVSRRVYFCLQTQKCSDSRRVAPVFAVNNETLQLSR YLAARQPIPLSALIASLDEAETQPLYRHLFRTPVLSPEHGGEVREFKHLVYFHHAAVLRHLNQVFLCPTSPSWFI SVFGHTEGQVLLTMAYYLFEGQYSTISTVEEYVRSFCTRDLGTIIPTHASMGEFARLLLGSPFRQRVSAFVAYA VARNRRDYTELEQVDTQINAFRERARLPDTVCVHYVYLAYRTALARARLLEYRRVVAYDADAAPEAQCTREP GFLGRRLSTELLDVMQKYFSLDNFLHDYVETHLLRLDESPHSATSPHGLGLAGYGGRIDGTHLAGFFGTSTQL ARQLERINTLSESVFSPLERSLSGLLRLCASLRTAQTYTTGTLTRYSQRRYLLPEPALAPLLERPLPVYRVHLPND QHVFCAVASETWHRSLFPRDLLRHVPDSRFSDEALTETVWLHDDDVASTSPETQFYYTRHEVFNERLPVFNF VADFDLRLRDGVSGLARHTVFELCRGLRRVWMTVWASLFGYTHPDRHPVYFFKSACPPNSVPVDAAGAPF DDDDYLDYRDERDTEEDEDGKENKNNVPDNGVFQKTTSSVDTSPPYCRCKGKLGLRIITPFPACTVAVHPSV LRAVAQVLNHAVCLDAELHTLLDPISHPESSLDTGIYHHGRSVRLPYMYKMDQDDGYFMHRRLLPLFIVPDA

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					YREHPLGFVRAQLDLRNLLHHHPPHDLPALPLSPPPRVILSVRDKICPSTEANFIETRSLNVTRYRRRGLTEVLA YHLYGGDGATAAAISDTDLQRLVVTRVWPPLLEHLTQHYEPHVSEQFTAPHVLLFQPHGACCVAVKRRDGA RTRDFRCLNYTHRNPQETVQVFIDLRTEHSYALWASLWSRCFTKKCHSNAKNVHISIKIRPPDAPVPPATAV
UL71	Forward	Tegument protein UL51	Involved in virion morphogenesis	361	MQLAQRLCELLMCRRKAAPVADYVLLQPSEDVELRELQAFLDENFKQLEITPADLRTFSRDTDVVNHLLKLLP LYRQCQSKCAFLKGYLSEGCLPHTRPAAEVECKKSQRILEALDILILKLVVGEFAMSEADSLEMLLDKFSTDQAS LVEVQRVMGLVDMDCEKSAYMLEAGAAATVAPPTPPAVVQGESGVREDGETVAAVSAFACPSVSDSLIPE ETGVTRPMMSLAHINTVSCPTVMRFDQRLLEEGDEEDEVTVMSPSPEPVQQQPPVEPVQQQPQGRGSHR RRYKESAPQETLPTNHEREILDLMRHSPDVPREAVMSPTMVTIPPPQIPFVGSARELRGVKKKKPTAAALLSS A
UL72	Reverse	Deoxyuridine triphosphatase	Enzymatically inactive; DURP family	388	MLTMLTDRIDSQLVLSRLPRSRFQRFWETPTLIMKEESASSSGSIILAEKSVNMRYCVRFASDSDFQTTFTLPQ STEEKYDKEQHPGEDEASSPLPSPLKVPYKWMPSSFIVKQCHTQLAFYNKHIIWLSRERKVPTSLGVSLYIPEG FFGITFYKCLDAQFVCMPELLESGLQVPQLDVVNLNDTFQSIFPGTIEGDIGVFPCFVPEPWQLMNLPPPNE HRFFSLRTRQTLVIGPGHTQTVYFDAAYVHAPGICALIVGVRQFSQSDLIIRPTIWLPGTAAGVTVVNTSHTTV CISPHTTVAKAVFTTHRFTYLPVGSHPLGQMIVPPTPDIGFTHTPEHALLQRTPSPVDDDVDETEEDEKSSDA ESPVNTNDVIFDVGPKPPRHP
UL73	Forward	Envelope glycoprotein N	Type 1 membrane protein; complexed with envelope glycoprotein M; involved in virion morphogenesis; involved in membrane fusion	135	MECNTLVLGLLVLSVVASSNNTSTASTPRPSSSTHASTTVKATTVATTSTTTATSTSSTTSAKPGSTTHDPNV MRPHAHNDFYNAHCTSHMYELSLSSFAAWWTMLNALILMGAFCIVLRHCCFQNFTATTTKGY
UL74	Reverse	Envelope glycoprotein O	Contains signal peptide; associated with envelope glycoprotein H and envelope glycoprotein L; involved in virion morphogenesis	467	MIMVKGIPKIMLLISITFLLLSLINCNVLVNSRGTRRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYRFLMY PSQQKFHAIVISMDKFPQDYILAGPIRNDSITHMWFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVPS MNCLSEMLNVSKRNDTGEKGCGNFTTFNPMFFNVPRWNTKLYIGSNKVNVDSQTIYFLGLTALLLRYAQRN CTRSFYLVNAMSRNLFRVPKYINGTKLKNTMRKLKRKQALVKEQPQKKNKKSQSTTTPYLSYTTSTAFNVTTN VTYSATAAVTRVATSTTGYRPDSNFMKSIMATQLRDLATWVYTTLRYRNEPFCKPDRNRTAVSEFMKNTHV LIRNETPYTIYGTLDMSSLYYNETMSVENETASDNNETTPTSPSTRFQRTFIDPLWDYLDSLLFLDKIRNFSLQL PAYGNLTPPEHRRAANLSTLNSLWWWSQ
UL74A	Forward	Envelope glycoprotein 24	Type 2 membrane protein	70	MALTRGGDPAPSVCLVVVWLACVYSLLILVVLLLIYRCCIGFQDDLVSRTLAVYRACIQGPICNQTHNSTS
UL75	Reverse	Envelope glycoprotein H	Type 1 membrane protein; possible membrane fusogen; complexed with envelope glycoprotein L; involved in cell entry; involved in cell-to-cell spread	742	MRPGLPSYLIVLAVCLLSHLLSSRYGAKAVSEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENA ISFNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRLNTYALVSKDLASYRSFSQQLKAQDSL GEQPTTVPPPIDLSIPHVWMPPQTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLID ELRYVKITLTEDFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYLKDP DFLDAALDFNYLDLSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRA LDRQAALLQIQEFMITCLSQTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWAL RQIADFALKLHKTHLASFLSAFARQELYLMGSLVHSMLVHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYL SDLYTPCSSSGRRDHSLERLTRLFPDATVPATVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHVSYVV TNQYLIKGISYPVSTTVVGQSLIITQTDSQTKCELTRNMHTTHSITAALNISLENCAFCQSALLEYDDTQGVINI MYMHDSDDVLFALDPYNEVVVSSPRTHYLMLLKNGTVLEVTDVVVDATDSRLLMMSVYALSAIIGIYLLYR MLKTC

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL76	Forward	Nuclear protein UL24		325	MPSGRGDDADSTGNALRRLPHVRKRIGKRKHLDIYRRLLRVFPSFVALNRLLGGLFPPELQKYRRRLFIEVRLS RRIPDCVLVFLPPDSGSRGIVYCYVIEFKTTYSDADDQSVRWHATHSLQYAEGLRQLKGALVDFDFLRLPRGG GQVWSVVPSLVFFQQKADRPSFYRAFRSGRFDLCTDSVLDYLGRRQDESVAHLLAATRRRLLRAARGKRAAL PRARASAVVGGRGGGNARRGLARGRAHGPGAQTVSASGAEGSGSQGADLLRGSRRARVRGGGAVEPAV RARRRTVAADAATTTVSSAFVVPRDRRGRSFRRPTRSL
UL77	Forward	DNA packaging tegument protein UL25	Located on capsid near vertices; possibly stabilizes the capsid and retains the genome; involved in DNA encapsidation	635	MSLLHTFWRLPVAVFFEPHEENVLRCPERVLRRLLEDAAVAMRGGGWREDVLMDRVRKRYLRQELRDLGH RVQTYCEDLEGRVSEAEALLNQQCELDEGPSPRTLLQPPCRPRSSSPGTGVAGASAVPHGLYSRHDAITGPA TPSDAATASAAAGASSTWLAQCAERPLPGNVPSYFGITQNDPFIRFHTDFRGEVVNTMFENASTWTFSFGI WYYRLKRGLYTQPRWKRVYHLAQMDNFSISQELLLGVVNALENVTVYPTYDCVLSDLEAAACLLAAYGHAL WEGRDPPDSVATVLGELPQLLPRLADDVSREIAAWEGPVAAGNNYYAYRDSPDLRYYMPLSGGRHYHPGT FDRHVLVRLFHKRGVIQHLPGYGTITEELVQERLSGQVRDDVLSLWSRRLLVGKLGRDVPVFVHEQQYLRSG LTCLAGLLLLWKVTNADSVFAPRTGKFTLADLLGSDAVAGGGLPGGRAGGEEEGYGGRHGRVRNFEFLVQY YIGPWYARDPAVTLSQLFPGLALLAVTESVRSGWDPSRREDSAGGGDGGGAVLMQLSKSNPVADYMFAQ SSKQYGDLRRLEVHDALLFHYEHGLGRLLSVTLPRHRVSTLGSSLFNVNDIYELLYFLVLGFLPSVAVL
UL78	Forward	Envelope protein UL78	Type 3 membrane protein; 7 transmembrane domains; putative chemokine receptor; possibly involved in intracellular signalling; GPCR family	431	MSSSAEKTTSVTDSIMLAIVNFKYMGPFEGYSMSADRAASDLLIGMFGSVSLVNLLTIIGCLWVLRVTRPPVS VMIFTWNLVLSQFFSIVATMLSKGIMLRGALNLSFCRLVLFVDDVGLYSTALFFLFLILDRLSAISYGRDLWHH ETRENAGVALYAVAFAWVLSIVAAVPTAATGSLDYRWLGCQIPIQYAAVDLTIKMWFLLGAPMIAVLANVV ELAYSDRRDHVWSYVGRVCIFYVTCLMLFVPYYCFRVLRGVLQPASAAGTGFGIMDYVELATRTLLTMRLGIL PLFIIAFFSREPTKDLDDSFDYLVERCQQSCHGHFVRRLVQALKRAMYSVELAVCYFSTSVRDVAEAVKKSSSR CYADATSAAVVVTTTTSEKATLVEHAEGMASEMCPGTTIDVSAESSSVLCTDGENTVATDATVTAL
UL79	Reverse	Protein UL79	Promotes accumulation of late transcripts	295	MMARDEENPAVPRVRTGKFSFTCANHLILQISEKMSRGQPLSSLRLEELKIVRLICVLLFHRGLETLLLRETMN NLGVSDHAVLSRKTPQPYWPHLYRELRQAFPGLDFEAAVFDETRAARLSQRLCHPRLSGGLLTRFVQRHTGL PVVFPEDLARNGNILFSLGTLYGHRLFRLAAFFTRHWGAEAYEPLIRIICQKMWYFYLIGTGKMRITPDAFEIQ RSRHETGIFTFIMEDYRTFAGTLSRHPHRPHPQQQQHHHPGPPHPPLSHPASSCLNPEAVLAARALHMPTL ANDV
UL80	Forward	Capsid maturation protease	Serine protease (N-terminal region); minor scaffold protein (remainder of protein, clipped near C terminus); involved in capsid morphogenesis	708	MTMDEQQPQAVTPVYVGGFLARYDQSPDEAELLLPRDVVEHWLHAQGQGQPSLSVALPLNINHDDTAVV GHVAAMQSVRDGLFCLGCVTSPRFLEIVRRASEKSELVSRGPVSPLQPDKVVEFLSGSYAGLSLSSRRCDDVE AATSLSGSETTPFKHVALCSVGRRRGTLAVYGRDPEWVTQRFPDLTAADRDGLRAQWQRCGSTAVDASGD PFRSDSYGLLGNSVDALYIRERLPKLRYDKQLVGVTERESYVKASVSPEAACDIKAASAERSGDSRSQAATPAA GARVPSSSPSPPVEPPSPVQPPALPASPSVLPAESPPSLSPSEPAEAASMSHPLSAAVTAATAPPGATVAGAS PAVPSLAWPHDGVYLPKDAFFSLLGASRSAAPVMYPGAVAAPPSASPAPLPLPSYPASYGAPVVGYDQLAA RHFADYVDPHYPGWGRRYEPAPSLHPSYPVPPPPSPAYYRRRDSPGGMDEPPSGWERYDGSHRGQSQKQ HRHGGSGGHNKRRKEAAAASSSSSDEDLSFPGEAEHGRARKRLKSHVNSDGGSGGHAGSNQQQQQRYDE LRDAIHELKRDLFAARQSSTLLSAALPAAASSSPTTTTVCTPTGELTSGGGETPTALLSGGAKVAERAQAGVV NASCRLATASGSEAATAGPSTAGSSSCPASVVLAAAAAQAAAASQSPPKDMVDLNRRIFVAALNKLE
UL80.5	Forward	Capsid scaffold protein	Clipped near C terminus; involved in capsid morphogenesis	373	MSHPLSAAVTAATAPPGATVAGASPAVPSLAWPHDGVYLPKDAFFSLLGASRSAAPVMYPGAVAAPPSAS PAPLPLPSYPASYGAPVVGYDQLAARHFADYVDPHYPGWGRRYEPAPSLHPSYPVPPPPSPAYYRRRDSPG GMDEPPSGWERYDGSHRGQSQKQHRHGGSGGHNKRRKEAAAASSSSSDEDLSFPGEAEHGRARKRLKSH

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					VNSDGGSGGHAGSNQQQQQRYDELRDAIHELKRDLFAARQSSTLLSAALPAAASSSPTTTTVCTPTGELTSG GGETPTALLSGGAKVAERAQAGVVNASCRLATASGSEAATAGPSTAGSSSCPASVVLAAAAAQAAAASQSP PKDMVDLNRRIFVAALNKLE
UL82	Reverse	Tegument protein pp71	Upper matrix protein; involved in gene regulation; transcriptional activator; targeted to ND10; targets Rb proteins for ubiquitin-independent proteosomal degradation; DURP family	559	MSQASSSPGEGPSSEAAAISEAEAASGSFGRLHCQVLRLITNVEGGSLEAGRLRLLDLRTNIEVSRPSVLCCFQ ENKSPHDTVDLTDLNIKGRCVVGEQDRLLVDLNNFGPRRLTPGSENNTVSVLAFALPLDRVPVSGLHLFQSQ RRGGEENRPRVEARAIIRRTAHYWAVRLTVTPNWRRRTDSSLEAGQIFVSQFAFRAGAIPLTLVDALEQLACS DPNTYIHKTETDERGQWIMLFLHHDSPHPPTSVFLHFSVYTHRAEVVARHNPYPHLRRLPDNGFQLLIPKSFT LTRIHPEYIVQIQNAFETNQTHDTIFFPENIPGVSIEAGPLPDRVRITLRVTLTGDQAVHLEHRQPLGRIHFFRR GFWTLTPGKPDKIKRPQVQLRAGLFPRSNVMRGAVSEFLPQSPGLPPTEEEEEEEEEDDEDDLSSTPTPTPLS EAMFAGFEEASGDEDSDTQAGLSRALILTGQRRRSGNNGALTLVIPSWHVFASLDDLVPLTVSVQHAALRPT SYLRSDMDGDVRTAADISSTLRSVPAPRPSPISTASTSSTPRSRPRI
UL83	Reverse	Tegument protein pp65	Lower matrix protein; involved in immune regulation; suppresses interferon response; inhibits NK cell cytotoxicity; binds NKp30; DURP family	561	MESRGRRCPEMISVLGPISGHVLKAVFSRGDTPVLPHETRLLQTGIHVRVSQPSLILVSQYTPDSTPCHRGDN QLQVQHTYFTGSEVENVSVNVHNPTGRSICPSQEPMSIYVYALPLKMLNIPSINVHHYPSAAERKHRHLPVA DAVIHASGKQMWQARLTVSGLAWTRQQNQWKEPDVYYTSAFVFPTKDVALRHVVCAHELVCSMENTRA TKMQVIGDQYVKVYLESFCEDVPSGKLFMHVTLGSDVEEDLTMTRNPQPFMRPHERNGFTVLCPKNMIIKP GKISHIMLDVAFTSHEHFGLLCPKSIPGLSISGNLLMNGQQIFLEVQAIRETVELRQYDPVAALFFFDIDLLLQR GPQYSEHPTFTSQYRIQGKLEYRHTWDRHDEGAAQGDDDVWTSGSDSDEELVTTERKTPRVTGGGAMAG ASTSAGRKRKSASSATACTAGVMTRGRLKAESTVAPEEDTDEDSDNEIHNPAVFTWPPWQAGILARNLVP MVATVQGQNLKYQEFFWDANDIYRIFAELEGVWQPAAQPKRRRHRQDALPGPCIASTPKKHRG
UL84	Reverse	Protein UL84	Involved in gene regulation; involved in DNA replication; DURP family	587	MPRVDPNLRNRARRPRARRGGGGGVGSNSSRHSGKCRRQRRALSAPPLTFLATTTTTTMMGVASTDDDS LLLKTPDELDKHSGSPQTILTLTDKHDIRQPRVHRGTYHLIQLHLDLRPEELRDPFQILLSTPLQLGEANGESQT APAMSQDEEAAISHEPEKKKEKEEKKEEEDEDDRNDDRERGILCVVSNEDSDVRPAFSLFPARPGCHILRSVI DQQLTRMAIVRLSLNLFALRIITPLLKRLPLRRKAAHHTALHDCLALHLPELTFEPTLDINNVTENAASVADAAE STDADLTPTLTVRVRHALCWHRVEGGISGPRGLTSRISARLSETTAKTLGPSVFGRLELDPNESPPDLTLSSLTL YQDGMLRFNVTCDRTEAPADPVAFRLRLRRETVRRPFFSDAPLPYFVPPRSGAADEGLEVRVPYELTLKNSH TLRIYRRFYGPYLGVFVPHNRQGLKMPVTVWLPRSWLELTVLVSDENGATFPRDALLGRLYFISSKHTLNRGC LSAMTHQVKSTLHSRSTSHSPSQQQLSVLGASIALEDLLPMRLASPETEPQDCKLTENTTGNTSPVTLAMVC GDL
UL85	Reverse	Capsid triplex subunit 2	Complexed 2:1 with capsid triplex subunit 1 to connect capsid hexons and pentons; involved in capsid morphogenesis	306	MAAMEANIFCTFDHKLSIADVGKLTKLVAAVVPIPQRLHLIKHYQLGLHQFVDHTRGYVRLRGLLRNMTLTL MRRVEGNQILLHVPTHGLLYTVLNTGPVTWEKGDALCVLPPLFHGPLARENLLTLGQWELVLPWIVPMPLA LEINQRLLIMGLFSLDRSYEEVKAAVQQLQTITFRDATFTIPDPVIDQHLLIDMKTACLSMSMVANLASELTM TYVRKLALEDSSMLLVKCQELLMRLDRERSVGEPRTPARPQHVSPDDEIARLSALFVMLRQLDDLIREQVVFT VCDVSPDNKSATCIFKG
UL86	Reverse	Major capsid protein	6 copies form hexons, 5 copies form pentons; involved in capsid morphogenesis	1370	MENWSALELLPKVGIPTDFLTHVKTSAGEEMFEALRIYYGDDPERYNIHFEAIFGTFCNRLEWVYFLTSGLAA AAHAIKFHDLNKLTTGKMLFHVQVPRVASGAGLPTSRQTTIMVTKYSEKSPITIPFELSAACLTYLRETFEGTIL DKILNVEAMHTVLRALKNTADAMERGLIHSFLQTLLRKAPPYFVVQTLVENATLARQALNRIQRSNILQSFKA KMLATLFLLNRTRDRDYVLKFLTRLAEAATDSILDNPTTYTTSSGAKISGVMVSTANVMQIIMSLLSSHITKETV SAPATYGNFVLSPENAVTAISYHSILADFNSYKAHLTSGQPHLPNDSLSQAGAHSLTPLSMDVIRLGEKTVIME

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					NLRRVYKNTDTKDPLERNVDLTFFFPVGLYLPEDRGYTTVESKVKLNDTVRNALPTTAYLLNRDRAVQKIDFVDALKTLCHPVLHEPAPCLQAFTERGPPSEPAMQRLLECRFQQEPMGGAARRIPHFYRVRREVPRTVNEMKQDFVVTDFYKVGNITLYTELHPFFDFTHCQENSETVALCTPRIVIGNLPDGLAPGPFHELRTWEIMEHMRLRPPPDYEETLRLFKTTVTSPNYPELCYLVDVLVHGNADAFLLIRAFIARCIVNMFHTRQLLVFAHSYALVTLIAEHLADGALPPQLLFHYRNLVAVLRLVTRISALPGLNNGQLAEEPLSAYVNALHDHRLWPPFVTHLPRNMEGVQVVADRQPLNPANIEARHHGVSDVPRLGAMDADEPLFVDDYRATDDEWTLQKVFYLCLIPAMTNNRACGLGLNLKTLLVDLFYRPAFLLMPAATAVSTSGTTSKESTSGVTPEDSVAAQRQAVGEMLTELVEDVATDAHTPLLQACRELFLAVQFVGEHVKVLEVRAPLDHAQRQGLPDFISRQHVLYNGCCVVTAPKTLIEYSLPVPFHRFYSNPTICAALSDDIKRYVTEFPHYHRHDGGFPLPTAFAHEYHNWLRSPFSRYSATCPNVLHSVMTLAAMLYKISPVSLVLQTKAHIHPGFALTAVRTDTFEVDMLLYSGKSCTSVIINNPIVTKEERDISTTYHVTQNINTVDMGLGYTSNTCVAYVNRVRTDMGVRVQDLFRVFPMNVYRHDEVDRWIRHAAGVERPQLLDTETISMLTFGSMSERNAAATVHGQKAVCELILTPVTMDVNYFKITNNPRGRASCMLGVDPYDTEAATKAIYDHREADAQTFAATHNPWASQAGCLSDVLYNTRHRERLGYNSKFYSPCAQYFNTEEIIAANKTLFKTIDEYLLRAKDCIRGDTDTQYVCVEGTEQLIENPCRLTQEALPILSTTTLALMETKLKGGAGAFATSETHFGNYVVGEIIPLQQSMLFNS
UL87	Forward	Protein UL87	Promotes accumulation of late transcripts	941	MAGAAPRRLGCDALIVVGGSAMPRRVLHVPVHVRACNLTQELSTGEDARFCRPRPVNVERVRAVFAALYRACPVHVRTEPERVKLVLGRLLLGPVAVPCFCDGEVEGHGEHLVPTTQFCRGPLLYVHRRCCCGSVTAGRALSYHVLENHVATHVLRGLLSLTEWNRELPGLFCDCPGGGGASGTEERYAMACLPRDLSLHLDDYPYLMVEIGRVLSVSEVDDYVTAVSGYLGEAAAPRIQVHYKLLFGLNVRPQAPCALDATRDFFLLELQKLWLGVEYHHEVTSEFFGRVLAQLHRDRARVMMALRLPEQTVCHLSTFVLSRFKRQVLYFKLQVSYGKCRTGHADRSGGGGNGGSQGHHNLLCYRRLSVTFADTDTVWRNLFYVYYELARDLGSHGTEDRPVSRGYGVSCAPRTSRLSPSESTVVSANGHALSSTALRTTSAGHKLSLPRDPAADRVRRYVCIISRLMFARYGERWRKHRRRRSETGEEEEEETVESGETDATPPFDFTGQQLRRAYQEHRRRKHLAVQRYAPCRRKLIGGMEFAEVTGVSLDRIAVNAFNTNRVINMKAALSSIAASGLGVRAPRLPKNMTHSFVMYKHTFKEPACTVSTFVSNDAVYINSLNVNIRGSYPEFLYSLGVYRLHVNIDHFFLPAVVCNSNSSLDVHGLEDQAVIRSERSKVYWTTNFPCMISHTNNVNVGWFKAATAIVPRVSGADLEAILLKELSCIKNMRDVCIDYGLHRVFTQLELRNSYQIPFLAKQLVLFLRACLLKLHGREKRLQLDRLVFEAAQRGLFDYSKNLTAHTKIKHTCALIGSRLANNVPKILARNKKVKLDHLGRNANVLTVCRHVEAHKIPRTRLKVLVEVLGALQSISGTPHTREVIHQTLFRLCSAAAATSGLCSSPPPLCVSLSSSAPSVPTSVSVDGSSEPTSPRARFASR
UL88	Forward	Tegument protein UL88		429	MMEAAAAAAAAFRPEERPTPGWHDAALLMDDGTVREHAFRNGPLSQLIRRVLPPPPDAEDDVVFASELCFYCSGRFNRRSSVFSIYWQKHSDLVYALTGITHCAKLVVECGQLGSSRLRWRDGDAGGEERQGDDDSRDELYDVPGIYMIRVNDGGSTGPRHVIWPGTSVLWAPDVVITTVQRRISAARALVNTFRQYFFLLERRSHEELVLCPPEMEERLAPLLQSATRGDSDMFDGVVASAYHRLRMSNIPRSSARLLEHCVGLAGAKKLLLLDVPRLENYFLCQVCLYELDEDEMGEEMLGMLAGKPEDAAVSGASGGFLLHRKTMKLAACLCLLLNSLHLHQEALEALDPPPPRVEENDLVNVVLRRYYRSHGGVQARTLAAARALLADYAETFSPLGSFTRLGYDRLVSADAGVSRRHLVALLRA
UL89	Reverse	DNA packaging terminase subunit 1	Contains ATPase domain; involved in DNA encapsidation	674	MLRGDSAAKIQERYAELQKRKSHPTSCISTAFTNVAALCRKRYQMMHPELGLAHSCNEAFLPLMAFCGRHRDYNSPEESQRELLFHERLKSALDELTFRPCSEEQRASYQKLDALTELYRDPQFQQINNFMTDFKKWLDGGFSTAVEGDAKAIRLEPFQKNLLIHVIFFIAVTKIPVLANRVLQYLIHAFQIDFLSQTSIDIFKQKATVFLVPRRHGKTWFIIPIISFLLKHMIGISIGYVAHQKHVSQFVLKEVEFRCRHTFARDYVVENKDNVISIDHRGAKSTALFASCYNTNSIRGQNFHLLLVDEAHFIKKEAFNTILGFLAQNTTKIIFISSTNTTSDATCFLTRLNNAPFDMLNVVSYVCEEHL

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					HSFTEKGDATACPCYRLHKPTFISLNSQVRKTANMFMPGAFMDEIIGGTNKISQNTVLITDQSREEFDILRYST LNTNAYDYFGKTLYVYLDPAFTTNRKASGTGVAAVGAYRHQFLIYGLEHFFLRDLSESSEVAIAECAAHMIISV LSLHPYLDELRIAVEGNTNQAAAVRIACLIRQSVQSSTLIRVLFYHTPDQNHIEQPFYLMGRDKALAVEQFISR FNSGYIKASQELVSYTIKLSHDPIEYLLEQIQNLHRVTLAEGTTARYSAKRQNRISDDLIIAVIMATYLCDDIHAIR FRVS
UL91	Forward	Protein UL91		111	MNSLLAELNRLGVAHATTEDVFIFVDRLFQHFSFLFQAEESGPRRLELVASVFEHLTVECVNDILDACSHPDV NVAETSNSCRPCPSPVPSAPKTVSGAQTSCATPRAPVT
UL92	Forward	Protein UL92		201	MCDASGACDMRHVQNAFTEEIQLHSLYACTRCFRTHLCDLGSGCALVSTLEGSVCVKTGLVYEALYPVARSH LLEPIEEAALDDVNIISAVLSGVYSYLMTHAGRYADVIQEVVERDRLKKQVEDSIYFTFNKVFRSMHNVNRISV PVISQLFIQLIIGIYSKQTKYDACVIKVSRKKREDALLKQMRSEYGNAPVFGSGV
UL93	Forward	DNA packaging tegument protein UL17	Capsid-associated; involved in DNA encapsidation; involved in capsid transport	594	METHLYSDLAFEARFADDEQLPLHLVLDQEVLSNEEAETLRYVYYRNVDSAGRSAGRAPGGDEDDAPASDD AEDAVGGDRAFDRERRTWQRACFRVLPRPLELLDYLRQSGLTVTLEKEQRVRMFYAVFTTLGLRCPDNRLS GAQTLHLRLVWPDGSYRDWEFLARDLLREEMEANKRDRQHQLATTTNHRRRGGLRNNLDNGSDHRLPEA AVASLETAVSTPFFEIPNGAGTSSANGGGRFSNLEQRVARLLRGDEEFIYHAGPLEPPSKIRGHELVQLRLDVN PDLMYATDPHDRDEVARTDEWKGAGVSRLREVWDVQHRVRLRVLWYVNSFWRSRELSYDDHEVELYRAL DAYRARVAVEYVLIRAVRDEIYAVLRRDGGALPQRFACHVPRNMSWRVVWELCRHALALWMDRADVRSC IIKALTPRLSRGAAAAAQRARRQRERPAPKPQELLFGPRNESGPPAERTWYADVVRCVRAQVDLGVEVRAA RCPRTGLWIVRDRRGRLRRWLSQPEVCVLYVTPDLDFYWVLPGGFAVFSRVTLHGLAQRALRDRFQNFEAV LARGMHVEAGRQEPETPRVSGRRLPFDDL
UL94	Forward	Tegument protein UL16	Possibly involved in virion morphogenesis	345	MAWRSGLCETDSRTLKQFLQEECMWKLVGKSRKHREYRAVACRSTIFSPEDDSSCILCQLLLLYRDGEWILCF CCNGRYQGHYGVSHVHRRRRRICHLPTLYQLSFGGPLGPASIDFLPSFSQVTSSMTCDGITPDVIYEVCMLVP QDEAKRILVKGHGAMDLTCQKAVTLGGAGAWLLPRPEGYTLFFYILCYDLFTSCGNRCDIPSMTRLMAAAT ACGQAGCSFCTDHEGHVDPTGNYVGCTPDMGRCLCYVPCGPMTQSLIHNEEPATFFCESDDAKYLCAVGS KTAAQVTLGDGLDYHIGVKDSEGRWLPVKTDVWDLVKVEEPVSRMIVCSCPVLKNLVH
UL95	Forward	Protein UL95	Promotes accumulation of late transcripts	527	MMAAAVVRAEVRRQRREERKKMAAARTTEDPPENDVVADVACGTGAVTRSSSSSLVVSSSSASGSDESSS ASPLSFPVSSPSTAVRSPGSAGVSTSLCSVERMVELSAQSPAADFSVSEAWRFEEAVNMALVACEAVSPYDR FRLIETPDENFLLVTNVIPRESAEVPVLDSSSSGGDSGPEDKKKNVGNKTAGEKNGGGSRAKRRRRRRAPKN DAATPSFLRRHDVLERFAAAAEPLPSLCVRDYALRNADRVTYDGELIYGSYLLYRKAHVELSLSSNKVQHVEA VLRQVYTPGLLDHHNVCDVEALLWLLYCGPRSFCARDTCFGREKNGCPFPALLPKLFYEPVRDYMTYMNLA ELYVFVWYRGYEFPAPTPQATTAGSGGGGGAGACAVETSASAGRVDDAGDEVHLPLKPVSLDRLREVLQA VRGRFSGREVPAWPASSRTCLLCALYSQNRLCLDLARDEARTVSYSPIVIQDCAAAVTDVTLSHILPGQSTVSL FPVYHVGKLLDALSLNDAGLITLNL
UL96	Forward	Tegument protein UL14	Involved in virion morphogenesis	127	MTSVNKQLLKDVMRVDLERQQHQFLRRTYGPQHRLTTQQALTVMRVAAREQTRYSQRTTQCVAAHLLEQ RAAVQQELQRARQLQSGNVDDALDSLTELKDTVDDVRATLVDSVSATCDLDLEVDDAV
UL97	Forward	Tegument serine/threonine protein kinase	Involved in protein phosphorylation; resistance to ganciclovir; high-level resistance to maribavir	707	MSSALRSRARSASLGTTTEGWDPPPLRRPSRARRRQWMREAAQAAAQAAVQAAQAAAAQVAQAHVDE DEVVDLMADEAGGGVTTLTTLSSVSTTTVLGHATFSACVRNDVMRDGEKEDAASDKENLRRPVVPSTSSRG SAASGDGYHGLRCRETSAMWSFEYDRDGDVTSVRRALFTGGSDPSDSVSGVRGGRKRPLRPPLVSLARTPL CRRRVGGVDAVLEENDVELRAESQDSAVASGPGRVPQPLSGSSGEESATAVEADSTSHDDVHCTCSNDQIIT

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					TSIRGLTCDPRMFLRLTHPELCELSISYLLVYVPKEDDFCHKICYAVDMSDESYRLGQGSFGEVWPLDRYRVVKVARKHSETVLTVWMSGLIRTRAAGEQQQPPSLVGTGVHRGLLTATGCCLLHNVTVHRRFHTDMFHHDQWKLACIDSYRRAFCTLADAIKFLNHQCRVCHFDITPMNVLIDVNPHNPSEIVRAALCDYSLSEPYPDYNERCVAVFQETGTARRIPNCSHRLRECYHPAFRPMPLQKLLICDPHARFPVAGLRRYCMSELSALGNVLGFCLMRLLDRRGLDEVRMGTEALLFKHAGAACRALENGKLTHCSDACLLILAAQMSYGACLLGEHGAALVSHTLRFVEAKMSSCRVRAFRRFYHECSQTMLHEYVRKNVERLLATSDGLYLYNAFRRTTSIICEEDLDGDCRQLFPE
UL98	Forward	Deoxyribonuclease	Involved in DNA processing	584	MWGVSSLDYDDDEELTRLLAVWDDEPLSLFLMNTFLLHQEGFRNLPFTVLRLSYAYRIFAKMLRAHGTPVAEDFMTRVAALARDEGLRDILGQRHAAEASRAEIAEALERVAERCDDRHGGSDDYVWLSRLLDLAPNYRQVELFQLLEKESRGQSRNSVWHLLRMDTVSATKFYEAFVSGCLPGAAAADGSGGGGSHYTGSRAGVSPGIQFGIKHEGLVKTLVECYVMHGREPVRDGLGLLIDPTSGLLGASMDLCFGVLKQGSGRTLLVEPCARVYEIKCRYKYLRKKEDPFVQNVLRRHDAAAVASLLQSHPVPGVEFRGERETPSAREFLLSHDAALFRATLKRARPLKPPEPLREYLADLLYLNKAECSEVIVFDAKHLNDDNSDGDATTTINASLGLAAGDAAGGGADHHLWGSPGDSPPPIPFEDENTPELLGRLNVYEVARFSLPAFVNPRHQYYFQMLIQQYVLSQYYIKKHPDPERIDFRDLPTVYLVSAIFREREESELGCELLAGGRVFHCDHIPLLLIVTPVVFDPQFTRHAVSTVLDRWSRDLSRKTNLPIWVPNSANEYVVSSVPRPVSP
UL99	Forward	Myristylated tegument protein	Envelope-associated; involved in virion morphogenesis	190	MGAELCKRICCEFGTTPGEPLKDALGRQVSLRSYDNIPPTSSSDEGEDDDDGEDDDNEERQQKLRLCGSSCGGNDSSSGSHREAAHDGPKKNAVRSTFREDKAPKPSKQSKKKKKPSKHHHHQQSSIMQETDDLDEEDTSIYLSPPPVPPVQVVAKRLPRPDTPRTPRQKKISQRPPTPGTKKPAASLPF
UL100	Reverse	Envelope glycoprotein M	Type 3 membrane protein; 8 transmembrane domains; complexed with envelope glycoprotein N; involved in virion morphogenesis; involved in membrane fusion	373	MAPSHVDKVNTRTWSASIVFMVLTFVNVSVHLVLSNFPHLGYPCVYYHVVDFERLNMSAYNVMHLHTPMLFLDSVQLVCYAVFMQLVFLAVTIYYLVCWIKISMRKDKGMSLNQSTRDISYMGDSLTAFLFILSMDTFQLFTLTMSFRLPSMIAFMAAVHFFCLTIFNVSMVTQYRSYKRSLFFFSRLHPKLKGTVQFRTLIVNLVEVALGFNTTVLAMALCYGFGNNFFVRTGHMVLAVFVVYAIISIIYFLLIEAVFFQYVKVQFGYHLGAFFGLCGLIYPIVQYGAFTIGDDYRTGISWSFGMLFFIWAMFTTCRAVRYFRGRGSGSVKYQALATASGEEVAALSHHDSLESRRLREEEDDDDDEDFEDA
UL102	Forward	Helicase-primase subunit	Involved in DNA replication	872	MTAQPPLHHRHHPYTLFGTSCHLSWYGLLEASVPIVQCLFLDLGGGRAEPRLHTFVVRGDRLPPAEVRAVHRASYAALASAVTTDADERRRGLEQRSAVLARVLLEGSALIRVLARTFTPVQIQTDASGVEILEAAPALGVETAALSNALSLFHVAKLVVIGSYPEVHEPRVVTHAAERVSEEYGTHAHKKLRRGYYAYDLAMSFRVGTHKYVLERDDEAVLARLFEVREVCFLRTCLRLVTPVGFVAVAVTDEQCCLLLQSAWTHLYDVLFRGFAGQPPLRDYLGPDLFETGAARSFFFPGFPPVPVYAVHGLHTLMRETALDAAAEVLSWCGLPDIVGSAGKLEVEPCALSLGVPEDEWQVFGTEAGGGAVRLNATAFRERPAGGDRRWLLPPLPRDDGDGENNVVEVSSSTGGAHPPSDDATFTVHVRDATLHRVLIVDLVERVLAKCVRARDFNPYVRYSHRLHTYAVCEKFIENLRFRSRRAFWQIQSLLGYISEHVTSACASAGLLWVLSRGHREFYVYDGYSGHGPVSAEVCVRTVVDCYWRKLFGGDDPGPTCRVQESAPGVLLVWGDERLVGPFNFFYGNGGAGGSPLHGVVGGFAAGHCGGACCAGCVVTHRHSSGGGGSGVGDADHASGGGLDAAAGSGHNGGSDRVSPSTPPAALGGCCCAAGGDWLSAVGHVLGRLPALLRERVSVSELEAVYREILFRFVARRNDVDFWLLRFQPGENEVRPHAGVIDCAPFHGVWAEQGQIIVQSRDTALAADIGYGVYVDKAFAMLTACVEVWARELLSSSTASTTACSSSVLSSALPSVTSSSSGTATVSPPSCSSSSATWLEERDEWVRSLAVDAQHAAKRVASEGLRFFRLNA

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL103	Reverse	Tegument protein UL7	Involved in virion morphogenesis	249	MEALMIRGVLEVHTDFTRQNVMIMEPQVLDFTVRGDKLWLHTEHGLLVSMAEYRSELLCTSAFLGYSAVFLLETEDAVTQVRLSDLRLKHRCGIVKADNLLHFALCTVISCVENCNLTRKCLHDLLQYLDAVNVRESFGRLLHHSARRLICSALYLLFEEKEPHIVQYVPATFVLFQQTRHTCLQLVARFFFRLTGQDEAHSFSLKLTERKTVDGWPVGLGLLDVLNANYPNLPSPPKLPPRWERGEEE
UL104	Reverse	Capsid portal protein	Dodecamer located at one capsid vertex in place of penton; involved in DNA encapsidation	697	MERNHWNEKSSGAKRSRERDLTLSTIRSILAADERLRIKASSYLGVGRGVDDEAVIDIFPTGQTMSFLRLLHGFLGTCRGQSMHQVLRDPCVLRKQLLYGVCKTLFDTITVRRVAEEWKLHAALFPYRALDEEDLEQYLLVWSASLRQSVQTGVLGALRDILYQYADNDDYGLYVDWCVTVGLVPLLDVKTKPSEAAERAQFVRAAVQRATETHPLAQDLLQANLALLLQVAERLGAVRVANAPEVRVFKKVRSERLEAQLRGKHIRLYVAAEPLAYERDKLLFTTPVAHLHEEILRYDGLCRHQKICQLLNTFPVKVVTASRHELNCKKLVEMMEQHDRGSDAKKSIMKFLLNVSDSKSRIGIEDSVESFLQDLTPSLVDQNRLLPARGPGGPGVVGPGGAVVGGPAGHVGLLPPPPGPAAPERDIRDLFKKQVIKCLEEQIQSQVDEIQDLRTLNQTWENRVRELRDLLTRYASRREDSMSLGARDAELYHLPVLEAVRKARDAAPFRPLAVEDNRLVANSFFSQFVPGTESLERFLTQLWENEYFRTFRLRRLVTHQGAEEAIVYSNYTVERVTLPYLCHILALGTLDPVPEAYLQLSFGEIVAAAYDDSKFCRYVELICSREKARRRQMSREAAGGVPERGTASSGGPGTLERSAPRRLITADEERRGPERVGRFRNGGPDDPRRAGGPYGFH
UL105	Forward	Helicase-primase helicase subunit	Involved in DNA replication	956	MSMTASSSTPRPTPKYDDALILNLSSAAKIERIVDKVKSLSRERFAPEDFSFQWFRSISRVERTTDNNPSAATTAAATTTVHSSASSSAAAAASSEAGGTRVPCVDRWPFFPFRALLVTGTAGAGKTSSIQVLAANLDCVITGTTVIAAQNLSAILNRTRSAQVKTIYRVFGFVSKHVPLADSAVSHETLERYRVCEPHEETTIQRLQINDLLAYWPVIADIVDKCLNMWERKAASASAAAAAAACEDLSELCESNIIVIDECGLMLRYMLQVVVFFYYFYNALGDTRLYRERRVPCIICVGSPTQTEALESRYDHYTQNKSVRKGVDVLSALIQNEVLINYCDIADNWVMFIHNKRCTDLDFGDLLKYMEFGIPLKEEHVAYVDRFVRPPSSIRNPSYAAEMTRLFLSHVEVQAYFKRLHEQIRLSERHRLFDLPVYCVVNNRAYQELCELADPLGDSPQPVELWFRQNLARIINYSQFVDHNLSSEITKEALRPAADVVATNNSSVQAHGGGGSVIGSTGGNDETAFFQDDDTTTAPDSRETLLTLRITYIKGSSVGVNSKVRACVIGYQGTVERFVDILQKDTFIERTPCEQAAYAYSLVSGLLFSAMYYFYVSPYTTEEMLRELARVELPDVSSLCAAAAATAAAPAWSGGENPINNHVDADSSQGGQSVPVSQRMEHGQEETHDIPCLSNHHDDSDAITDAELMDHTSLYADPFFLKYVKPPSLALLSFEETVHMYTTFRDIFLKRYQLMQRLTGGRFATLPLVTYNRRNVVFKANCQISSQTGSFVGMLSHVSPAQTYTLEGYTSDNVLSLPSDRHRIHPEVVQRGLSRLVLRDALGFLFVLDVNVSRFVESAQGKSLHVCTTVDYGLTSRTAMTIAKSQGLSLEKVAVDFGDHPKNLKMSHIYVAMSRVTDPEHLMMNVNPLRLPYEKNTAITPYICRALKDKRTTLIF
UL111A	Forward	Interleukin-10	Contains signal peptide; involved in immune regulation	176	MLSVMVSSSLVLIVFFLGASEEAKPATTTTIKNTKPQCRPEDYATRLQDLRVTFHRVKPTLQREDDYSVWLDGTVVKGCWGCSVMDWLLRRYLEIVFPAGDHVYPGLKTELHSMRSTLESIYKDMRQCPLLGCGDKSVISRLSQEAERKSDNGTRKGLSELDTLFSRLEEYLHSRK
UL112	Forward	Protein UL112	Involved in gene regulation; involved in DNA replication	684	MDLPTTVVRKYWTFANPNRILHQSVNQTFDVRQFVFDTARLVNCVDGDGKVLHLNKGWLCATIMQHGEASAGAKTQQGFMSIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRHYLKNHGESSGGGGGCGAASTASAVCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGGGGGGAGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQKERRPPPSSENDGSPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAVSSSSNNHHHHHHHHNAVTDVAAGTDGALLLPIERGAVVSSPSSTSPSSLLSLPRPGSAHSAGETVQESEAAATAAAAGLMMMRRMRRAPAEAAEAPPQSEEENDSTTPVSNCRVPPNSQESAAPQPPRSPRFDDIIQSLTKMLNDCKEKRLCDLPLVSSRLLPETSGGTVVV

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					NHSSVARTAAAVSTAGVGPPAAACPPLVTTGVVPSGSVAGVAPVAAAVETPVAPPRPVCEIKPYVVNPVVA TAAAASNSSSSSSAPLPPPPPPPSGRRGRARNNTRGGGGGGGGRNSRRQAASSSSSSSRRSRRRNNRHEDE EDNDPLLRLSQVAGSGRRRGPSFLEDGLEIIDPSEEAAIAAASIAAFFDD
UL114	Reverse	Uracil-DNA glycosylase	Involved in DNA repair	250	MALKQWMLANIADNKGSLLTPDEQARVFCLSADWIRFLSLPDHDTVLLRDTVAAVEGARQLEMVYPAPEH VHRWSYLCPPEQVRVVIVGQDPYCDGSASGLAFGTLAGRPPPPSLNNVFRELARTVDGFQRPASGCLDAW ARRGVLLLNTVFTVVHGQPGSHRHLGWQTLSNHVIRRLSERREHLVFMLWGADAHTCEYLIDRRRHLVLKS CHPSPRNTTRAFVGNDHFILANAYLDTHYRETMDWRLCG
UL115	Reverse	Envelope glycoprotein L	Contains signal peptide; complexed with envelope glycoprotein H; involved in cell entry; involved in cell-to-cell spread	278	MCRRPDCGLSFSPGPVILLWCCLLLPIVSSAAVSVAPTAAEKVPAECPELTRRCLLGEVFQGDKYESWLRPLV NVTGRDGPLSQLIRYRPVTPEAANSVLLDDAFLDTLALLYNNPDQLRALLTLLSSDTAPRWMTVMRGYSECG DGSPAVYTCVDDLCRGYDLTRLSYGRSIFTEHVLGFELVPPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPEGI TLFYGLYNAVKEFCLRHQLDPPLLRHLDKYYAGLPPELKQTRVNLPAHSRYGPQAVDAR
UL116	Reverse	Protein UL116	Contains signal peptide	313	MKRRRRWRGWLLFLALCFCLLCEAVETNTTTVTSTTAAAATTNTTVATTGTTTTSPNVTSTTSNTVTTPTTVS SVSNLTSSATSIPISTSTVSGTNTGNNNTTTIGTNATSPSPSVSILTTVTPAATSTTSNNEDVTSDYTPTFHLENI TTTRAPTRPPAQDLCSHNLSIILYEEESQSSVDITVDEEESELEEDDDEEYDELWFPLYFEAECNRNYTLHVNHS CDYSVRQSSVSFPPWRDIDSVTFVPRNLSNCSAHGLAVIVAGNQTWYVNPFSLAHLLDAIYNVLGIEDLSAN FRRQLAPYRHTLIVPQT
UL117	Reverse	Protein UL117		424	MVMFSQDHVQIVYGSTRICKSLAPANKRKTHRTIVVAPRRGFLRIPPDGQDVNHVKIVPTTSSSLAPPRDDE RRPTPPLRPPLTVYPYGTSLIRRSARDAKLRSKLIVFHITRPALGQHPQNPGISGPAAMDHSEFLTSFRREVDR QTVLTAESAPATVEVCLGDALPGGVMGGGGLPAGVGSASAAVAAAAAAVAGVPVAANPVMPATATVTTP PMIDLTSHHRPLTLFTPASAAAAPAVATNGGNATYILPADCRYAPLFASKYKYVFEEVSRLMRLHDSTAVQLQ ISASCGNAFQALKSALLKLHNVTVLAGQQLITQTMPHTPQAVATFKFFHQDPNRVLDCIRPVVPRSTSYHET GVYQMWVSGATKKDLFDAVTLCASIVEKQPDVFNINVSLLTYPSIAAPHLPLYNEFTSFRLPTS
UL119	Reverse	Membrane glycoprotein UL119	Type 1 membrane protein; IgG Fc-binding; similar to OX-2; involved in immune regulation; possibly related to UL120 family	351	MCPALAIALAAALLSNTHPGMGSSTTSAVTSPNITVTFTTSISTPNNVTSTVTTTVQTSTFTSGSTSTSASTSVA TTTQKEGHLYNVNCEASYSYDQVSLNATCKVNLLNNTKNPDILSVTCYARTNCKGPFIQVGYLSAFPSDDKGK LHLSYNATAQELLISGLRPQETTEYTCSFFSWGRHHNATWDLFTYPIYAVYGTRLNATTMRVRVLLQEHEHCL LNGSSLYHPNSTVHLHQGNQLIPPWNISNVTYNGQRLREFVFYLNGTYTVVRLHVQIAGRSFTTTYVFIKSDP LFEDRLLAYGVLAFLVFMVIILLYVTYMLARRRDWSYKRLEEPVEEKKHPVPYFKQW
UL120	Reverse	Membrane protein UL120	Type 1 membrane protein; UL120 family	197	MYRAGVTLLVVAVISVGRWDVVAVAIGWYEPQVSVAYMYEHNDTNLTIFCNTTACDSPFLASGMMISIPQ NTQFLTSKVNYSDDMYNDRKNYTHELKRMLTGAPGAYVNGSVTCWGSNGTFGAKTFLVKSMINTTAGNT STSIIHFVQQDELVENPAYFRRSNHRAFMIVILTQVVFVVFIINASFIWSWTFKRHKR
UL121	Reverse	Membrane protein UL121	Type 1 membrane protein; UL120 family	180	MWGCGWSRILVLLPLMCMALMARGTHGAYICSPNPGRLRISCALSVLDQRLWWEIQYSSGRLTRVLVFHD EGDEGDDLHLTDTRHCTSCTHPYVISLVTPLTINATLRLLIRDGMYGRGEKELCIAHLPTLRDIRTCRVDADLGL LYAVCLILSFSIVAAALWKVDYDRSVAVVSKSYKS
UL122	Reverse	Regulatory protein IE2	Interacts with basal transcriptional machinery and cellular transcription factors; specific DNA-binding protein; involved in gene regulation	580	MESSAKRKMDPDNPDEGPSSKVPRPETPVTKATTFLQTMLRKEVNSQLSLGDPLFPELAEESLKTFEQVTED CNENPEKDVLAELGDILAQAVNHAGIDSSSTGPTLTTHSCSVSSAPLNKPTPTSVAVTNTPLPGASATPELSPR KKPRKTTRPFKVIIKPPVPPAPIMLPLIKQEDIKPEPDFTIQYRNKIIDTAGCIVISDSEEEQGEEVETRGATASSP STGSGTPRVTSPTHPLSQMNHPPLPDPLARPDEDSSSSSSSSCSSASDSESESEEMKCSSGGGASVTSSHHGR GGFGGAASSSLLSCGHQSSGGASTGPRKKKSKRISELDNEKVRNIMKDKNTPFCTPNVQTRRGRVKIDEVSR

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					MFRNTNRSLEYKNLPFMIPSMHQVLEEAIKACKTMQVNNKGIQIIYTRNHEVKNEVDQVRCRLGTMCNLALSTPFLMEHTMPVTHPPEVAQRTADACNEGVKAAWSLKELHTHQLCPRSSDYRNMIIHAATPVDLLGALNLCLPLMQKFPKQVMVRIFSTNQGGFMLPIYETAAKAYAVGQFEQPTETPPEDLDTLSLAIEAAIQDLRNKSQ
UL123	Reverse	Regulatory protein IE1	Enhances activation by IE2; interacts with basal transcriptional machinery and cellular transcription factor; disrupts ND10; involved in gene regulation	491	MESSAKRKMDPDNPDEGPSSKVPRPETPVTKATTFLQTMLRKEVNSQLSLGDPLFPELAEESLKTFEQVTEDCNENPEKDVLAELVKQIKVRVDMVRHRIKEHMLKKYAQTEEKFTGAFNMMGGCLQNALDILDKVHEPFEEMKCIGLTMQSMYENYIVPEDKREMWMACIKELHDVSKGAANKLGGALKAKARAKKDELRRKMMYMCYRNIEFFTKNSAFPKTTNGCSQAMAALQNLPQCSPDEIMAYAQKIFKILDEERDKVLTHIDHIFMDILTTCVETMCNEYKVTSDACMMTMYGGISLLSEFCRVLCCYVLEETSVMLAKRPLITKPEVISVMKRRIEEICMKVFAQYILGADPLRVCSPSVDDLRAIAEESDEEDAIVAYTLATAGASSSDSLVSPPESPVPATIPLSSVIVAENSDQEESEQSDEEEEEGAQEEREDTVSVKSEPVSEIEEVAPEEEEDGAEEPTASGGKSTHPMVTRSKADQ
UL124	Forward	Membrane protein UL124	Type 1 membrane protein	152	MERNSLLVCQLLCLVARAAATSTAQTTLPSTVNSTATGVTSDSYQNTTTQLPASSSAAALSLPSASAVRAQSPSSFSDTYPTATALCGTLVVVGIVLCLSLASTVRSKELPSDHEPLEAWDQGSDVEAPPLPEKSPCPEHVPEIRVEIPRYV
UL128	Reverse	Envelope protein UL128	Contains signal peptide; similar to CC chemokines; complexed with envelope glycoproteins H and L; essential in non-fibroblast cells; involved in cell entry	171	MSPKDLTPFLTALWLLLGHSRVPRVRAEECCEFINVNHPPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIVTTMTHSLTRQVVHNKLTSCNYNPLYLEADGRIRCGKVNDKAQYLLGAAGSVPYRWINLEYDKITRIVGLDQYLESVKKHKRLDVCRAKMGYMLQ
UL130	Reverse	Envelope glycoprotein UL130	Contains signal peptide; complexed with envelope glycoproteins H and L; essential in non-fibroblast cells; involved in cell entry	214	MLRLLLRHHFHCLLLCAVWATPCLASPWSTLTANQNPSPPWSKLTYSKPHDAATFYCPFLYPSPPRSPLQFSGFQQVSTGPECRNETLYLLYNREGQTLVERSSTWVKKVIWYLSGRNQTILQRMPQTASKPSDGNVQISVEDAKIFGAHMVPKQTKLLRFVVNDGTRYQMCVMKLESWAHVFRDYSVSFQVRLTFTEANNQTYTFCTHPNLIV
UL131A	Reverse	Envelope protein UL131A	Contains signal peptide; complexed with envelope glycoproteins H and L; essential in non-fibroblast cells; involved in cell entry	129	MRLCRVWLSVCLCAVVLGQCQRETAEKNDYYRVPHYWDACSRALPDQTRYKYVEQLVDLTLNYHYDASHGLDNFDVLKRINVTEVSLLISDFRRQNRRGGTNKRTTFNAAGSLAPHARSLEFSVRLFAN
UL132	Reverse	Envelope glycoprotein UL132	Type 1 membrane protein	270	MPAPRGLLRVTFLVLIAFGLLLHMDFSDATNMTSSTNVPTSTSSRNTVESATSSEPTTETNMTTAHESSVHNAHNDEIMKVLAILFYIVTGTSIFSFIAVLIAVVYSSCCKHPGRFRFVDEEAVNLLDDTDDSGGSSPFGSGSRRGSQIPAGFCSSSPYQRLETRDWDEEEEASAARERMKHDPENVIYFRKDGNLDTSFVNPNYGRGSPLTIESHLSDNEEDPIRYYVSVYDELTASEMEEPSNSTSWQIPKLMKVAMQPVSLRDPEYD
UL148	Reverse	Membrane protein UL148	Type 1 membrane protein	316	MLRLLFTLVLLVLHGPSVNASRDYVHVRLLSYRGDPLVFKHTFSGVRRPFTELGWAACRDWDSMHCTPFWSTDLEQMTDSVRRYSTVSPGKEVTLQLHGNQTVQPSFLSFTCRLQLEPVVENVGLYVAYVVNDGERPQQFFTPQVDVVRFALYLETLSRIVEPLESGRLAVEFDTPDLALAPDLVSSLFVAGHGETDFYMNWTLRRSQTHYLEEMALQVEILKPRGVRHRAIIHHPKLQPGVGLWIDFCVYRYNARLTRGYVRYTLSPKARLPAKAEGWLVSLDRFIVQYLNTLLITMMAAIWARVLITYLVSRRR
UL147A	Reverse	Membrane protein UL147A	Type 1 membrane protein	75	MSLFYRAVALGTLSALVWYSTSILAEINENSCSSSSVDHEDCEEPDEIVREEQDYRALLAFSLVICGTLLVTCVI
UL147	Reverse	Chemokine vCXCL2	Contains signal peptide; putative secreted glycoprotein; possibly involved in immune regulation; CXCL family	159	MRPTWSYHPIPNSHVKLLSVRYLYLTAYVLLSTCPLFVYLLELEDYDKRCRCNNQILLNTLPVGTQLLKPIAASESCNRQEVLAILKDKGTKCLNPNAQAVRRHINRLFFRLVLDEEQRIYDVVSTNIEFGAWPVPTAYKAFLWKYAKKLNYHYFRLRW

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL146	Reverse	Chemokine vCXCL1	Contains signal peptide; putative secreted glycoprotein; involved in immune regulation; CXCL family	125	MRFIFGLLIIFLAYMYYYEVNGTELRCKCPGDKKLPRTRIMLGDFWTHRESGGPGCNGYQYLLYFNNGGKHG RGVCLAPDHHISKWLDTHNDGRWYNVNITKQPGRRTGGRGPGQVNITLIAVKQ
UL145	Reverse	Protein UL145	RL1 family	130	MYGVLAHYYSFISSPSVMVNFKHHNAVQLLCARTRDGTAGWERLTHHASYHANYGAYAVLMATSQRKSLV LHRYSAVTAVALQLMPVEMLRKLDQSDWVRGAWIVSETFPTSDPKGFWSDDDSSMGGSED
UL144	Reverse	Membrane glycoprotein UL144	Type 1 membrane protein; similar to TNFR; interacts with BTLA and inhibits T cell proliferation; does not bind TNF superfamily ligands; activates NFkappaB via TRAF6-dependent mechanism to upregulate CCL22; involved in immune regulation	176	MKPLVMLICFGVFLLQLGGSKMCKPDEVKLGNQCCPPCGSGQKVTKVCTENSGITCTLCPNGTYLTGLYNCT NCTQCDDTQITVRNCTSTNNTICASKNYTLFSTPGVQHHKQRQQNHTAHVTVKQGKSGRHTLAWLSLFIFL VGIILLILYLIAAYRSERCQQCCSIGKIFYRTL
UL142	Reverse	Membrane glycoprotein UL142	Type 1 membrane protein; similar to MHC-I; inhibits NK cell cytotoxicity; downregulates MICA; acts on activating receptor NKG2D; involved in immune regulation; MHC family	306	MRIEWAYWLFGYFVSSVGSERSLSYRYHLESNPSTDVVCNGNISVFVNGTLGVRYNITVGISSSLLIGHLTIQV LESWFTPWVQNKSYIKQPLGDTETLYNIDSENIHRVSQYFHTRWIKFLQENHTCDLTNSTPTYTYQVNVNNT NYLTLTSSGWQDRLNYTVINSTHFNLTESNITNIQKYLNTTCIERLRNYTLESVYTTTVPQNVTTSQHATTTLHT IPPNTITIQDTTQSHTVQTPSFNDTHNVTEHTLNISYVLSQKTNNTTSPWVYAIPMGATATIGAGLYIGKHFTP VRFVYEVWRGQ
UL141	Reverse	Membrane glycoprotein UL141	Type 1 membrane protein; inhibits NK cell cytotoxicity; sequesters CD155; acts on activating receptor DNAM-1, CD96; involved in immune regulation; UL14 family	338	MCRRESLRTLPWLFWALLSCPRLLEYSSSSFPFATADIAEKMWAENYETTSPAPVLVAEGEQVTIPCTVMTH SWPMVSIRARFCRSHDGSDELILDAVKGHRLMNGLQYRLPYATWNFSQLHLGQIFSLTFNVSMDTAGMYE CVLRNYSHGLIMQRFVILTQLETLSRPDEPCCTPALGRYSLGDQIWSPTPWRLRNHDCGMYRGFQRNYFYIG HADAEDCWKPACPDEEPDRCWTVIQRYRLPGDCYRSQPHPPKFLPVTPAPPADIDTGMSPWATRGIAAFL GFWSIFTVCFLCYLCYLQCCGRWCPTPGRGRRGGEGYRRLPTYDSYPGVRKMKR
UL140	Reverse	Protein UL140	Contains potential transmembrane domain	191	MTPAQTNATTTVHPHDAKNGSGGSALPTLVVFGFIVTLLFFLFMLYFWNNDVFRKLLRCAWIQHCCDRFDA WQDEVIYRRPSRRSQSDDESRTNSVSSYVLLSPASDGGFDNPALTEAVDSVDDWATTSVFYATSDETADAE RRDSQQLLIELPPEPLPPDVVAAMQKAVKRAVQNALRHSHDSWQLHQTL
UL139	Reverse	Membrane glycoprotein UL139	Type 1 membrane protein	135	MLWILVLFALAASASETTTGTSSNSSQSTSATANTTVSTCINASNGSSWTVPQLALLAASGWTLSGLLLLFTCC FCCFWLVRKICSCCGNSSESESKTTHAYTNAAFTSSDATLPMGTTGSYTPPQDGSFPPPPR
UL138	Reverse	Protein UL138	Contains potential transmembrane domain; possibly involved in latency	169	MDDLPLNVGLPIIGVMLVLIVAILCYLAYHWHDTFKLVRMFLSYRWLIRCCELYGEYERRFADLSSLGLGAVRR ESDRRYRFSERPDEILVRWEEVSSQCSYASSRITDRRAGSSSSSSVHVASQRNSVPPPDMAVTAPLTDVDLLK PVTGSATQFTTVAMVHYHQEYT
UL136	Reverse	Protein UL136	Contains potential transmembrane domain	240	MSVKGVEMPEMTWDLDVGNKWRRRKALSRIHRFWECRLRVWWLSDAGVRETDPPRPRRRPTWMTAV FHVICAVLLTLMIMAIGALIAYLRYYHQDSWRDMLHDLFCGCHYPEKCRRHHERQRRRRRAMDVPDPELG DPARRPLNGAMYYGSGCRFDTVEMVDETRPAPPALSSPETGDDSNDDAVAGGGAGGVTSPATRTTSPNAL LPEWMDAVHVAVQAAVQATVQVSGPRENAVSPAT
UL135	Reverse	Protein UL135	Contains potential transmembrane domain	308	MVWLWLGVGLLGGTGLASLVLAISLFTQRRGRKRSDETSSRGRLPAAASDKRGACACCYRNPKEDVVEPLD LELGLMRVATHPPTPQVPRCTSLYIGEDGLPIDKPEFPPARFEIPDVSTPGTPTSIGRSPSHCSSSSSLSSSTSVD TVLHQPPPSWKPPPPPGRKKRPPTPPVRAPTTRLSSHRPPTPIPAPRKNLSTPPTKKTPPPTKPKPVGWTPPV TPRPFPKTPTPQKPPRNPRLPRTVGLENLSKVGLSCPCPRPRTPTEPTTLPIVSVSELAPPPRWSDIEELLEQAV QSVMKDAESMQMT

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
UL133	Reverse	Protein UL133	Contains potential transmembrane domain	257	MGCDVHDPSWQCQWGVPTIIVAWITCAALGIWCLVGSPNTFSGPGIAAVVGCSVFMIFLCAYLIRYREFFK DSVIDVFTCRWVRYCSCSCKCSCKCISGPCSRCCSACYKETMIYDMVQYGHRRRPGHGDDPDRVICEIVEGP PVSAPTVSVPPPSEESHQPVIPPQPPAPTSEPKPKKGRAKDKPKGRPKDKPPCEPTVSSQLPLQPTAMPGGP PDAPPPAMPQMPPGVAEAVQAAVQAAVTAALQQQQQQHQTGT
UL148A	Reverse	Protein UL148A	Contains potential transmembrane domain	79	MSSQSLDPWIPVCVVVVMTSVVLFAGLHVYLWYVRRQLVAFCLEKVCVRCCGKDETTPLVEDAEPPAELEM VEVSDECY
UL148B	Reverse	Protein UL148B	Contains potential transmembrane domain	80	MALDTLAGLAICVGLVMGVTVIASCALLVFYYCDEREDGRPSKLLQRSIRRWRHGLSTESLTAILPDGSSTEQE IYHTRL
UL148C	Forward	Protein UL148C	Contains potential transmembrane domain	77	MLTPAVFPAVLYLLALVVWVEMFCLVAVAVVEREIAWALLLRMLVVGLMVEVGAAAAWTFVRCLAYQRSF PVLTAFP
UL148D	Forward	Protein UL148D	Contains potential transmembrane domain	62	MTAPKCVTTTTYLVKTKERPWWPDNAIRRWWISVAIVIFIGVCLVALMYFTQQQARSGSSSG
UL150	Reverse	Protein UL150	Contains signal peptide	639	MLTLYLFTATCCFVYALPRQTPADTAICHHTWPMCSDVPHPTSSAHNTALSSPHAISIEHRLLHSPSPPRENIR HSMRCRRRDMASSASTPVSHTQPLAANHRRSRITYATTDPTNSPTASPAKSDKLEADADPALHRRPASLLRH LFQPCHAQRGTSNRATSHRASLNAVHHKLCGAMISSSCSTTCTPLIMDLPSLSVELSAGHKKKETPTEGGWG GEEGEDDVLATIRNTLSAPTSPAAATTHRLSFPGESTFCLTAVSECSQRRTSTAALTPPPPTVAAAFSFSSTVSE TATFPQSTTGRTRVDDTAVVTAGDPRSPVTHVTLLQIFRLRSSLLTSRSGGALRGGEHEAIPKVASLFWTLLKA TQIVEMTHKTPSGNSHRNPQKYTDRPQRLLLTALAIWQRTYNDTRGAHAPQVRLLGDILTYRRPQTATASTK AHTQQQPEEPKGQQIWTQTAGQAAPHGDEPHSDGELRHESHSAPPTSRTLPDTILAVKRRSVAQRSHVRL DAKPGLNERDGFRRRLLLPLSGYFRANELRNQQFMGYGTKNGLKNTWLTRPLGVAGGVRETIGERQDGNV ADSATQRVFHTLYAALQTVRVWYTALGTAWRTSGSRTRESLFDGPRRRDRQAARLRRLEL
UL150A	Forward	Protein UL150A		271	MASCSPPRRAPPDLLVSSELRRRNIWRRVTCVTGERGSPAVTTAVSSTRVRPVVLCGKVAVSETVDEKENAA ATVGGGGVNAAVDVRRCEHSETAVRQKVDSPGNDSRCVVAAADSSTDSDGKSIISGVQVVEHDEDIIAPQS LWCTAFKEALWDVALLEVPRWAWQGWKRWRNSEAGRRWSAGSASASSLSDLAGEAVGELVGSVVAYVI LERLWLAARGWVCETGVEAEEAMSRRRQRMLWRMFSRGGDGECSRRCSMEMACGEESAVL
IRS1	Forward	Tegument protein IRS1	Transcriptional activator; blocks phosphorylation of eIF2alpha and host shutoff of protein synthesis; binds dsRNA; involved in gene regulation; involved in translational regulation; US22 family	858	MAQRNGMSPRPPPLGRGRGAGGPSGVGSSLSSSCVSLGATSTAGTSTSTAGTATPGHGVHRIEPRGPPGA PPGSGNNSNFWHGPERLLLSQIPVERQALTELEYQAMGAVWRAAFLANSTGRAMRKWSQRDAGTLLPLG RPYGFYARVTPRSQMNGVGATDLRQLSPRDAWIVLVATVVHEVDPAADPTVGDKAGHPEGLCAQDGLYL ALGAGFRVFVYDLANNTLILAARDADEWFRHGAGEVVRLYRCNRLGVGTPRATLLPQPALRQTLLRAEEATA LGRELRRRWAGTTVALQTPGRRLQPMVLLGAWQELAQYEPFASAPHPASLLTAVRRHLNQRLCCGWLALG AVLPARWLGCAAGPATGTAAGTTTMPAGTTSPPGASGTETEAAGGDAPCAIAGAVGSAVTLPPQPYGAA GGSAICAPNADAHAVLGADATAAAAAAAAAPTVMVGPTAMAGPAASGTVPRAMLVVVLDELGAVFGYC PLDGHVYPLAAELSHFLRAGVLGALALGRESAPAAEAARRLLPELDREQWERPRWDALHLHPRAALWAREP HGQWEFMFREQRGDPISDPVAFRISDARTLGLDLTTVMTERQSQLPEKYIGFYQIRKPPWLMEQPPPPSRQ TKPDAATMPPPLSAQASVSHALRYDDESWRPLSTVHDHKAWLDLDESHWVLGDSRPDDIKQRRLLKATQR RGAEIDRPMPVVPEECYDQRFTTEGHQVIPLCASEPEDDDEDPTYDELPSRPPQKHKPPDKPPRLCKTGPGP PPLPPKQRHGSTDGKVSAPRQSEHHKRQTRPPRPPPPKFGDRTAAHLSQNMRDMYLDMCTSSGHRPRPP APPRPKKCQTHAPHHVHH

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
US1	Reverse	Protein US1	Contains duplicated TT virus ORF2 motif; US1 family	156	MASGLGDLSVGVSSLPMRELAWRRVADDSHDLWCCCMDWKAHVEYAHPASELRPGSGGWPEHAEAQWRQQVHAAHDVWCNCGDWQGHALRSRSRTAESGRSSLSSSVSVLSDGDQQPWWRRLRVKRPKFPSWARRWTQRHDSEERASQQAENDSTS
US2	Reverse	Membrane glycoprotein US2	Contains potential transmembrane domain; causes selective degradation of MHC-I and MHC-II; involved in immune regulation; US2 family	199	MNNLWKAWVGLWTSMGPLIRLPDGITKAGEDALRPWKSTAKHPWFEIEDNRCYIDNGKLFARGSIVGNMSRFVFDPKADYGGVGENLYVHADDVEFVPGESLKWNVRNLDVMPIFETLALRLVLQGDVIWLRCVPELRVDYTSSAYMWNMQYGMVRKSYTHVAWTIVFYSINITLLVLFIVYVTVDCNLSMMWMRFFVC
US3	Reverse	Membrane glycoprotein US3	Type 1 membrane protein; inhibits processing and transport of MHC-I and MHC-II; involved in immune regulation; US2 family	186	MKPVLVLAILAVLFLRLADSVPRPLDVVVSEIRSAHFRVEENQCWFHMGMLHYKGRMSGNFTEKHFVSVGIVSQSYMDRLQVSGEQYHHDERGAYFEWNIGGHPVPHTVDMVDITLSTRWGDPKKYAACVPQVRMDYSSQTINWYLQRSIRDDNWGLLFRTLLVYLFSLVVLVLLTVGVSARLRFI
US6	Reverse	Membrane glycoprotein US6	Type 1 membrane protein; inhibits TAP-mediated peptide transport; involved in immune regulation; US6 family	183	MDLLIRLGFLLMCALPTPGERSSRDPKTLLSLSPRQQACVPRTKSHRPICYNDTGDCTDADDSWKQLGEDFAHQCLLAAKKRPKTHKSRPNDRNLEGKLTCQRVSRLLPCDLDIHPSHRLLTLMNNCVCDGAVWNAFRLIERHGFFAVTLYLCCGITLLVVILALLCSITYESTGRGIRRCGS
US7	Reverse	Membrane glycoprotein US7	Type 1 membrane protein; possibly involved in immune regulation; US6 family	225	MRIQLLLVSTLVAFIVATRVEDMATFRTEKQWQQDLQYRREFVKRQLAPKPKSNIVVSHTVSCVIDGGNMTSVWRFEGQFNAHIASEVILHDTSGLYDVPHEVQNDGQVLTVTVKRSVPANIAKVLISLKPVKLSSGQYECRPQLQLPWIPKPSSFMYDSYRLWYEKRWLTIILYVFMWTYLVMLLQYCIVRFIGTRLFYFLQRNITIRFTGKPTYNLLTYPVKG
US8	Reverse	Membrane glycoprotein US8	Type 1 membrane protein; binds to MHC-I; involved in immune regulation; US6 family	226	MRRWLRLLVGLGCCWVTLAHAGNPYEDDYYYYREDEPRQHGEPNYVAPPARQFRFPPLNNVSSYQASCVVKDGVLDAVWRVQGTFYPEKEIVARVGWSGRHGRKWGRLQAPECLVETTEAVFRLRQWVPTDLDHLTLHLVPCSKCKPMWCQPRYHIRYFSYGNSVDNLRRLHYEYRHLELGVVIAIQLAMVLLLGYVLARTVYRVSSAHYLRWHACVPQKCEKSLC
US9	Reverse	Membrane glycoprotein US9	Type 1 membrane protein; possibly involved in immune regulation; US6 family	232	MILWSPSTCSFFWHWCLIAVSVLSSRSKESLRLSWSSDESSASSSSRICPLSDSKSVRLPQYPRGFGDVSGYRVSSSVSECYVQHGVLVAAWLVRGNFSDTAPRAYGTWGNERSATHFKVGAPQLENDGALRYETELPQVDARLSYVMLTVYPCSACNRSMLHCRPASRLPWLPLRATPSDLERLFAERRYLTFLYVVLVQFVKHVALFSFGVQVACCVYLRWIRPWVQGRG
US10	Reverse	Membrane glycoprotein US10	Type 1 membrane protein; delays trafficking of MHC-I; involved in immune regulation; US6 family	184	MLRRGSLRNPLATCLLWLGVVAAATEETREPTYFTCGCVIQNHVLKGAVKLYGQFPSPKTLRVSTWLHDGENHERHRQPILVEGTATATEALYILLPTELSSPEGNRPRNYSVTLTLASRDCYERFVCPVYDSGTPMGVLMNLTYLWYLGDYGAILKIYFGLFCGACVITRSLLLICGYYPPRE
US11	Reverse	Membrane glycoprotein US11	Type 1 membrane protein; causes selective degradation of MHC-I; involved in immune regulation; US6 family	215	MNLVMLILALWAPVAGSMPELSLTLFDEPPPLVETEPLPPLSDVSEYRVEYSEARCVLRSGGRLEALWTLRGNLSVPTPTPRVYYQTLEGYADRVPTPVEDVSESLVAKRYWLRDYRVPQRTKLVLFYFSPCHQCQTYYVECEPRCLVPWVPLWSSLEDIERLLFEDRRLMAYYALTIKSAQYTLMMVAVIQVFWGLYVKGWLHRHFPWMFSDQW
US12	Reverse	Membrane protein US12	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	281	MVQIQFHQGEPLGHKKEKPPPVSPPSPPPIRRVTVITKDEDTLRSVQHFLWMVRLYGTVVFQTSATIATTILFMLIPWRVTTPYLRDTLPFWSTLLPCALRCHAYWLERRRRPGTLMLVMVYTTLTTISVSTIGLCFDRTVVIQAYVLSSMLCVWCTGLAWLMAWNMQRRLAILCLLSFMLPILWLFIAVQSWEPYQRIILALTVSFIYGLKIVLIRDTLTVLYRSPSNCYTDGDLLRTAMLLYMDQVIMFLLVVVPLTAPIWYPNYAGALGRTAHWLFHK
US13	Reverse	Membrane protein US13	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	261	MDPPLPSLHSPQWASLLQLHHGLMWLRRFAVLVRVYALVVFHIAISTAFCGMIWLGIPDSHNICQHESSPLLLVFAPSLLWCLVLIQGERHPDDVVLTMGYVGLLSVTTVFYTWCSDLPAILIDYTLVLTLWIACTGAVMVGDSFRAKRWELICSRVLTSVFFITLWVIGDQTVFHHQRILLYGYGAIVFLMMTVTFYGTRYIRDELPAAQTLRGSLLIYVGLVTMFKITLIVLSPNLWRLPWTTVFAAFRSSYCEGGGGS

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
US14	Reverse	Membrane protein US14	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	310	METVSTQRETASSETERTREAASAETTDATFRSLEEGSTISSRYSETASTASEDAVCWLRRTAIVLRVYGLLTLETAFSVLISALVWLGYPSLGYECRDDPSPLLLSCTPVLVLGALELTDHRHPSNGLVFALYVALLSFTTAGLNLCATAPIGISSLILTWTLFVACNGVAWEHRLSSVWRDALFTSTLLTVMVSVLASTYTWLHKTLLCLYTVFVGCILAVLFQDVRYIATKMPVSHVIRSSLVLYATETLIYHTTLLMLTPVVWSARWDQMFSYLAKLGTYHHYRVDNGTLSVILNSTTATFQSRVA
US15	Reverse	Membrane protein US15	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	262	MRREKGFQVPTDGTVIYVPPGIQETRLATRSLAWVDCCRVALHTYGAVGWQLAGLTALLSAFCYAAPATWFHHSRLCLTESSPSLVFVIPVTSVIFIHCYETSHPSNIGVLLFYTLLHVPPLIVICLCLDGTLVISAALFTLLAFLSCTGVALLAPERTVRRQIVVVHALITLTFTAIVVVILRRGWSWCFKIVLSFSVLITCLAVSHFHEAALAVRYETPLERALLAAVKVFLSLVFTLLMVLRIMTLRTFLQTYFSSDKL
US16	Reverse	Membrane protein US16	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	309	MGLRFPTATQRQIVFRRLFDSGNNDDYDEAAVVAVLGWVHRFEVVVRIAGLLLFQISAAVAVLGSFSLVFPTATLKSRPGFPCHVVWAPEVLLLVPVASALFVYFRYERPVLAQRNRHPRCRRPFRQLVLLLAGLLAHIPALGVTCACQEPREVLTSFVLTLVITLLCAEVVFICRDNCTLSDQFTLINGVWVVVFLANVLIVFTRPWTWPLRLLLGFYSTVGLIFAGHFSQQVLFVRHVLMPRDVAHTSLQLFITFISLFFLILRIRNCQDLLSDLRLLELPSSDAMTLTPNDLSHASSSTPLSTLSP
US17	Reverse	Membrane protein US17	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	293	MSPNSEATGTAWAPPPPRPSRGVIMISSVSTNDVRRFLLCMRVYSTVAVQGTCTFLLCLGLVLAFPHLKSTVFLCCTGFMPPLSLVVPTICLALLHGKRDEGSFTSPPSPGLLTIYSVLTTLSVIVASACSSSTLVTFSGLLACVLFSLCSCVTGLAAHNHRRWQVIITLFVIGVIAFLIALYLQPVPLGHKLFLGYYAMALSFMLVVTVFDTTRLFEIAWSEADLLTLCLYENLVYLYLLILILFTTEDSLDKLIAWMTWLSSRATGATNAASISGCDLLREVQRNLTRTMA
US18	Reverse	Membrane protein US18	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	274	MGDTASVSEHHESPTVTIVPLHRSHALVAEQQLFQWLKRFKLLMEVYHGLVWQLACTLTVCLLAWLAFPDVQGQCANGIVPALSSIVPVSTLAMLRGFAEFRPHTTNFAHLTVACLLINTGITVCTGFCGERRVIGLSFALVMVFFVLCSGLTYLAGNNPTRWKVIGIGYGWSVIVFYLLLYFSPVLWVSKIYSGLYVLVVTAASAVLIYETLDLIYQRGTLSKNSVCVSVVLYTIVMSLLNMSVAIFSGHVWVQQYAEKHGGRIDGVSLLSLL
US19	Reverse	Membrane protein US19	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	240	MLHVVPLEWTVEEVVPYVERLAVWLRASVLVAFQLTATVALSVLSWWLMPPPVAELCERGRDDDPPPLSHLSLVVPVGCLFLLLRGPSLDRCPRKLPLLLAYCLPHALAFLTLLMCQPSPQAFLGAALLALAVDLGCLGASLLGCDPGASLRRLWLPSVLSLLCATALGLWLLRAAAPFFLGLHATTLLTVTLMLIHDLSLITCQSSFPESFQPSLRLYVENVALFIGMYHLLRLWLWSP
US20	Reverse	Membrane protein US20	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	254	MQAQEANALLLSRMEALEWFKKFTVWLRVYAIFIFQLAFSFGLGSVFWLGFPQNRNFCVENYSFFLTVLVPIVCMFITYTLGNEHPSNATVLFIYLLANSLTAAIFQMCSESRVLVGSYVMTLALFISFTGLAFLGGRDRRRWKCISCVYVVMLLSFLTLALLSDADWLQKIVVTLCAFSISFFLGILAYDSLMVIFFCPPNQCIRHAVCLYLDSMAIFLTLLLMLSGPRWISLSDGASLDNGTLTAASTTGTS
US21	Reverse	Membrane protein US21	Type 3 membrane protein; 7 transmembrane domains; possibly involved in virion morphogenesis; US12 family	243	MSLRGQVQIARSVFLLRIYILIWVQCLILMSVCAFCWLVLPHRLEQLFSSVRLTLSCLMISIVCLGLLRWAEPNFPKNVWILLTYTLLTSVAVTASGFHFSHRSVIYAMVATVTLFCFLTLATYLFARDVELQRSLLTGASTLILLLFAVFSLFPEAVSEILVMIAGLAVIVTSVVCDTQDILHDIEYESYIPGALCLYMDLMYLFVSVLYFMPSEPDSAHTAQTTVATTVAATAAASPQFVS
US22	Reverse	Tegument protein US22	US22 family	576	MSLLTKAAAEAWGTYLRQRDERCEDAIRCDYGVFQFRNTVFQKTLSTLQGLYLRQYDPPALRTYVQRHQGTTVALRNPANWFLVMREQAAIPQIYARSLAADYLCCDDTLEAVGVLAVRPPDSDLTRNTKQAQELPCVLMLSHYGTVYVYDWETDGLYEVASDIKAFSKNGLLWCEYVYRHPQTPFATTEPRYHVQKFLCTDPTDAAAVAKTAREMSGLNLVIRTPGRTEVEPLLMLGSIEGLRACRPFDHMPAADFRDLLNFIRQRLCCEWYVVGLVGYYLAYG

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					PFVPSGLVLLDKFGVVYLHKIEDSDLYRIADNFHMFLKCGLLKLRGLCRFDRGLRGECRLEELPVCHHTLKRDV LRWHGALGTITRSQLESALDWFLRPTRGTDKVPNNSSAWGRTDLLPTGALQDNQNWAFPSHSVETLPALQ GGLWEDNDETTQTVDGQRCFRMPKFFPPPMCRAPYNECGVELPGGDSSDEDESGRPRRIANRIGDTPETP CNSENEDDTTVEGTSGGPEAMDSQTPYPSEDSPTTEKEAWLQRGRRAKAMHARGHTLKQVPIPEPDQMG DDGPDPGP
US23	Reverse	Protein US23	US22 family	592	MWRTRWEDGAPTFTRNDEFLYCHTRYETFLRVMGDFQGIFECQYSADVLRDWVRNHVDQVLSLGIPHN WFLQVRPGSTMPELRDQLLDDVICCPERLIVLGKCVIMVEDHYEETELVLCMGGGTRLYIYEPSQEILLLCARH LDELARYGMMYTEAVYRQPQTPFATRVPHDVVAMLLRHGHDADALAACVGEHHGRDVNFHTPGRHAKT LKLLTSFGCLTDCWPFEVAPAARLAECEMYVTLQLRCRWYLLGAVGSYRAGGFFDTSFLLIFDRFCRFYVVIVK SHLDRSPPLQRLAGEIYRLADSLEELFRAGLMKVYVRRRYEHGLRRAARLERNGGCVHMGEAARLHFTMFD SGVDRDYARQFRWLCRGDRFRAEMLNNWDGWDAFTIWQARVVRGDFAERRRPRSLGDGEEEDEGNDG RAMPVVRRRPPPMPRDDDEDNHVVPDNQNLEVIHDALADDEEQGEDDDDSGAEPMEPEENNVVPNVE RRGGEDAVAARMAAGHESDDDEWEDLGFDLEEDTVFDLKDVDEWFEQRRLAEKERWHLGQRIVNAYRTE AEVSEAEVEARRINLNTDLSPEWVKSFDFREHFV
US24	Reverse	Tegument protein US24	US22 family	501	MMDPAAGSGPDGAAVVPPELPALPVAAEDPMALYRQVLRDFKELFFCLEPMEITRYVHRNEGRCLSLGPPK GWHVMLRTEDGIITAAKQAASKLICCREPLTPLGYAVILLPEPRRDHHDGMVATPYVVFMGRFSRVYAYDTR EKYMVLVSHNLDELARYGVSRSEIAYRDVIHTTLRRMTVPVPRRYPKGARTMHVLFLNDTTPEGSYATAERIL GCDVKLHTPGYGTVIMRLMKTVQQLHRIWPFCALTEVESRRWWWAVRANLATPWYVLGVTGRPRPGRS FVAEVLVLLDWFGAVYAIQMDDPNHYVRRVANTITEFFRMGLLKMVFRHRRFERERQRQTRMEHRHLCP HHHERAVDHKRDILFNEDAALPDERRERERRILQQQYDWLCLTERFDPHEGAWERLDPNTLVLHRYDTNSQ SYVLDPDIVGVEAAEREAAGHQDDTGPRLHCLVTTRSSTREGAERVITALVHQSRLVTYSDPFPLKSLTGVRE YIQI
US26	Reverse	Protein US26	US22 family	603	MRQSYRYASGAVVRRTLKGLRKLILCQDLRQDIRHLVRSYADMNISLPVSAPPGWRLDFVEFEDIFGSAAVT DGPETPWGQLICCEESLESLGVLQFSTTVLPRVHGPRSSSEDEDSDDDDFFVYVEEIEPPSQARLVLLLGRYET VWCLDRDRGVLYYLAHSLDDFARHGLLHCEAIYGEQMRTPLLTTQPDHIICDLRLHDNSISELQRVTCRYRGE CVPLRTPGEMTRPLLLCGQAENLKGVWPFICMETEQFNDLLKFFVDRLCCETMIMGVVGESLPSGVFHADF VILVDRACEFFYFDVSRREIWRLADSVDMLLTVGLLKIYQAGRRFHYAVDDAERLEVPGRCPHENFPFWDRF GTVERVRASTRHHELRYKWLIRKDRFIVRPDWCSMRNSLDEVSGTADVSWDPRIRPDYPQTSDLECAKQY WQELNDHVREQTARYGPVRRYSVWCGMSSRLERAVKRLQQRIPRQNLMNPSLMNQGLCVYYSDEEEDQ EEDDTSDDDDQEEETENPQNNIGSLTRTPSSPGSLEGVEERMLNVMKEAVAEQDRKKTQKKHKIDTAQRR VLTRRAARAAVLEGRPAPKPTMPHPVSYLPFWM
US27	Forward	Envelope glycoprotein US27	Type 3 membrane protein; 7 transmembrane domains; putative chemokine receptor; possibly involved in intracellular signalling; GPCR family	362	MTTSTNNQTLTQVSNMTNHTLNSTEIYQLFEYTRFGVWLMCIVGTFLNVLVITTILYYRRKKKSPSDTYICNLA VADLLIVVGLPFFLEYAKHHPKLSREVVCSGLNACFYICLFAGVCFLINLSMDRYCVIVWGVELNRVRNNKRAT CWVVIFWILAALMGMPHYLMYSHTNNECVGEFANETSGWFPVFLNTKVNICGYLAPIVLMAYTYNRMVRF IINYVGKWHMQTLHVLLVVVVSFASFWFPFNLALFLESIRLLAGTQNETLQTVITFCLYVGQFLAYVRACLNP GIYILVGTQMRKDMWTTLRVFACCCVKQEIPYQDIDIELQKDIQRRAKHTKRTHYDRKNAPMESGEEEFLL
US28	Forward	Envelope protein US28	Type 3 membrane protein; 7 transmembrane domains; broad spectrum CC chemokine receptor;	354	MTPTTTTAELTTEFDYDEAATPCVFTDVLNQSKPVTLFLYGVVFLFGSIGNFLVIFTITWRRRIQCSGDVYFINL AAADLLFVCTLPLWMQYLLDHNSLASVPCTLLTACFYVAMFASLCFITEIALDRYYAIVYMRYRPVKQACLFSI

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
			constitutive signalling; mediates cellular migration; involved in intracellular signalling; GPCR family		FWWIFAVIIAIPHFMVVTKKDNQCMTDYDYLEVSYPIILNVELMLGAFVIPLSVISYCYYRISRIVAVSQSRHKGRIVRVLIAVVLVFIIFWLPYHLTLFVDTLKLLKWISSSCEFERSLKRALILTESLAFCHCCLNPLLYVFVGTKFRQELHCLLAEFRQRLFSRDVSWYHSMSFSRRSSPSRRETSSDTLSDEVCRVSQIIP
US29	Forward	Membrane protein US29	Contains potential transmembrane domain	462	MRCFRWWLYSGWWWLTFGCARTVTVGFVAPTVRAQSTVVRSEPAPPSETRRDNNDTSYFSGTSFHSSVSPATSVDRQFRRTTYDRWDGRHWLRTRYGNASACVTGTQWSTNFFFSQCEHYPSFVKLNGVQRWTPVRRPMGEVAYYGGCCMVGGGNRAYVILVSGYGTASYGNALRVDFGRGNCTAPKRTYPRRLELHDGRTNPSRCDPYQVYFYGLQCPEQLVITAHGGVGMRRCPTGSRPTPSRPHRHDLENELHGLCVDLLVCVLLLALLLLELVPMEAVRHPLLFWRRVALSPSTSKVDRAVKLCLRRMLGLPPPPSVAPPGEKKELPAQAALSPPLTTWSLPPFPSTRIPDSPPPPYQLRHATSLVTVPTLLLYTSSDIGDTASETTCVAHATYGEPPEPARSTATVQECTVLTAPNCGIVNNDGAVSEGQDHGDAVHHSLDVVSQCTADTGVVNASE
US30	Forward	Membrane protein US30	Type 1 membrane protein	349	MGNPRSPLDRRLRFRNVPFLPPRIAASSTTTARSLKAKTMEMRFTIAWMWFPSVLLILGLLTPPSNGCTVDVGRNMSIREQCRLRNGATFSKGDIEGNFSGPVVVELDYEDVDITGERQRLRFHISGLGCPTKEKIRKDNKSDVNGGIRWALYIQTGDAKYGIRNQHLSIRLMYPGEKDTQQLLGSDFSCERHRRPSTPLGNNAKVSFTTRTSSTYGVLSAFVVWIGSGLNIIWWTGIVLLAADALGLGERWLRLALSHRDKHHASRTAALQCQRDMLLRQRRRARRLHAVSEGKLQEEKKRQSALVWNVEARPFPSTHQLIVLPPPVASAPPAVPSQPPEYSSVFPPV
US31	Forward	Protein US31	Contains duplicated TT virus ORF2 motif; US1 family	161	MSLLEREESWRRVVDYSHNLWCTCGNWQSHVEIQDEEPNCEQPEPAHWLEYVAVQWQARVRDSHDRWCLCNAWRDHALRGRWGTAYSSGSSASSSGFVAESKFTWWKRLRHSTRRWLFRRRRARYTPSNCGESSTSSGQSSGDESNCSLRTHGVYTRGEQH
US32	Forward	Protein US32	Contains duplicated TT virus ORF2 motif; US1 family	183	MAMYTSESERDWRRVIHDSHGLWCDCGDWREHLYCVYDSHFQRRPTTRAERRAANWRRQMRRLHRLWCFCQDWKCHALYAEWDGKESDDESSASSSGEAPEQQVPAWKTVRAFSRAYHHRINRGLRGTPPPRNLPGYEHASEGWRFCSRRERREDDLRTRAEPDRVVFQLGGVPPRRHRETYV
US33A	Forward	Protein US33A		57	MSLRFPERAGYEKLGYRPHAKRVRVHDSLGLTRFIMRQLMMYPLVLPFTFPFYVPRS
US34	Forward	Protein US34	Contains signal peptide	163	MNLEQLINVLGLLVWIAARAVSRVGPHGSGLVYRELHDFYGYLQLDLLGPVVAGNRSVRTWREQADRAKGTFARRSGLNTSHILPVGSMYRGSDTLSAGLYRPEEEVFLLLNRCHGPLSTPKNACLAEVGVANSTFLSRFNVGDFHGASWENGTAPDGELGVC
US34A	Forward	Protein US34A	Contains potential transmembrane domain	64	MLKFLLRFRKRRRPVVVPRFVRFIVYVVLFTVAVQRVKQERDAHLRRYEERLRKNRARRRQSFP
TRS1	Reverse	Tegument protein TRS1	Transcriptional activator; blocks phosphorylation of eIF2alpha and host shutoff of protein synthesis; binds dsRNA; involved in gene regulation; involved in translational regulation; US22 family	800	MAQRNGMSPRPPPLGRGRGAGGPSGVGSSLSSSCVSLGATSTAGTSTSTAGTATPGHGVHRIEPRGPPGAPPGSGNNSNFWHGPERLLLSQIPVERQALTELEYQAMGAVWRAAFLANSTGRAMRKWSQRDAGTLLPLGRPYGFYARVTPRSQMNGVGATDLRQLSPRDAWIVLVATVVHEVDPAADPTVGDKAGHPEGLCAQDGLYLALGAGFRVFVYDLANNTLILAARDADEWFRHGAGEVVRLYRCNRLGVGTPRATLLPQPALRQTLLRAEEATALGRELRRRWAGTTVALQTPGRRLQPMVLLGAWQELAQYEPFASAPHPASLLTAVRRHLNQRLCCGWLALGAVLPARWLGCAAGPATGTAAGTTTMPAGTTSPPGASGTETEAAGGDAPCAIAGAVGSAVTLPPQPYGAAGGSAICAPNADAHAVLGADATAAAAAAAAAPTVMVGPTAMAGPAASGTVPRAMLVVVLDELGAVFGYCPLDGHVYPLAAELSHFLRAGVLGALALGRESAPAEAARRLLPELDREQWERPRWDALHLHPRAALWAREPHGQLAFLLRPGRGEAEVLTLATKHPAICANVEDYLQDARRRADAQALGLDLATVVMEAGGQMIHKKTKKPKGKEDESLMKGKHSRYTRPTEPPLTPQASLGRALRRDDEDWKPPRLPGEDSWYDLDETFWVLGSNRKNDV

Gene	Orientation	Product	Product Features / Known functions	AA length	AA Sequence
					YQRRWKKTVLRCGLEIDRPMPTVPKGCRPQTFTHEGVQLMGGATQEPLDTGLYAPSHVTSAFVPSVYMPP TVPYPDPAARLCRDMRRVTFSNVATHYHYNAQ

Appendix 13. This table outlines the annotated genes in ZM240 and the amino acid sequences of their products.

APPENDIX 14: MANUSCRIPT 1

HIV/AIDS MAJOR ARTICLE

Human Cytomegalovirus Infant Infection Adversely Affects Growth and Development in Maternally HIV-Exposed and Unexposed Infants in Zambia

U. A. Gompels,[1] N. Larke,[3] M. Sanz-Ramos,[1] M. Bates,[1] K. Musonda,[1,4] D. Manno,[2] J. Siame,[5] M. Monze,[4] S. Filteau,[2] and the CIGNIS Study Group[a]

Departments of [1]Pathogen Molecular Biology, [2]Nutrition and Public Health Interventions, and [3]Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, University of London, United Kingdom; and [4]Virology Unit and [5]Department of Pediatrics, University Teaching Hospital, Lusaka, Zambia

Background. Human immunodeficiency virus (HIV) and human cytomegalovirus (HCMV) coinfections have been shown to increase infant morbidity, mortality, and AIDS progression. In HIV-endemic regions, maternal HIV-exposed but HIV-uninfected infants, which is the majority of children affected by HIV, also show poor growth and increased morbidity. Although nutrition has been examined, the effects of HCMV infection have not been evaluated. We studied the effects of HCMV infection on the growth, development, and health of maternally HIV-exposed and unexposed infants in Zambia.

Methods. Infants were examined in a cohort recruited to a trial of micronutrient-fortified complementary foods. HIV-infected mothers and infants had received perinatal antiretroviral therapy to prevent mother-to-child HIV transmission. Growth, development, and morbidity were analyzed by linear regression analyses in relation to maternal HIV exposure and HCMV infection, as screened by sera DNA for viremia at 6 months of age and by antibody for infection at 18 months.

Results. All HCMV-seropositive infants had decreased length-for-age by 18 months compared with seronegative infants (standard deviation [z]-score difference: −0.44 [95% confidence interval {CI}, −.72 to −.17]; $P = .002$). In HIV-exposed infants, those who were HCMV positive compared with those who were negative, also had reduced head size (mean z-score difference: −0.72 [95% CI, −1.23 to −.22]; $P = .01$) and lower psychomotor development (Bayley test score difference: −4.1 [95% CI, −7.8 to −.5]; $P = .03$). HIV-exposed, HCMV-viremic infants were more commonly referred for hospital treatment than HCMV-negative infants. The effects of HCMV were unaffected by micronutrient fortification.

Conclusion. HCMV affects child growth, development, and morbidity of African infants, particularly in those maternally exposed to HIV. HCMV is therefore a risk factor for child health in this region.

Human cytomegalovirus (HCMV) commonly infects children, establishing a lifelong latent infection. It is mainly a pathogen in an immunocompromised setting, particularly for human immunodeficiency virus (HIV)/AIDS and transplantation patients, where HCMV can reactivate. HCMV is the main infectious cause of mental retardation and neurodevelopmental impairment in neonates; this includes hearing loss, chiefly from congenital infection in utero [1–4]. In transplantation patients, HCMV-associated pneumonitis has a high mortality rate that is controlled primarily with ganciclovir. Prior to antiretroviral therapy (ART) in adult HIV/AIDS patients, HCMV was a major

Received 30 June 2011; accepted 7 October 2011.

[a] The members of the CIGNIS Study Group are detailed in the Acknowledgments.

Correspondence: Ursula A. Gompels, PhD, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, Keppel St, London WC1E 7HT, UK (ursula.gompels@lshtm.ac.uk).

Clinical Infectious Diseases 2012;54(3):434–42

© The Author 2011. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please email:journals.permissions@oup.com. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1093/cid/cir837

opportunistic infection that correlated with retinitis, leading to blindness as well as disseminated infections and in turn mortality [5]. In childhood HIV, HCMV coinfection can be particularly severe; studies in the United States showed increased neurological disease, AIDS progression, and mortality [6, 7]. In HIV-endemic sub-Saharan Africa, pediatric ART is restricted and HCMV complications are only beginning to be addressed; even the extent of HCMV effects on the general child population has not been established. In Kenyan infants, HIV infection was associated with impaired control of HCMV replication [8]. In Zambia, autopsy studies of HIV-positive children who died from respiratory disease identified HCMV as a frequent infection, often together with tuberculosis [9, 10].

HCMV may also affect infants born to HIV-positive mothers, even if the infants remain HIV-negative. HIV-positive mothers can have reduced maternal passive immunity to protect their infants against HCMV disease. This includes protective cytokines or antibodies, immunoglobulin G (IgG) transcytosed across the placenta in utero to the fetus, or immunoglobulin A (IgA) secreted in breast milk to protect against gut infection in the newborn. Moreover, immunosuppression in the HIV-positive mother can lead to increased HCMV levels due to reactivation or secondary infection. Little is known about how these effects impact childhood HCMV in HIV-endemic African populations. The number of children who are maternally HIV exposed but not infected is increasing due to the epidemic spread of HIV and also ART introduction to decrease mother-to-child HIV transmission.

We previously reported that in Zambia, where HIV is endemic, children are exposed to a complex collection of HCMV strains, as exists elsewhere, but also show increased prevalence of higher loads transmitted from HIV-positive mothers [10]. HIV-exposed infants do not grow as well as unexposed infants, have increased morbidity, and are particularly vulnerable when they transition from breastfeeding [11, 12]. Here we report the effects of HCMV infection on growth, development, and health of HIV-exposed and unexposed infants in Zambia who were recruited to a double-blind, randomized controlled trial of micronutrient-fortified infant foods [13]. The results identify HCMV as a pathogen associated with poor development including evidence of increased morbidity, irrespective of diet.

METHODS

Study Design, Population, and Recruitment

A total of 811 infants were recruited for a trial of micronutrient-fortified complementary foods. The trial was conducted in Chilenje at the University Teaching Hospital (UTH), Lusaka, Zambia, from October 2005 to July 2009, registered as ISRCTN37460449 (www.controlled-trials.com/mrct), with 77% completion rates as described [13]. The infants were representative of the region, as mothers were recruited during their infant's standard care checks for weighing or vaccinations, and coverage for these services is high in Lusaka.

Follow-up, Anthropometry, and Sera Collection

Anthropometry and standard deviation z scores were calculated using World Health Organization (WHO) growth reference data, as described [13]. Stunting was recorded as length/age $z < -2$. Venous blood was collected in plain vacutainers at age 6 months and 18 months; sera were separated within 2 hours, aliquoted, and stored at $-80°C$.

HIV Antibody Testing

Serum samples at 18 months were tested for HIV antibodies using Determine HIV 1/2 (Inverness Medical). Negative results were recorded; positive results were further tested using Unigold HIV 1/2 (Trinity Biotech). If a negative, discordant result was shown, a third test, SD-Bioline HIV 1/2 (Standard Diagnostics), was used.

Human Cytomegalovirus Antibody and DNA Testing

The ETI-CYTOK-G PLUS enzyme-linked immunosorbent assay kit (DiaSorin) for detection of anti-HCMV IgG in human serum was used in accordance with the manufacturer's instructions. Standard curves were plotted; control sera standards were used to interpolate sample IgG titers with positive antibody titers read above a cutoff of 0.4 IU/mL. Qualitative and quantitative polymerase chain reaction (PCR) was performed on DNA extracted from 200 µL of sera using the QIAamp DNA Blood Mini Kit (Qiagen). PCR was used to amplify HCMV glycoprotein gB (UL55), as described previously [10], with baseline and high-load cutoff sensitivities of 50 and 1000 copies/mL of sera, respectively.

Morbidity Assessments

The clinical officer (J. S.) examined children during scheduled or voluntary visits, with diagnoses and treatment given based on WHO guidelines for Integrated Management of Childhood Illness [13]. Prescriptions for antibiotics or antimalarials were provided at the study clinic. Hospital referrals were defined as those made to the local tertiary facility (UTH), with severe symptoms (inability to drink or breastfeed, severe vomiting, convulsions, respiratory distress, loss of consciousness, or severe lethargy), for surgery, or with an illness requiring consultation with a specialist.

Bayley Testing of Development

Child development tests were administered by 2 trained psychologists using Bayley scales of infant development (BSID II). The mental development index (MDI) and psychomotor development index (PDI) were used as standardized in the United States [14] and age-normalized to a population mean of 100 and SD of 15. Some items were adapted to local settings (doll

appearance or house pictures) while keeping underlying constructs. The test was translated to local languages (Bemba and Nyanja). If a child was sick at the time of assessment, the mother was asked to bring the child back 1–2 weeks after recovery. For logistical reasons, testing was conducted on a subset, including all with HIV-positive mothers and alternate infants with HIV-negative mothers.

Statistical Analysis

Growth and reported illness were compared between infants with and without HCMV using χ^2 tests. The impact of HCMV infection on length-for-age (LAZ), weight-for-age (WAZ), head circumference-for-age (HCZ), and arm circumference-for-age (ACZ) *z* scores at 18 months was analyzed as a continuous variable using linear regression and 95% confidence intervals (CIs) for regression coefficients. The impact on incidence of severe morbidity, defined as hospital referral or death, was estimated using rate ratios (RRs) and 95% CIs obtained using random-effects Poisson regression to account for repeated referrals among infants. Any planned surgeries were excluded (5 circumcisions, 4 hernias, and 4 other congenital abnormalities). It was decided a priori to perform analyses stratified by maternal HIV status; this excluded the 70 infants of HIV-unknown women. Maternal education, socioeconomic status (SES), and diet intervention arm were determined a priori as potential confounders. Adjustments were made for these a priori confounders as well as for breastfeeding duration (shorter or longer than 6 months), which was determined to be associated with HCMV infection in risk factor analyses. SES was measured by an asset index, as described previously [13]. Interobserver reliability of the Bayley test was obtained in 663 testing sessions during which the same child was scored by the 2 psychologists. The intraclass correlation coefficient (ICC) of the scores was calculated using 1-way analysis of variance. A good agreement between scorers was found for both MDI (ICC = 0.91 [95% CI, .89–.92]) and PDI (ICC = 0.91 [95% CI, .90–.92]). MDI and PDI scores were analyzed as continuous variables and checked for normality on a normal probability plot. The association of Bayley test scores and HCMV seropositive antibody at 18 months was examined using linear regression with subsequent adjustment for SES, mother's education, diet intervention arm, and breastfeeding duration. A 2-sided *P* value ≤.05 was considered statistically significant. Stata software (version 11.1, StataCorp) was used in all analyses.

Multiple Imputation Analyses

Among the 811 children enrolled in the study, 561 (69%) were available for HCMV DNA at 6 months, 430 (53%) for HMCV DNA at 18 months, and 460 (57%) for HCMV antibody at 18 months. Any missing data were mainly due to limitations on the infant serum sample volume after the large number of previous metabolic assays, which were used to interpret the primary outcome analyses, as described [13]. Ninety samples were missing due to participant withdrawal from the study. In order to account for infants who were missing data, we used multiple imputations with chained equations methods using Stata software. Factors predictive of missing data (small child), risk factors for HCMV (mother's education, SES, breastfeeding duration, and maternal HIV status), and the micronutrient diet intervention arm were used to impute 10 data sets. All continuous variables used in the imputation were normally distributed; SES and mother's education were treated as ordered categorical variables. In addition to the main analyses of observed data, complete analyses were conducted with all imputed data.

Ethics

The study received approval from the ethics committees of the University of Zambia and the London School of Hygiene and Tropical Medicine. The funding source, the Bill and Melinda Gates Foundation, was not involved in the study or its interpretation.

RESULTS

Growth Effects

In order to analyze the effects of HCMV on growth, 2 measurements were made: HCMV viremia at recruitment (age 6 months; sera DNA) and HCMV infection at the end of the study (age 18 months; sera antibody). The antibody was used to measure infection occurring in the cohort by 18 months; the DNA assay detects HCMV viremia on recruitment at 6 months, either primary or earlier, uncontrolled infections.

Among all children there was no significant effect of HCMV viremia on growth from age 6 months to 18 months. However, when the population was stratified for maternal HIV exposure, there were significant associations, but in opposing directions (Table 1). Among HIV-exposed children with HCMV viremia at 6 months, there was reduced LAZ ($P = .04$), with evidence for interaction of HCMV viremia with maternal HIV on LAZ (interaction, $P = .007$). Similar but nonsignificant directions were seen for WAZ and HCZ. With multiple imputations, the effect on LAZ in HIV-exposed infants remained (regression coefficient of −0.40 to −0.28; interaction, $P = .07$). When HIV-positive children were removed from the analysis, the reduced LAZ effect remained for the maternally HIV-exposed uninfected infants.

In contrast, among children of HIV-negative mothers, HCMV viremia at 6 months was associated with increased LAZ and HCZ. Effects of HCMV viremia in all infants were independent of nutrition, either micronutrient intervention or breastfeeding duration, as adjustments did not alter the associations. Adjustment for hemoglobin had no effect (data not shown).

Next, the effects of HCMV infection on growth by 18 months (sera antibody) were analyzed. HCMV infection affected LAZ,

Table 1. Effect of Human Cytomegalovirus Viremia (Sera DNA) at 6 Months on Anthropometric *z* Scores at 18 Months

	Mean (SD)							
	HCMV Negative	HCMV Positive	Unadjusted Regression Coefficient (95% CI)	*P* Value	Adjusted Regression Coefficient[a] (95% CI)	*P* Value	Adjusted Regression Coefficient[b] (95% CI)	*P* Value
All children	n = 336	n = 226						
Length	−1.05 (1.17)	−1.03 (1.07)	0.02 (−.19 to .23)	.87	0.05 (−.15 to .25)	.61	0.06 (−.14 to .26)	.55
Weight	−0.50 (1.23)	−0.50 (1.14)	0.00 (−.23 to .22)	.99	0.03 (−.18 to .25)	.77	0.04 (−.18 to .25)	.72
Head circumference	0.57 (0.97)	0.61 (1.06)	0.05 (−.14 to .23)	.64	0.08 (−.11 to .26)	.40	0.10 (−.09 to .28)	.30
Arm circumference	0.20 (1.15)	0.01 (1.06)	−0.20 (−.41 to .01)	.07	−0.17 (−.37 to .04)	.11	−0.16 (−.36 to .05)	.13
Maternal HIV negative	n = 241	n = 152						
Length	−1.06 (1.19)	−0.81 (1.00)	0.25 (.00–.51)	.06	0.24 (.00 to .48)	.05	0.24 (.00 to .48)	.05
Weight	−0.48 (1.25)	−0.28 (1.07)	0.19 (–.08 to .47)	.17	0.19 (–.07 to .45)	.16	0.19 (−.07 to .46)	.15
Head circumference	0.62 (1.00)	0.84 (1.01)	0.22 (−.01 to .45)	.06	0.23 (.00 to .46)	.05	0.24 (.01 to .46)	.04
Arm circumference	0.27 (1.13)	0.16 (0.95)	−0.11 (−.37 to .15)	.40	−0.10 (−.35 to .15)	.43	−0.10 (−.35 to .15)	.43
Maternal HIV positive	n = 65	n = 55						
Length	−1.09 (1.07)	−1.54 (1.14)	−0.45 (−.89 to –.02)	.04	−0.40 (−.80 to .01)	.05	−0.40 (−.80 to .00)	.05
Weight	−0.58 (1.16)	−0.94 (1.30)	−0.36 (−.82 to .10)	.13	−0.31 (−.75 to .13)	.16	−0.32 (−.76 to .13)	.16
Head circumference	0.42 (0.91)	0.25 (1.09)	−0.17 (−.56 to .22)	.40	−0.13 (−.51 to .25)	.50	−0.12 (−.50 to .26)	.54
Arm circumference	0.01 (1.28)	−0.36 (1.27)	−0.36 (−.80 to .07)	.10	−0.33 (−.75 to .09)	.13	−0.33 (−.75 to .09)	.12

P values for interaction HCMV viremia with maternal HIV: length-for-age = .007, weight-for-age = .05, head circumference-for-age = .11, and arm circumference-for-age = .35. Infants of women with unknown HIV status were excluded from the stratified analyses.

Abbreviations: CI, confidence interval; HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; SD, standard deviation.

[a] Adjusted for maternal socioeconomic status and education.

[b] Additionally adjusted for diet treatment arm and breastfeeding duration.

with reduced measures in all children (Table 2). The effects remained after adjusting for SES, education, breastfeeding duration, and diet treatment arm (adjusted regression coefficient, −0.33 [95% CI, −.60 to −.06]; *P* = .01; adjusted regression coefficient was 0.04 after Bonferroni correction). The results were similar after multiple imputations for missing data (data not shown). There was 3-fold more stunting (LAZ <−2) associated with HCMV infection in maternally HIV-exposed infants and unexposed infants (Figure 1). The other anthropometric measures were also negatively associated with HCMV infection by 18 months; only results for HCZ were statistically significant, with the greatest effects among HIV-exposed children (regression coefficient, −0.71; interaction *P* = .02). This HCZ effect was similar after multiple imputations, although significance was borderline (regression coefficient, −0.57; *P* = .09). Because maternal HIV exposure has been shown to increase the prevalence of congenital HCMV infection [7, 15], and this may affect head size, an additional adjustment was made for viremia at 6 months, and this would include congenital infection. Adjustment strengthens the association (regression coefficient, −0.80 [95% CI, −1.40 to −.20]), suggesting that overall HCMV infection from 6 to 18 months is a main effect.

Morbidity Effects

Table 3 shows effects of HCMV viremia and infection on infant hospitalization or death. Although no associations were statistically significant, the directions were consistent with HCMV viremia at recruitment (age 6 months), with fewer hospitalizations among HIV-unexposed infants and increased hospitalizations for HIV-exposed or HIV-infected infants. The increased hospitalizations were concentrated in infants coinfected with HIV and HCMV. Similar results are shown using a higher threshold for copy number of HCMV DNA in sera of 1000 copies/mL and after further sensitivity analyses (*P* = .09).

Development Effects

Mental and psychomotor development was assessed using Bayley tests. HIV-exposed infants with HCMV viremia at 6 months showed a borderline significance for decreased psychomotor skills (PDI) (Table 4). By 18 months, HIV-exposed infants with HCMV infection detected by antibody had significantly

Table 2. Association of Infant Human Cytomegalovirus Infection (Antibody) on Anthropometric *z* Scores at 18 Months

	Mean (SD)		Unadjusted Mean		Adjusted Mean		Adjusted Mean	
	HCMV Negative	HCMV Positive	Difference (95% CI)	*P* Value	Difference[a] (95% CI)	*P* Value	Difference[b] (95% CI)	*P* Value
All children	n = 76	n = 384						
Length	−0.72 (1.07)	−1.17 (1.11)	−0.44 (−.72 to −.17)	.002	−0.31 (−.58 to −.05)	.02	−0.33 (−.60 to −.06)	.01
Weight	−0.34 (1.10)	−0.64 (1.15)	−0.30 (−.58 to −.01)	.04	−0.15 (−.42 to .13)	.29	−0.17 (−.45 to .11)	.22
Head circumference	0.79 (1.00)	0.49 (0.99)	−0.31 (−.55 to −.06)	.01	−0.21 (−.45 to .04)	.10	−0.25 (−.49 to .00)	.05
Arm circumference	0.05 (1.14)	0.03 (1.06)	−0.01 (−.28 to .25)	.92	0.11 (−.15 to .38)	.40	0.06 (−.21 to .33)	.65
Maternal HIV negative	n = 49	n = 263						
Length	−0.63 (1.06)	−1.13 (1.13)	−0.49 (−.83 to −.15)	.01	−0.38 (−.71 to −.06)	.02	−0.38 (−.71 to −.06)	.02
Weight	−0.26 (1.05)	−0.59 (1.15)	−0.33 (−.68 to .03)	.07	−0.19 (−.53 to .16)	.29	−.19 (−.53 to .16)	.29
Head circumference	0.74 (0.95)	0.59 (1.00)	−0.15 (−.46 to .15)	.33	−0.05 (−.35 to .26)	.77	−0.04 (−.34 to .27)	.82
Arm circumference	0.18 (1.07)	0.13 (1.00)	−0.06 (−.39 to .28)	.74	0.08 (−.25 to .40)	.65	0.08 (−.25 to .42)	.62
Maternal HIV positive	n = 19	n = 86						
Length	−0.85 (1.11)	−1.35 (1.10)	−0.49 (−1.05 to .06)	.08	−0.29 (−.82 to .23)	.27	−0.29 (−.83 to .25)	.30
Weight	−0.56 (1.20)	−0.79 (1.22)	−0.23 (−.81 to .34)	.43	−0.02 (−.57 to .53)	.93	−0.08 (−.65 to .49)	.78
Head circumference	0.95 (1.19)	0.25 (0.98)	−0.71 (−1.21 to −.21)	.01	−0.60 (−1.08 to −.12)	.02	−0.72 (−1.23 to −.22)	.01
Arm circumference	−0.23 (1.38)	−0.21 (1.23)	0.02 (−.52 to .56)	.93	0.17 (−.35 to .69)	.52	0.07 (−.48 to .62)	.80

P values for interaction HCMV antibody with maternal HIV: length-for-age – .76, weight-for-age – .76, head circumference-for-age – .02, and arm circumference-for-age – .97. Infants of women with unknown HIV status were excluded from the stratified analyses.

Abbreviations: CI, confidence interval; HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; SD, standard deviation.

[a] Adjusted for maternal socioeconomic status and education.

[b] Additionally adjusted for diet treatment arm and breastfeeding duration.

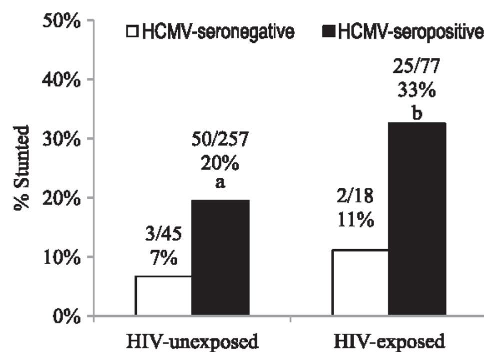


Figure 1. Human cytomegalovirus (HCMV) seroprevalence correlates with stunting at 18 months. Stunting (length/age $z < -2$) prevalence at 18 months is tripled in HCMV-seropositive infants. Significance is by Pearson χ^2: [a] Human immunodeficiency virus (HIV)–unexposed, $P = .037$; [b] HIV-exposed, $P = .071$.

lower PDI after adjustment for SES, education, breastfeeding duration, and diet intervention (regression coefficient, −4.1 [95% CI ,−7.8 to −.5]; $P = .03$) (Table 5). HCMV did not affect PDI in HIV-unexposed infants or mental development in either subgroup.

DISCUSSION

The results show that HCMV is a widespread pathogen that affects healthy growth and development of infants in Zambia. In the maternally HIV-exposed infants, HCMV infection was also associated with increased prevalence of stunting, reduced head size, and decreased psychomotor development. Previous analyses of this cohort showed that reduced growth in the HIV-exposed infants was unaffected by micronutrient fortification but was affected by socioeconomic factors [13, 16]. The effects of HCMV on growth were also influenced by socioeconomic factors; however, the effects still existed after adjustments were made for socioeconomic factors, education level, breastfeeding duration, and diet. In Zambia we showed that HCMV infections occurred earlier, that is, during infancy, and were more prevalent (83% IgG seropositive in this cohort) than in many other geographic regions. Therefore, effects of HCMV on growth could affect the majority of the infant population. Moreover, growth effects at such an early age have been linked with lifetime risks of adult chronic disease and thus HCMV infection is a general public health concern [17]. In the United States, children studied from age 6 years, showed reduced growth associated with increased infectious burden, including HCMV (30%–50% seropositive at outset). This resulted in health disparities that persisted to middle age [18, 19].

HIV-unexposed infants with HCMV viremia at 6 months had higher length and head circumference *z* scores, whereas among maternally HIV-exposed infants, HCMV viremia was associated with lower *z* scores. This suggests that HIV-uninfected mothers, but not infected, provide some protective maternal immunity, permitting infants to tolerate HCMV infection. In the HIV-unexposed infants, HCMV may provide beneficial immune stimulation at this age, possibly contributing to the

Table 3. Effect of Human Cytomegalovirus Viremia at 6 Months (Sera DNA) on Hospitalization

	No./Person-Years (Rate per 100 py)							
	HCMV Negative	HCMV Positive	Unadjusted RR (95% CI)	*P* Value	Adjusted RR (95% CI)[a]	*P* Value	Adjusted RR (95% CI)[b]	*P* Value
Standard threshold (50 copies/mL)								
All children	61/288.5 (21.1)	43/204.6 (21.0)	1.00 (.65–1.53)	>.99	1.04 (.68–1.59)	.86	1.04 (.68–1.59)	.85
Maternal HIV								
Negative	44/202.4 (21.7)	21/134.2 (15.7)	0.72 (.41–1.25)	.24	0.75 (.43–1.30)	.31	0.74 (.43–1.29)	.29
Positive	15/59.9 (25.0)	19/52.0 (36.5)	1.49 (.70–3.16)	.30	1.62 (.77–3.43)	.21	1.53 (.72–3.27)	.27
Child HIV								
Negative	11/50.3 (21.9)	7/43.1 (16.2)	0.76 (.29–1.96)	.57	0.90 (.34–2.36)	.83	0.96 (.36–2.57)	.94
Positive	1/4.5 (22.4)	5/6.3 (79.9)	3.56 (.41–.80)	.25	3.11 (.35–27.40)	.31	5.08 (.55–47.30)	.15
High threshold (>1000 copies/mL)								
All children	61/288.5 (21.1)	6/31.0 (19.4)	0.93 (.38–2.31)	.88	0.81 (.32–2.03)	.65	0.86 (.34–2.18)	.75
Maternal HIV								
Negative	44/202.4 (21.7)	2/18.5 (10.8)	0.49 (.11–2.16)	.35	0.47 (.10–2.12)	.33	0.48 (.11–2.12)	.33
Positive	15/59.9 (25.0)	4/11.5 (34.8)	1.48 (.43–5.13)	.54	1.06 (.30–3.80)	.93	1.28 (.36–4.54)	.70

Abbreviations: CI, confidence interval; HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; py, person-years; RR, rate ratio.

[a] Adjusted for mother's socioeconomic status and education.

[b] Additionally adjusted for breastfeeding duration (<6 months/≥6 months) and diet treatment arm. Maternal HIV-exposure interaction with HCMV in a fully adjusted model has a *P* value = .10.

Table 4. Effect of Human Cytomegalovirus Viremia at 6 Months (Sera DNA) on Development at 18 Months

	Mean (SD)		Unadjusted Regression Coefficient[a] (95% CI)	*P* Value	Adjusted Regression Coefficient[b] (95% CI)	*P* Value
	HCMV Negative	HCMV Positive				
All children	n = 155	n = 121				
MDI	89.2 (7.8)	89.2 (7.9)	0.0 (−1.9 to 1.8)	.97	−0.1 (−1.8 to 1.9)	.93
PDI	91.0 (5.3)	90.4 (7.0)	−0.6 (−2.0 to .9)	.46	−0.7 (−2.1 to .8)	.35
Maternal HIV negative	n = 78	n = 57				
MDI	89.3 (7.9)	90.6 (6.9)	1.3 (−1.3 to 3.9)	.32	1.0 (−1.6 to 3.6)	.45
PDI	91.1 (5.2)	91.7 (5.9)	0.6 (−1.3 to 2.5)	.53	0.2 (−1.7 to 2.2)	.82
Maternal HIV positive	n = 55	n = 49				
MDI	88.0 (6.9)	87.4 (8.9)	−0.6 (−3.7 to 2.4)	.68	−0.6 (−3.8 to 2.6)	.73
PDI	90.9 (5.7)	88.6 (8.2)	−2.3 (−5.0 to .4)	.09	−2.5 (−5.4 to .3)	.08

P values for interaction between human cytomegalovirus viremia and maternal human immunodeficiency virus on mental development index = 0.03, on psychomotor development index = 0.07.

Abbreviations: CI, confidence interval; HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; MDI, mental development index; PDI, psychomotor development index; SD, standard deviation.

[a] Mean difference.

[b] Adjusted for socioeconomic status, mother's education, diet intervention arm, and breastfeeding duration.

high HCMV seroprevalence in this population. There are similar examples with other sequential infections; for example, prior influenza infection affects outcome for subsequent respiratory syncytial virus infection [20]. HCMV encodes a number of immunomodulatory proteins that favor virus persistence, but these proteins may also affect responses to other infections and host survival [21–23]. However, even if beneficial in the short term, long-term HCMV immune stimulation during infancy can be detrimental by driving early immune senescence [24]. In West Africa, Malawi, and Gambia, HCMV-infected infants had expanded HCMV-specific T lymphocytes with markers of aged anergic cells, which have been previously identified in elderly European populations where HCMV is acquired later [25, 26]. HCMV immune senescence may compromise the ability to respond to new infections during infancy in Africa, as suggested for the elderly in Europe [24–26]. Therefore, increased overall burden of all infections may affect healthy growth. This is supported by the 18-month HCMV antibody data, which shows that when overall HCMV infection is considered at times when maternal immunity wanes, there is a negative effect on growth.

HCMV may also directly affect growth and development, depending on the route of infection. HCMV frequently reactivates in HCMV IgG-positive mothers during lactation, and the virus

Table 5. Effect of Human Cytomegalovirus Infection (Antibody) on Development at 18 Months

	Mean (SD)		Unadjusted Regression Coefficient[a] (95% CI)	*P* Value	Adjusted Regression Coefficient[b] (95% CI)	*P* Value
	HCMV Negative	HCMV Positive				
All children	n = 48	n = 239				
MDI	89.2 (7.3)	88.8 (7.8)	−0.4 (−2.8 to 2.0)	.73	−0.1 (−2.7 to 2.5)	.97
PDI	92.2 (4.3)	90.4 (6.2)	−1.8 (−3.6 to .0)	.05	−2.0 (−4.0 to .0)	.05
Maternal HIV negative	n = 22	n = 121				
MDI	90.9 (6.6)	88.9 (8.2)	−2.0 (−5.7 to 1.7)	.28	−0.5 (−4.2 to 3.3)	.79
PDI	92.5 (4.4)	90.8 (6.3)	−1.7 (−4.5 to 1.1)	.22	−1.0 (−3.8 to 1.8)	.48
Maternal HIV positive	n = 19	n = 86				
MDI	87.5 (6.7)	87.8 (7.1)	0.3 (–3.2 to 3.8)	.86	0.8 (-3.4 to 5.0)	.71
PDI	92.1 (4.1)	89.2 (6.5)	−2.8 (−5.9 to .2)	.07	−4.1 (−7.8 to −.5)	.03

P values for interaction between human cytomegalovirus exposure and maternal human immunodeficiency virus on mental development index = 0.12, on psychomotor development index = 0.13.

Abbreviations: CI, confidence interval; HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; MDI, mental development index; PDI, psychomotor development index; SD, standard deviation.

[a] Mean difference.

[b] Adjusted for socioeconomic status, mother's education, diet intervention arm, and breastfeeding duration.

is detectable in milk whey. Furthermore, HCMV is a pathogen of the gastrointestinal tract of HIV/AIDS patients and so may alter nutrient absorption, and can invade the central nervous system, possibly affecting development [3, 4, 27, 28]. HCMV infection acquired in utero can cause neurological disease, while later postnatal infections, primarily via breast milk, are generally asymptomatic [5]. In utero infection may account for HCMV effects on reduced head size and psychomotor development in HIV-exposed infants because adjustments for breastfeeding increased the effect. However, in separate analyses, we showed that the major route for infection in this cohort is via breastfeeding and that its duration over age 6 months is a risk factor, particularly for HIV-exposed infants. Studies in Europe of preterm births show symptomatic HCMV infection acquired via breast milk is associated with a growth disadvantage [29, 30]. High HCMV load and prolonged excretion in breast milk are risk factors for acquisition [27]; both factors were apparent for this HIV-exposed Zambian cohort.

Maternal HIV exposure increased some adverse effects of HCMV. Our previous analyses indicated that maternal HIV exposure led to a greater prevalence of high-load viremic infant HCMV infections [10]. High-load HCMV viremia can correlate with symptomatic infection [4, 31, 32]. Studies in Kenya of HIV-positive mother/child pairs showed that detection of maternal HCMV plasma DNA was associated with infant mortality [33]. In our Zambian cohort, there was some evidence of increased hospitalization with HCMV viremia in the HIV-infected children, but this was not significant with maternal HIV exposure. Therefore, by affecting other diseases, the effects of HCMV on growth and development could be partially indirect; the effects appear to be a combination of direct and indirect causes. A limitation of our study was that the maternal CD4 count or HCMV status was not available and we could not determine the effects of HIV on maternal immunity or HCMV. Study strengths include analyses of HCMV effects in infants using both 6-month DNA for viremia and 18-month antibody for infection, with comparisons between maternally HIV-exposed and HIV-unexposed infants, controls for major confounding factors, and a large population-based study, the first in this region.

What may limit the effects of HCMV on infant growth and development in this region? Diagnostics for symptomatic congenital infections and their treatment with ganciclovir or its derivatives may be appropriate [34, 35]. Recent evidence shows that high-avidity HCMV IgG can cross the placenta and contribute to viral replication control [36]. This is an indication that hyperimmune globulin could also limit placental infection [35], and so improving maternal immunity could limit infection. In France, ART of HIV-positive women decreased HCMV congenital infection [37]. For HIV-positive mothers, limiting breastfeeding duration over 6 months could reduce infant HCMV infection; however, further study is required to assess the effects. In Zambia, HIV-positive women with advanced HIV/AIDS improved infant survival by decreasing breastfeeding, while those with less advanced HIV/AIDS improved survival of HIV-positive infants by increasing breastfeeding duration to longer than 4 months [38, 39]. Because HCMV is frequently secreted in urine, improved hygiene could limit horizontal transmission. Preliminary analyses showed lower HCMV infant infection with the presence of private indoor toilets compared with outdoor toilets. Ultimately, HCMV vaccination could control new infections or reduce viremia in those already infected. Early trials of a subunit vaccine show promise [40, 41].

In conclusion, HCMV infection in Zambia resulted in poor growth and development of maternally HIV-exposed infants and poor growth of HIV-unexposed children. Other African countries also have high HCMV childhood infection and endemic HIV. Therefore, efforts to improve child health will need to consider HCMV, particularly for effects on growth and development.

Notes

Acknowledgments. We thank all the mothers and children who participated in the study and are grateful for the support of Lusaka District Health staff who permitted the project. Thanks are extended to the entire CIGNIS study team: Principal Investigator: Suzanne Filteau, London School of Hygiene and Tropical Medicine (LSHTM); Zambian Lead Investigator: Lackson Kasonka, University Teaching Hospital (UTH), Lusaka; Senior Investigators: Rosalind Gibson, University of Otago, New Zealand; Ursula A. Gompels, LSHTM; Shabbar Jaffar, LSHTM; Emmanuel Kafwembe, Tropical Diseases Research Centre, Ndola; Mwaka Monze, UTH; Moses Sinkala, Catholic Relief Services, Zambia; Andrew Tomkins, Institute of Child Health, University College, London; Rodah Zulu, National Institute of Science and Industrial Research, Zambia; Clinic Coordinator: Molly Chisenga; Clinical Officer: Joshua Siame; Data Manager: Hildah Banda Mabuda; Statisticians: Kathy Baisley, Helen Dale, Natasha Larke, Daniela Manno, Andrea Rehman; Research Fellows: Matthew Bates, Anne Mullen, Kunda Musonda, Marta Sanz-Ramos; Clinic Staff: Hellen Kangwa Bwalya, Margaret Chileshe, Priscilla Kangwa Kowa, Mabvuto Kumwenda, Munalula Likando, Sydney Mambwe, Mutinta Muzyamba, Anne Mwale, Lungowe Nyaywa; Laboratory Staff: Humphrey Bima, Julia Chibumbya, Laura Gosset, Louise Hackett, Abigail Jackson, Mirriam Kapambwe, Sydney Mwanza, Ida Ndumba, Eric Njunju; Data Entry: Concillia Kabanga, Natalia Shampwaya; Drivers and Cleaners: John Chobo, Winford Kapumba, Charity Musonda, Philip Soko.

Financial support. This work was supported by the Bill and Melinda Gates Foundation, grant ID 37253.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Foulon I, Naessens A, Foulon W, Casteels A, Gordts F. A 10-year prospective study of sensorineural hearing loss in children with congenital cytomegalovirus infection. J Pediatr **2008**; 153:84–8.
2. Griffiths PD, Walter S. Cytomegalovirus. Curr Opin Infect Dis **2005**; 18:241–5.
3. Koyano S, Inoue N, Nagamori T, et al. Dried umbilical cords in the retrospective diagnosis of congenital cytomegalovirus infection as a cause of developmental delays. Clin Infect Dis **2009**; 48:e93–5.
4. Rosenthal LS, Fowler KB, Boppana SB, et al. Cytomegalovirus shedding and delayed sensorineural hearing loss: results from longitudinal

follow-up of children with congenital infection. Pediatr Infect Dis J **2009**; 28:515–20.
5. Boppana SB, Fowler KB. Persistence in the population: epidemiology and transmission. In: Arvin A, Campadelli-Fiume G, Mocarski ES Jr, et al, eds. Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge: Cambridge University Press, **2007.**
6. Kovacs A, Schluchter M, Easley K, et al. Cytomegalovirus infection and HIV-1 disease progression in infants born to HIV-1-infected women. Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted HIV Infection Study Group. N Engl J Med **1999**; 341:77–84.
7. Doyle M, Atkins JT, Rivera-Matos IR. Congenital cytomegalovirus infection in infants infected with human immunodeficiency virus type 1. Pediatr Infect Dis J **1996**; 15:1102–6.
8. Slyker JA, Lohman-Payne BL, John-Stewart GC, et al. Acute cytomegalovirus infection in Kenyan HIV-infected infants. AIDS **2009**; 23:2173–81.
9. Chintu C, Mudenda V, Lucas S, et al. Lung diseases at necropsy in African children dying from respiratory illnesses: a descriptive necropsy study. Lancet **2002**; 360:985–90.
10. Bates M, Monze M, Bima H, Kapambwe M, Kasolo FC, Gompels UA. High human cytomegalovirus loads and diverse linked variable genotypes in both HIV-1 infected and exposed, but uninfected, children in Africa. Virology **2008**; 382:28–36.
11. Filteau S. The HIV-exposed, uninfected African child. Trop Med Int Health **2009**; 14:276–87.
12. Koyanagi A, Humphrey JH, Ntozini R, et al. Morbidity among human immunodeficiency virus-exposed but uninfected, human immunodeficiency virus-infected, and human immunodeficiency virus-unexposed infants in Zimbabwe before availability of highly active antiretroviral therapy. Pediatr Infect Dis J **2011**; 30:45–51.
13. CIGNIS Study Team. Micronutrient fortification to improve growth and health of maternally HIV-unexposed and exposed Zambian infants: a randomised controlled trial. PLoS One **2010**; 5:e11165.
14. Lennon EM, Gardner JM, Kamel B. Bayley scales of infant and early childhood development. In: Encyclopedia of infant and early childhood development. **2008**:145–56.
15. Duryea EL, Sanchez PJ, Sheffield JS, et al. Maternal human immunodeficiency virus infection and congenital transmission of cytomegalovirus. Pediatr Infect Dis J **2010**; 29:915–8.
16. Filteau S, Baisley K, Chisenga M, Kasonka L, Gibson RS. CIGNIS Study Team. Provision of micronutrient-fortified food from 6 months of age does not permit HIV-exposed, uninfected Zambian children to catch up in growth to HIV-unexposed children: a randomised controlled trial. J Acquir Immune Defic Syndr **2011**; 56:166–75.
17. Barker DJ, Osmond C, Kajantie E, Eriksson JG. Growth and chronic disease: findings in the Helsinki Birth Cohort. Ann Hum Biol **2009**; 36:445–58.
18. Dowd JB, Aiello AE, Alley DE. Socioeconomic disparities in the seroprevalence of cytomegalovirus infection in the US population: NHANES III. Epidemiol Infect **2009**; 137:58–65.
19. Dowd JB, Zajacova A, Aiello A. Early origins of health disparities: burden of infection, health, and socioeconomic status in U.S. children. Soc Sci Med **2009**; 68:699–707.
20. Walzl G, Tafuro S, Moss P, Openshaw PJ, Hussell T. Influenza virus lung infection protects from respiratory syncytial virus-induced immunopathology. J Exp Med **2000**; 192:1317–26.
21. Mocarski ES Jr. Immune escape and exploitation strategies of cytomegaloviruses: impact on and imitation of the major histocompatibility system. Cell Microbiol **2004**; 6:707–17.
22. Wilkinson GW, Tomasec P, Stanton RJ, et al. Modulation of natural killer cells by human cytomegalovirus. J Clin Virol **2008**; 41:206–12.
23. Boeckh M, Geballe AP. Cytomegalovirus: pathogen, paradigm, and puzzle. J Clin Invest **2011**; 121:1673–80.
24. Pawelec G, Derhovanessian E, Larbi A, Strindhall J, Wikby A. Cytomegalovirus and human immunosenescence. Rev Med Virol **2009**; 19:47–56.
25. Ben-Smith A, Gorak-Stolinska P, Floyd S, et al. Differences between naive and memory T cell phenotype in Malawian and UK adolescents: a role for cytomegalovirus? BMC Infect Dis **2008**; 8:139.
26. Miles DJ, van der Sande M, Jeffries D, et al. Cytomegalovirus infection in Gambian infants leads to profound CD8 T-cell differentiation. J Virol **2007**; 81:5766–76.
27. Jim WT, Shu CH, Chiu NC, et al. High cytomegalovirus load and prolonged virus excretion in breast milk increase risk for viral acquisition by very low birth weight infants. Pediatr Infect Dis J **2009**; 28:891–4.
28. Meier J, Lienicke U, Tschirch E, Kruger DH, Wauer RR, Prosch S. Human cytomegalovirus reactivation during lactation and mother-to-child transmission in preterm infants. J Clin Microbiol **2005**; 43: 1318–24.
29. Hamprecht K, Maschmann J, Jahn G, Poets CF, Goelz R. Cytomegalovirus transmission to preterm infants during lactation. J Clin Virol **2008**; 41:198–205.
30. Hamprecht K, Maschmann J, Vochem M, Dietz K, Speer CP, Jahn G. Epidemiology of transmission of cytomegalovirus from mother to preterm infant by breastfeeding. Lancet **2001**; 357:513–8.
31. Kidd IM, Fox JC, Pillay D, Charman H, Griffiths PD, Emery VC. Provision of prognostic information in immunocompromised patients by routine application of the polymerase chain reaction for cytomegalovirus. Transplantation **1993**; 56:867–71.
32. Lanari M, Lazzarotto T, Venturi V, et al. Neonatal cytomegalovirus blood load and risk of sequelae in symptomatic and asymptomatic congenitally infected newborns. Pediatrics **2006**; 117:e76–83.
33. Slyker JA, Lohman-Payne BL, Rowland-Jones SL, et al. The detection of cytomegalovirus DNA in maternal plasma is associated with mortality in HIV-1-infected women and their infants. AIDS **2009**; 23:117–24.
34. Nassetta L, Kimberlin D, Whitley R. Treatment of congenital cytomegalovirus infection: implications for future therapeutic strategies. J Antimicrob Chemother **2009**; 63:862–7.
35. Adler SP, Nigro G, Pereira L. Recent advances in the prevention and treatment of congenital cytomegalovirus infections. Semin Perinatol **2007**; 31:10–8.
36. Nozawa N, Fang-Hoover J, Tabata T, Maidji E, Pereira L. Cytomegalovirus-specific, high-avidity IgG with neutralizing activity in maternal circulation enriched in the fetal bloodstream. J Clin Virol **2009**; 46(Suppl 4):S58–63.
37. Guibert G, Warszawski J, Le Chenadec J, et al. Decreased risk of congenital cytomegalovirus infection in children born to HIV-1-infected mothers in the era of highly active antiretroviral therapy. Clin Infect Dis **2009**; 48:1516–25.
38. Kuhn L, Aldrovandi GM, Sinkala M, et al. Differential effects of early weaning for HIV-free survival of children born to HIV-infected mothers by severity of maternal disease. PLoS One **2009**; 4:e6059.
39. Kuhn L, Aldrovandi GM, Sinkala M, et al. Effects of early, abrupt weaning on HIV-free survival of children in Zambia. N Engl J Med **2008**; 359:130–41.
40. Pass RF, Zhang C, Evans A, et al. Vaccine prevention of maternal cytomegalovirus infection. N Engl J Med **2009**; 360:1191–9.
41. Sabbaj S, Pass RF, Goepfert PA, Pichon S. Glycoprotein B vaccine is capable of boosting both antibody and CD4 T-cell responses to cytomegalovirus in chronically infected women. J Infect Dis **2011**; 203:1534–40.

APPENDIX 15: MANUSCRIPT 2

Vaccine 31 (2013) 2042–2049

Contents lists available at SciVerse ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Reduced Poliovirus vaccine neutralising-antibody titres in infants with maternal HIV-exposure

Marta Sanz-Ramos [a,1], Daniela Manno [b,1], Mirriam Kapambwe [c], Ida Ndumba [c], Kunda G. Musonda [a,c], Matthew Bates [a], Julia Chibumbya [c], Joshua Siame [d], Mwaka Monze [c], Suzanne Filteau [b], Ursula A. Gompels [a,*], CIGNIS study team

[a] *Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, UK*
[b] *Department of Nutrition & Public Health Intervention Research, London School of Hygiene and Tropical Medicine, UK*
[c] *Virology Unit, University Teaching Hospital, Lusaka, Zambia*
[d] *Chilenje Clinic, University Teaching Hospital, Lusaka, Zambia*

ARTICLE INFO

Article history:
Received 21 September 2012
Received in revised form 4 February 2013
Accepted 25 February 2013
Available online 6 March 2013

Keywords:
Maternal HIV exposure
Childhood HIV
Polio vaccine
Human cytomegalovirus
Micronutrient
Breast-feeding

ABSTRACT

Background: Maternally HIV-exposed (mHIV-EU) infants have poor health even without HIV-1 infection. The responses to vaccination are less well defined. Immunity to oral Poliovirus vaccine (OPV) was studied in Zambian infants participating in a randomised controlled trial of micronutrient fortification to improve child health.

Method: Maternally HIV-unexposed and mHIV-EU infants were recruited at 6 months age and randomised to basal or enriched micronutrient-fortified diets for 12 months. HIV-exposed mother–infant pairs had received perinatal nevirapine to prevent mother-to-child-transmission. In the cohort of 597 infants, neutralising-antibody titres to OPV were analysed at 18 months with respect to micronutrient fortification, maternal or infant HIV-1 infection, and human cytomegalovirus (HCMV) infection detected by antibodies and viraemia (serum DNA). Vaccine protection was defined as $\log_2$ titre > 3.

Results: Compared to uninfected children, HIV-1-infected children had reduced neutralising antibody titres to OPV, irrespective of diet: $\log_2$ titre difference (95% confidence interval) −3.44 (−2.41; −4.46), $P < 0.01$. OPV antibody titres were lower in HIV-infected children with HCMV viraemia compared to those without viraemia at 18 months, but did not reach significance: difference −2.55 (−6.10; 1.01), $P = 0.14$. Breast-feeding duration was independently associated with increasing OPV titre (P-value < 0.01). In mHIV-EU children there were reduced neutralising antibody titres to Poliovirus compared with maternally HIV-unexposed, irrespective of diet, maternal education and socioeconomic status: $\log_2$ titre difference (95% confidence interval) −0.56 (−0.98; −0.15), $P < 0.01$. This difference was noticeably decreased after adjusting for breast-feeding duration, suggesting that in our study population less breast-feeding by HIV-positive mothers could explain the reduced OPV titres in mHIV-EU infants.

Conclusion: The mHIV-EU infants had reduced polio vaccine antibody titres which were associated with reduced breast-feeding duration. This has important implications for polio eradication and control of vaccine-preventable diseases, in countries where childhood HIV-1 infection and maternal exposure are public health threats.

© 2013 Published by Elsevier Ltd.

1. Introduction

HIV-1 exposure in infants is a major public health challenge in developing countries, particularly in Sub-Saharan Africa, which has 90% of infected paediatric cases worldwide. Limited access to preventive methods and antiretroviral treatments increases HIV-1 transmission and worsens disease outcome [1]. Children with HIV-1/AIDS may show lower protection against vaccine-preventable infections relative to HIV-1 uninfected children. Causes include lower vaccine coverage [2], and impaired capacity to produce or maintain adequate levels of antibodies against vaccines [3–7]. Additional threats are opportunistic pathogens. Human cytomegalovirus (HCMV) is a major viral opportunistic infection in HIV-1 positive infants, and has been associated with HIV-1/AIDS progression, severe morbidity and mortality [8,9]. Furthermore, in Europe, in elderly populations, HCMV is associated with 'senescence' of the immune system leading to decreased naïve

* Corresponding author at: Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, University of London, Keppel St., London WC1E 7HT, UK. Tel.: +44 02079272315.
E-mail address: ursula.gompels@lshtm.ac.uk (U.A. Gompels).
[1] These authors had equal contributions.

0264-410X/$ – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.vaccine.2013.02.044

T-lymphocytes, increased differentiated memory T-lymphocytes, and reduced ability to respond to new infections [10]. Leukocytes with these phenotypes are also observed in African countries, in children from the Gambia and Malawi, after HCMV infections which here are earlier, during infancy, and more prevalent, suggesting effects on these cells at an early age could affect response to other infections during childhood [11,12]. In studies of infants with congenital HCMV infections, there were reduced lymphocyte responses to HCMV, but not to herpes simplex virus [13]. It is possible that the modified leukocytes identified in postnatally acquired HCMV African infants may also affect immunity to other challenges, including vaccines.

Recent studies show HIV-1-negative infants born to HIV-1-positive mothers (maternally exposed uninfected, mHIV-EU) also display altered immunity [14–16] and increased disease risk, including higher prevalence of high-load HCMV infections, which in some patient groups are linked with pathology [8,9,15]. Nevertheless, few investigations address their vaccine responses. For example, one study shows after vaccination, mHIV-EU infants presented lower anti-tetanus antibody titres and increased non-responders to hepatitis B vaccine, than unexposed infants, but no difference in diphtheria vaccine responses [17]. Another investigation showed little difference in mHIV-EU infants antibody responses after routine vaccination against Haemophilus Influenzae type b, pneumococcus, tetanus toxoid and hepatitis B surface antigen [18]. This apparent contradiction and the lack of knowledge on this subject highlights the need for further investigation.

Poliomyelitis is a priority vaccine-preventable disease worldwide. Despite great success in eradication, three countries remain with endemic Poliovirus (Afghanistan, Nigeria and Pakistan), and several more with imported Poliovirus [19,20]. A current major concern is that the oral polio vaccine, OPV, efficacy at inducing protective-antibodies against Poliovirus is suboptimal in developing countries [21]. Furthermore, immunogenicity may be reduced in transition to inactivated Poliovirus vaccine, IPV, in countries trying to maintain polio-free status. Diverse factors have been suggested to contribute to this deficient antibody response, such as enteric pathogens interference with OPV responses, high levels of maternal antibodies, vaccination during rainy season, and immunodeficiency syndromes [20,22]. Nonetheless, the role of HIV-1 in reduced OPV immune-responses in developing countries remains controversial, in contrast to characterised effects of weaker responses to vaccines, such as measles, the combined diphtheria, tetanus and whole-cell pertussis (DTwP) [3,23]. Only limited studies have investigated childhood HIV-1 effects and impaired immune responses to OPV [5,23]. Moreover, vaccine efficacy has not been evaluated in infants with maternal HIV-1 exposure but without HIV-1 infection (mHIV-EU) nor in HIV-exposed infants using anti-retroviral (ART) treatments, under new programs to prevent mother-to-child-transmission (MTCT). These mHIV-EU infants are the majority of children affected by maternal exposure to HIV-1. Where antenatal HIV-1 infection is between 20 and 25% as in Zambia, mHIV-EU infants are a large proportion of all births [24,25].

We have evaluated the neutralising-antibody response to OPV in HIV-infected, mHIV-EU and HIV-unexposed Zambian infants included in a randomised double-blind controlled micronutrient trial. Zambia is a polio-free country, but is in danger of importing the virus from polio-endemic countries. In addition, there is high HIV-1/AIDS prevalence, increasing the importance of ensuring complete OPV coverage and high levels of protective antibodies. The main aim of the trial was to investigate the efficacy of micronutrient fortification to prevent stunting of infants, 6 to 18 months age [26]. Here, as a planned secondary outcome we analysed neutralising antibody responses to OPV. In addition, we have investigated how maternal HIV-exposure, infant HIV-1, HCMV infection and breast-feeding influence the OPV neutralising-antibody response, given the interactions described above. We hypothesised that infant HCMV and HIV-1 co-infection, and possibly HCMV infection in mHIV-EU children, could have synergistic negative effects on OPV protective-antibody responses.

2. Methods

2.1. Study design and population

The Chilenje Infant Growth, Nutrition and Infection Study (CIGNIS) was a randomised double-blind controlled trial comparing two micronutrient food-fortifications, 'rich' or 'basal' (Registered clinical trial ISRCTN37460449) [26]. The trial was conducted in Chilenje, Lusaka, Zambia, from October 2005 to July 2009. Children were randomised to one of the two fortified-porridges at age 6 months, and were followed for 12 months. A total of 811 infants were enrolled in the study with 743 correctly randomised [26].

2.2. Poliovirus vaccination and assay for Poliovirus antibody-neutralisation titres

The OPV schedule followed national campaigns with inoculations at birth, 6, 10 and 14 weeks age. Infants missing the first dose were given a booster at 9 months age. All study children had a further boost at 12 months. Poliovirus type 2 neutralising-antibody titres were determined at 18 months as described [27] with modifications. Serial two-fold dilutions of sera, inactivated at 56 °C for 30 min, were incubated with 200 $TCID_{50}$ of Poliovirus type 2 challenge virus (MEF, vaccine reference strain of Poliovirus 2) in 96-microwell plates at 36 °C for 1 h. After the incubation, 100 μl of L20B cell suspension was added to each well, and incubated at 36 °C for minimum 5 days. Presence or absence of cytopathic effect (CPE) was recorded and antibody titres calculated as the highest dilution of serum protecting 50% of the cultures against challenge virus. Positive correlates of protection were neutralisation-antibody titres ≥8 (1:8 dilution, $\log_2 3$) cut-offs [27].

2.3. HIV-1 testing

Maternal HIV-1 status was obtained from government health service antenatal HIV-1 antibody testing records. HIV-1 positive mothers had received short course intrapartum Nevirapine, as local standard of care to reduce childhood transmission, MTCT [28]. Seventy women had unknown antenatal HIV status. All children at 18 months were tested for HIV-antibody using a serial testing algorithm. Samples were tested using Determine HIV 1/2 (Inverness Medical, Japan) with any negative results recorded, positive results were re-tested using Unigold HIV 1/2 (Trinity Biotech plc, Ireland), with confirmed-positive results recorded. If negative, i.e. discordant, a third test SD-Bioline HIV 1/2 (Standard Diagnostics, Korea) was used as final result.

2.4. HCMV antibody and DNA testing

The ETI-CYTOK-G PLUS ELISA Kit (DiaSorin) was used for detection of anti-HCMV IgG in human serum. Standard curves were plotted with control serum standards; sample IgG titres above a cut-off 0.4 IU/ml were considered positive as shown [29]. Qualitative and quantitative polymerase chain reactions (PCR) were performed using HCMV glycoprotein gB (UL55) specific-primers on sera DNA extracted with QIAamp DNA Blood Mini Kit (Qiagen) as described. Possible variability in DNA yields by hemolysis in sera collection were analysed by quantitative real time assays for human GAPDH gene, this showed no evidence for effects of hemolysis. Furthermore specificity and reproducibility were checked by

Fig. 1. Conceptual framework guiding the analyses which shows the main pathways maternal HIV status can influence antibody response to OPV. The overall effect (pathway $a+b+c$), was explored including HIV-positive infants in the analysis, while the effect of maternal HIV infection not dependent on infant HIV status (pathway $b+c$), was obtained restricting the analysis to HIV-uninfected infants only. In HIV-uninfected infants, we then explored the effect of maternal HIV infection on Poliovirus neutralising-antibodies not dependent on breast-feeding duration adjusting by breast-feeding duration in the multivariable analysis (pathway c).

reconstructions of virus genes in sera establishing standard curves for the real time analyes [8,29].

2.5. Statistical analyses

Data were double-entered and verified in Access. Range and logical checks were run in Access and in Stata (Version 12, Statacorp, College Station, TX, USA). OPV neutralising-antibody titre was $\log_2$ transformed and compared between diet and HIV exposure groups using the unpaired *t*-test. In addition, the proportion of infants with seroprotective levels of OPV specific neutralising-antibody ($\log_2$ titre < 3 = un-protective and $\geq$3 protected) was compared between groups using chi-squared test or Fisher's exact test when smallest expected-value was <5. Socioeconomic status (SES) was measured using an asset index, created by combining data on possessions and housing characteristics using principal component analysis [26]. To guide analyses, a conceptual framework was developed to identify the main pathways by which maternal-HIV status can influence OPV neutralising-antibody response (Fig. 1). Multivariable analysis was conducted using linear regression and logistic regression when the outcome was $\log_2$ OPV neutralising-antibody titre and seroprotective levels of OPV antibody, respectively.

2.6. Multiple imputation analyses

Among the 811 infants enrolled in the trial, 597 (73%) infants at 18 months had available OPV neutralising-antibody results. Missing data were mainly due to loss to follow up at 18 months (23%) and inadequate infant serum sample volume (an additional 5%) (Fig. 2). HCMV IgGs and HMCV DNA were measured respectively in available subsample of 460 (57%) and 428 (53%) infants at 18 months age. Only 444 (54%) infants had both HCMV IgGs and OPV antibodies response available, while 414 (51%) had HCMV DNA and OPV response available. Multiple imputations with chained equations methods were used to account for any missing data. Known covariates including those associated with loss to follow up (maternal age and HIV status) and Poliovirus-antibody response (child HIV-status, anaemia, stunting, mother's education, socioeconomic status, breast-feeding duration) (see Table 1) were used to impute 10 data-sets. All continuous variables used in the imputation were normally distributed; socioeconomic status, mother's education and breast-feeding duration were treated as ordered categorical variables. In addition to the main analyses of observed data, complete analyses were conducted with all imputed data.

2.7. Ethics statement

All mothers gave informed, written consent for the study. Ethical approval for the trial was given by the Ethics committees of University Teaching Hospital, Lusaka, Zambia and London School of Hygiene and Tropical Medicine.

3. Results

3.1. Micronutrient-fortification and OPV antibody response

A total of 743 infants were recruited and randomised (Fig. 2); of those 373 (50.2%) were in the 'basal' and 370 (49.8%) in the 'rich' micronutrient-fortified arm. Trial arms were substantially balanced at baseline (supplemental Table 1), with no difference in loss to follow-up by 18 months ($P = 0.88$). The percentage of infants of HIV-uninfected women was higher among those lost to follow-up (77%) than among those who completed the study (68%), but these groups were otherwise similar [26].

Overall and in sub-groups defined by maternal and child HIV-serostatus, there was no evidence ($P > 0.05$) of an effect of rich-fortified, compared with basal-fortified arm, on OPV neutralising-antibody titres (Fig. 3) or in the proportion of infants with un-protective levels of antibodies ($\log_2$ titre < 3) after OPV (2% in basal and 3% in rich micronutrient-fortified arm, $P = 0.74$). Results were confirmed after multiple-imputations analysis (data not shown).

3.2. Breast-feeding and other characteristics associated with OPV antibody response

There was strong evidence ($P < 0.01$) of increasing Poliovirus neutralising-antibody titres with increasing breast-feeding duration. Smaller effects were seen in titres of Poliovirus-antibodies, which were slightly lower in infants who were stunted (length for age < −2 Z) ($P < 0.01$) or anaemic (Hb < 105 g/L) ($P = 0.01$) at 18 months. There was also some evidence for lower OPV antibody-titres in low birth weight infants ($P = 0.06$). Infants of more educated mothers had better antibody responses to OPV, with evidence of linear trend across maternal education categories ($P = 0.02$). Socioeconomic status of the mother, sex, HCMV seropositivity and HCMV DNA positivity at 18 months of the infants were not associated with titre of Poliovirus antibodies and proportion of infants with un-protective antibodies levels, respectively. The above results were confirmed using imputed datasets (data not shown).

3.3. Child HIV infection

Among the 597 infants with available OPV antibody response, 17 (2.8%) tested HIV positive (Fig. 2, Table 1). In univariable analysis, titres of Poliovirus antibodies were significantly lower ($P < 0.01$) in HIV-seropositive compared with HIV-seronegative infants. The proportion of infants with un-protective levels of antibodies ($\log_2$ titre < 3) was also higher (35%) in HIV-seropositive, than in HIV-seronegative infants (1.6%, $P < 0.01$) (Table 1). Almost all HIV positive infants were breastfed (Table 1S), highlighting

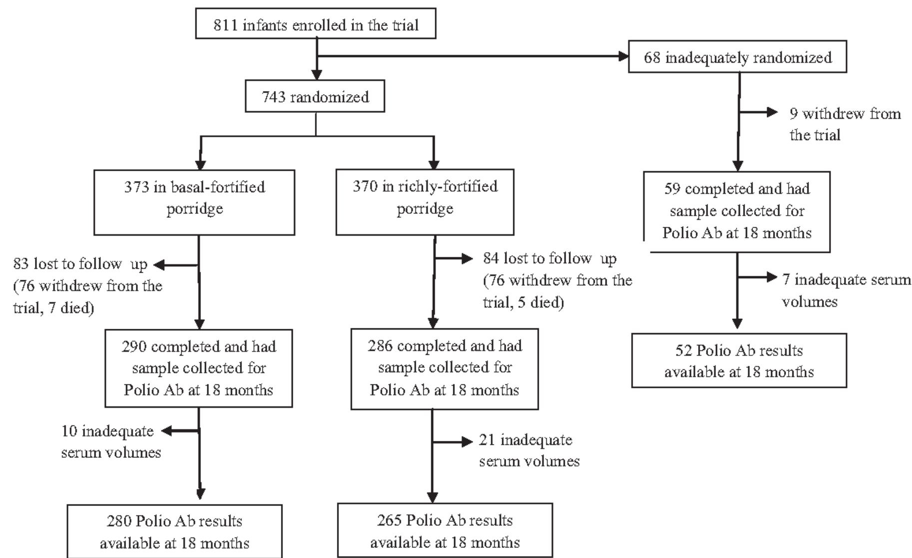


Fig. 2. Flow diagram displaying infants recruited in CIGNIS trial, those lost to follow up and those tested at 18 months for OPV neutralising-antibody response by treatment arm. 743/811 were correctly randomised [26], and those with OPV titres are included in the trial analyses. The 68 non-randomised children are also included where the trial was not the priority focus; 52 of these had available poliovirus-antibody results. Therefore, to assess the effect of child and maternal HIV infection on OPV antibody response, these were included in the analysis (Fig. 1).

the risk of infection through this route. Women who were HIV-positive showed some awareness of this and had modified their own breast-feeding patterns compared to HIV-uninfected women; 25% of HIV-infected women never breastfed whereas almost all HIV-uninfected women breastfed [30]. These differences were further evaluated in effects of maternal HIV exposure on OPV titres.

3.4. Maternal HIV exposure

Among the 547 mothers with known HIV-serostatus, 150 (27.4%) were HIV-positive (Table 1, Fig. 2). The overall effect of maternal HIV status on antibody response to OPV (pathway $a+b+c$ in Fig. 1, included HIV-positive infants) showed that maternal HIV-serostatus was significantly associated with both lower titres of Poliovirus-antibodies ($P<0.01$) and a higher proportion of infants with un-protective Poliovirus-antibody levels ($\log_2$ titre < 3) after OPV ($P<0.01$) (Table 1).

The effect of maternal HIV infection on Poliovirus-antibody titres not dependent on infant HIV status (pathway $b+c$ in Fig. 1), was then analysed including only HIV-uninfected infants (Table 2). Maternal-HIV was still associated with reduced

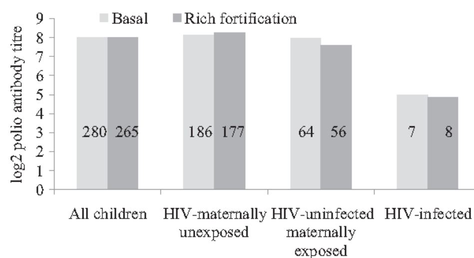


Fig. 3. Log_2 antibody titre by micronutrient-fortification treatment arm, stratified by HIV-infection and maternal HIV-exposure in uninfected infants. Numbers on the bars are sample sizes in each group.

Poliovirus-antibody titres, even after adjusting for maternal education and socioeconomic status (adjusted difference −0.56, 95%CI [−0.98; −0.15], $P<0.01$). We then explored the effect of maternal HIV infection on Poliovirus-antibodies not dependent on breast-feeding duration (pathway c in Fig. 1). Duration of breast-feeding was significantly shorter in HIV-positive mothers compared to HIV-negative (median 6 months vs 15 months, $P<0.01$). Longer breast-feeding duration was associated with increased antibody titre to OPV, independent of socioeconomic status, maternal education and trial arm (Table 2). After adjusting for breast-feeding duration, the effect of maternal HIV exposure on Poliovirus antibody titres was removed (adjusted difference −0.03, 95%CI [−0.59; 0.64], $P=0.93$) (Table 2). These results were confirmed after multiple imputations using imputed data-sets. The imputed regression coefficient for a linear trend between breast-feeding and Poliovirus antibody titre after adjusting for trial arm, maternal HIV-status, education and socioeconomic status was 0.42 [95%CI 0.12; 0.72], $P<0.01$. The association with maternal HIV was again removed after further adjustment for breast-feeding duration: imputed regression coefficient of 0.29 [95%CI −0.50; 1.09], $P=0.47$).

Regarding the effect of maternal-HIV status on proportion of HIV-uninfected infants with un-protective levels of antibodies ($\log_2$ titre < 3) (pathway $b+c$ in Fig. 3), after adjusting for maternal education and socioeconomic status, mHIV-EU infants had slightly, but not significantly, increased odds of having an unprotective level of antibodies after OPV (OR 2.68, 95%CI [0.65; 11.10], $P=0.17$). This also decreased to 1.04, 95%CI [0.14; 7.61], $P=0.97$, after breast-feeding duration adjustment.

Therefore, the overall analyses show breast-feeding duration was able to explain most of the observed differences in antibody titres between mHIV-EU and unexposed-uninfected infants.

3.5. HCMV antibody and DNA positivity

Among the 597 infants with available OPV antibody response, 414 also had HCMV serum DNA results, and 414 HCMV antibody

Table 1
Univariable association OPV antibody response and characteristics of the infants and their mothers.

	n (%) *N* = 597[a]	Mean poliovirus antibody titre (SD)	Difference (95% CI)	*P*[b]	Vaccine Failure (%)[c]	*P*[d]
Infant's characteristics						
Sex						
Male	281 (47.1)	7.91 (2.21)			7 (2.5)	
Female	316 (52.9)	8.17 (2.16)	0.26 (0.60; −0.09)	0.17	8 (2.5)	0.99
Birth weight[e]						
≥2.5 kg	508 (85.7)	8.13 (2.13)			12 (2.4)	
<2.5 kg	85 (14.3)	7.64 (2.41)	−0.49 (−0.99; 0.01)	0.06	3 (3.5)	0.52
Stunting at 18 months[f]						
No	476 (79.9)	8.18 (2.17)			12 (2.5)	
Yes	120 (20.1)	7.51 (2.23)	−0.67 (−1.10; −0.23)	<0.01	3 (2.5)	0.99
Anaemia at the assessment[g]						
No	409 (69.2)	8.20 (2.13)			7 (1.7)	
Yes	182 (30.8)	7.70 (2.29)	−0.50 (−0.12; −0.88)	0.01	8 (4.4)	0.05
HIV serostatus						
Negative	580 (97.2)	8.15 (2.09)			9 (1.6)	
Positive	17 (2.8)	4.71 (2.97)	−3.44 (−2.41; −4.46)	<0.01	6 (35.3)	<0.01
HCMV serostatus at 18 months[h]						
Negative	75 (16.9)	7.87 (2.29)			3 (4.0)	
Positive	369 (83.1)	8.13 (2.09)	0.26 (0.79; −0.27)	0.33	5 (1.4)	0.14
HCMV DNA at 18 months[i]						
Negative	266 (64.3)	8.05 (2.21)			5 (1.9)	
Positive	148 (35.7)	7.87 (2.29)	−0.04 (0.40; −0.49)	0.85	5 (3.4)	0.35
Breast-feeding duration						
Never	42 (7.0)	6.86 (2.28)			1 (2.4)	
<6 months	56 (9.4)	7.52 (2.32)	0.66 (−0.27; 1.59)		3 (5.4)	
6–11 months	70 (11.7)	7.77 (2.54)	0.91 (−0.03; 1.86)		7 (9.9)	
12–17 months	242 (40.6)	8.27 (2.12)	1.41 (−0.71; 2.12)		4 (1.7)	
≥18 months	187 (31.3)	8.30 (2.12)	1.44 (−0.76; 2.13)	<0.01[j]	0 (0.0)	0.23[k]
Mother's characteristics						
HIV status[l]						
Negative	397 (72.6)	8.27 (2.12)			4 (1.0)	
Positive	150 (27.4)	7.31 (2.36)	−0.96 (−0.55; −1.37)	<0.01	10 (6.7)	<0.01
Education						
None/primary	187 (31.3)	7.77 (2.11)			5 (2.7)	
Secondary	236 (39.5)	8.09 (2.17)	0.32 (−0.10; 0.74)		6 (2.5)	
College/university	174 (29.2)	8.29 (2.29)	0.52 (0.08; 0.98)	0.02[j]	4 (2.3)	0.81[j]
Socioeconomic status						
Low	191 (32.0)	8.02 (2.11)			4 (2.1)	
Middle	234 (39.2)	7.91 (2.28)	−0.11 (−0.52; 0.32)		7 (3.0)	
High	172 (28.8)	8.27 (2.16)	0.25 (−0.20; 0.71)	0.25	4 (2.3)	0.82

[a] Including 52 infants with available Poliovirus antibodies results among those not appropriately randomised.
[b] *P*-value obtained by linear regression adjusting for treatment arm.
[c] Number (%) infants with OPV antibody $\log_2$ titre < 3.
[d] *P*-value obtained by logistic regression adjusting for treatment arm.
[e] Information was missing for 4 infants.
[f] Length for age < −2 *Z*; not available for one infant.
[g] Haemoglobin <105 g/L; information was missing for 6 infants.
[h] Not available for 153 infants.
[i] Not available for 183 infants.
[j] Test for trend. Test for departure from linear trend *P*-value > 0.05.
[k] Test for trend *P*-value. Test for departure from linear trend *P*-value = 0.02.
[l] Reported or from clinical records; for 50 mothers HIV status was unknown.

results at 18 months age. Of these, 148 (35.7%) had HCMV viraemia (positive HCMV DNA) indicating active infection, while 369 (83.1%) had antibodies against HCMV documenting a previous exposure to the virus. In our study, we found no evidence of an association between antibody response to OPV and both HCMV DNA positivity and serostatus (Table 1). However, in the 5 HIV-infected infants with HCMV viraemia at 18 months there was a suggestion of lower OPV-antibody titres compared to the 8 infants without HCMV viraemia (Table 3). There was no OPV-antibody titre difference between mHIV-EU or HIV-unexposed infants who were HMCV DNA-positive or negative. HCMV seropositivity at 18 months had no effect on OPV titres in HIV-positive (*P* = 0.49) and mHIV-EU infants (*P* = 0.98) (Table 3). While in HIV-unexposed infants, there was a suggestion of increased OPV antibody titre in HCMV seropositive infants (*P* = 0.11).

4. Discussion

Our study has documented that infant HIV-1 infection is associated with an impaired neutralising antibody response to OPV. Notably, over 35% of the HIV-positive children did not have protective-antibody levels against Poliovirus infection at 18 months of age despite receiving an extra vaccine-booster at 12 months age. In childhood HIV-1 primary infections, the loss of CD4$^+$ T-lymphocytes linked to the peak of viraemia, and subsequent erosion of the immune system, may affect induction and maintainance of immunological memory to OPV [3–5,23]. Furthermore, HIV-1-infected infants are highly susceptible to infections that may either interfere with the vaccine antibody-response, or may alternatively exacerbate HIV-1 effects, as our results show an indication of this from HCMV co-infection.

Table 2
Multivariable association mean $\log_2$ antibody titre after OPV and maternal HIV and breast-feeding in HIV-uninfected infants.[a]

	n (%) N = 530[a]	Difference (95% CI)	P[b]	Adjusted difference (breast-feeding excluded) (95% CI)	P[b]	Adjusted difference (breast-feeding included) (95% CI)	P[b]
Mother HIV status[c]							
Negative	397 (74.9)						
Positive	133 (25.1)	−0.62 (−1.04; −0.21)	<0.01	−0.56 (−0.98; −0.15)	<0.01	−0.03 (−0.59; 0.64)	0.93
Maternal education							
None/primary	162 (30.6)	0 (baseline)		0 (baseline)		0 (baseline)	
Secondary	208 (39.2)	0.41 (−0.02; 0.84)		0.36 (−0.09; 0.80)		0.42 (−0.03; 0.87)	
College/university	160 (30.2)	0.61 (0.15; 1.07)	0.01[d]	0.51 (0.00; 1.03)	0.05[d]	0.58 (0.06; 1.09)	0.03[d]
Socioeconomic status							
Low	171 (32.3)	0 (baseline)		0 (baseline)		0 (baseline)	
Middle	205 (38.7)	0.08 (−0.35; 0.51)		−0.10 (−0.55; 0.34)		−0.03 (−0.48; 0.41)	
High	154 (29.0)	0.32 (−0.15; 0.78)	0.37	0.03 (−0.48; 0.54)	0.91[d]	0.14 (−0.37; 0.65)	0.60[d]
Breast feeding							
Never	38 (7.2)	0 (baseline)				0 (baseline)	
<6 months	51 (9.6)	0.58 (−0.30; 1.46)		–		0.67 (−0.22; 1.56)	
6–11 months	59 (11.1)	1.01 (0.15; 1.86)		–		1.03 (0.13; 1.92)	
12–17 months	217 (41.0)	1.20 (0.47; 1.92)		–		1.23 (0.30; 2.15)	
≥18 months	165 (31.1)	1.28 (0.54; 2.02)	<0.01[d]	–		1.39 (0.46; 2.33)	<0.01[d]

[a] Model based on 530 HIV-negative infants with available $\log_2$ antibody titre.
[b] P-value obtained by linear regression adjusting also for treatment arm.
[c] Reported or from clinical records.
[d] Test for trend. Test for departure from linear trend P-value > 0.05.

Antibody response to OPV at 18 months was not influenced by the provision from 6 to 18 months of micronutrient-fortified infant-food, in the overall study population or any maternal/child HIV sub-groups. In our trial, rich compared to basal micronutrient-fortification was able to improve iron status, reduced anaemia, and increased serum selenium, but did not reduce stunting and was associated with increased occurrence of lower respiratory tract infections [26,31].

Importantly, this study shows that maternal HIV-1 exposure was associated with reduced neutralising-antibody response to OPV also in HIV-uninfected infants. Breast-feeding duration was independently associated with these reduced titres. This is important since many HIV-infected women stop breast-feeding early; 25% HIV-infected mothers, compared to 0.5% HIV-uninfected mothers, never breastfed with medians of 6 vs 15 months breast-feeding, respectively. With adjustment for breast-feeding duration, most differences in OPV titre were removed, suggesting that in our study population, lower breast-feeding duration by HIV-positive mothers could explain reduced OPV titres in mHIV-EU infants.

Breast-feeding was previously associated with better humoral response to OPV and other vaccines, such as diphtheria, tetanus toxoids and Haemophilus influenzae type b vaccine, although HIV-effects were not analysed [32,33]. Despite evidence that formula-fed babies compared to breastfed babies had increased responses to poliovirus type 3 in Brazil [34], the effects of maternal HIV-status were not assessed and there were no differences with type 2 which were analysed here. Reduced OPV titres in mHIV-EU infants could be due to reduced antibody transfer via breast-feeding

Table 3
Effect of HCMV DNA or antibody positivity at 18 months on OPV neutralising-antibody response stratified by infant HIV and maternal HIV-exposure.

	$\log_2$ poliovirus antibody titre					
	N (%)	Mean (SD)[a]	Difference (95% CI)	P value	Vaccine failures[b] (%)	P value
HIV-positive infants	N = 13[c]					
HCMV DNA positive	5 (38.5)	3.20 (1.30)			2/5 (25.0)	
HCMV DNA negative	8 (61.5)	5.75 (3.41)	−2.55 (−6.10; 1.01)	0.14	2/8 (40.0)	1.00
HIV-mEU infants	N = 84[c]					
HCMV DNA positive	28 (33.3)	7.79 (2.33)			1/28 (3.6)	
HCMV DNA negative	56 (66.7)	7.54 (2.05)	0.25 (−0.74; 1.24)	0.62	2/56 (3.6)	1.00
HIV-maternally unexposed	N = 284[c]					
HCMV DNA positive	103 (36.3)	8.29 (2.14)			2/103 (1.9)	
HCMV DNA negative	181 (63.7)	8.25 (2.15)	0.04 (−0.48; 0.55)	0.89	1/181 (0.6)	0.30
HIV-positive infants	N = 9[d]					
HCMV seropositive	7 (77.8)	4.43 (2.37)			2/7 (28.5)	
HCMV seronegative	2 (22.2)	6.00 (4.24)	−1.57 (-6.72; 3.58)	0.49	0/2 (0.0)	1.00
HIV-mEU infants	N = 94[d]					
HCMV seropositive	76 (80.9)	7.79 (2.00)			2/76 (2.6)	
HCMV seronegative	18 (19.1)	7.78 (2.13)	0.01 (−1.04; 1.06)	0.98	1/18 (5.6)	0.47
HIV-maternally unexposed	N = 302[d]					
HCMV seropositive	255 (84.4)	8.31 (2.07)			1/255 (0.4)	
HCMV seronegative	47 (15.6)	7.77 (2.36)	0.53 (−0.12; 1.20)	0.11	2/47 (4.3)	0.06

[a] Mean of $\log_2$ polio neutralising antibody titres and the standard deviation (SD) is in brackets.
[b] Titres below correlate of protection, $\log_2 < 3$.
[c] Polio Ab titre and HCMV DNA were only available in 13 HIV-positive infants, 84 HIV-mEU infants and 284 HIV-maternally unexposed infants. For 33 HIV-uninfected infants with available Polio Ab titre and HCMV DNA, maternal HIV status was unknown, therefore they are excluded from the table.
[d] Polio Ab titre and HCMV serostatus were only available in 9 HIV-positive infants, 94 HIV-mEU infants and 302 HIV-maternally unexposed infants. For 39 HIV-uninfected infants with available Polio Ab titre and HCMV serostatus, maternal HIV status was unknown, therefore they are excluded from the table.

in HIV-1 positive mothers, from reduced duration or reduced antibody production. Breast-feeding also promotes immune system development [35]. Alternatively, the OPV titres may be affected from high exposure of mHIV-EU infants to infectious agents that can be transmitted via breast milk, or via saliva during close contacts between mother and child, both routes for HCMV.

HCMV was further investigated as a factor affecting OPV titres in HIV-infected or exposed infants. HCMV is a main viral opportunistic infection in HIV-infected children, can account for pathology leading to morbidity/mortality, and is a marker for AIDS progression [9]. HCMV has also been linked with immune alterations, which may affect the ability to respond to new infections [10–12]. In our study, there appeared negative synergistic effects between infant HIV infection and HCMV active infection but the analyses were limited by low numbers. In this cohort, HCMV infection also associated with reduced growth [29], but with no evidence of a synergistic effect between maternal HIV-exposure and HCMV previous or current infection on OPV antibody response. However, there was a suggestion of higher OPV titres in maternally HIV-unexposed children who had previous HCMV infection, suggesting a potential immunostimulatory effect, possibly a factor in breastmilk; this was not seen amongst maternally HIV-exposed children.

The impaired immune response to OPV of HIV-1-positive and mHIV-EU infants has important implications not only in the control of poliomyelitis from wild-type Poliovirus, but also in outbreaks caused by vaccine-derived Polioviruses (VDPVs). OPV is a live-attenuated vaccine that can cause infection in immunocompromised individuals. Replication in the host may lead to revertants, recombinant and/or mutant VDPVs with increased pathogenicity, which together with the continuous shedding of virus in infected individuals provide sources for new outbreaks [36]. Where there is reduced Poliovirus immunisation the risk of wild-type Poliovirus and VDPV outbreaks increases. Thus, in regions with high HIV-1/AIDS prevalence, such as Sub-Saharan Africa, it is of great importance to ensure good Poliovirus vaccine coverage and adequate immune responses. For example, increasing doses of Poliovirus vaccine (OPV or IPV) in those children with unprotective neutralising-antibody titres, could improve their vaccine immune response [5]. Although weaker immune responses to OPV were observed, robust titres above the correlate of protection cut-off were in the majority of the cohort (97.5%, 582/597), with unprotective titres mainly in the HIV-infected infants (35%, 6/17) compared to mHIV-EU (4%, 5/133) and the unexposed (1%, 4/397).

The limitations to the interpretation of the study, also point towards an underestimate of the effects of maternal HIV-1 exposure on efficacy of OPV. Although we did not follow immunisation adherence, all study infants were given an additional OPV dose at 12 months age, therefore maternal HIV-exposure effects on OPV may be more pronounced on standard regimens; while in Zambia polio vaccine coverage is high (96%, WHO survey 2009). Responses were only measured to serotype 2, which are characteristically strong; serotype 3 responses, usually weaker than to serotype 2 after OPV [27], could be more affected by HIV-exposure. Finally, only the attenuated-live OPV was used; the killed vaccine, IPV, in use for some eradication programmes for safety, may be more sensitive to maternal HIV-1 exposure. The specific effects of maternal HIV-exposure and HIV-1 infection on different vaccines and the underlying mechanisms, such as negative effects of early HCMV infection with HIV-1, or immune-benefit of breast-feeding, need further investigation. Mechanisms may include effects on antibody, such as IgA, as well as benefit from innate regulation through NK cell and cytokine components of breastmilk. Although breast-feeding was associated with better OPV responses and has a clear public health message, this has to be in tandem with measures to prevent pathogens such as HIV or HCMV transmitted in breastmilk.

Clinical recommendations from this study are that maternally HIV-exposed in addition to HIV-infected children should be monitored for Poliovirus vaccine efficacy, and additional doses given as required, particularly in the absence of breast-feeding. This may apply generally to vaccines for mHIV-EU infants. Moreover, in countries where highly prevalent and early infant infections with HCMV are seen, such as several African countries in addition to Zambia, the effects of childhood HIV-1 infection may be more severe in reduction of OPV efficacy and need consideration in vaccination programs.

Acknowledgements

This work was supported by the Bill and Melinda Gates Foundation, grant ID 37253. We thank all the mothers and their children who participated in the study, Lusaka District Health staff for support, and all members of the CIGNIS study team: Principal Investigator: Suzanne Filteau, London School of Hygiene and Tropical Medicine (LSHTM); Zambian Lead Investigator: Lackson Kasonka, University Teaching Hospital (UTH), Lusaka; Senior Investigators: Rosalind Gibson, University of Otago, New Zealand; Ursula A. Gompels, LSHTM; Shabbar Jaffar, LSHTM; Emmanuel Kafwembe, Tropical Diseases Research Centre, Ndola; Mwaka Monze, UTH; Moses Sinkala, Catholic Relief Services, Zambia; Andrew Tomkins, Institute of Child Health, University College, London; Rodah Zulu, National Institute of Science and Industrial Research, Zambia; Clinic Coordinator: Molly Chisenga; Clinical Officer: Joshua Siame; Data Manager: Hildah Banda Mabuda; Statisticians: Kathy Baisley, Helen Dale, Natasha Larke, Daniela Manno, Andrea Rehman; Research Fellows: Matthew Bates, Anne Mullen, Kunda Musonda, Marta Sanz-Ramos; Clinic Staff: Hellen Kangwa Bwalya, Margaret Chileshe, Priscilla Kangwa Kowa, Mabvuto Kumwenda, Munalula Likando, Sydney Mambwe, Mutinta Muzyamba, Anne Mwale, Lungowe Nyaywa; Laboratory Staff: Humphrey Bima, Julia Chibumbya, Laura Gosset, Louise Hackett, Abigail Jackson, Mirriam Kapambwe, Mazyanga Liewe, Sydney Mwanza, Ida Ndumba, Eric Njunju; Data Entry: Concillia Kabanga, Natalia Shampwaya; Drivers and Cleaners: John Chobo, Winford Kapumba, Charity Musonda, Philip Soko.

Conflict of interest statement: The authors report no conflict of interest in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.vaccine.2013.02.044.

References

[1] Granich R, Crowley S, Vitoria M, Lo YR, Souteyrand Y, Dye C, et al. Highly active antiretroviral treatment for the prevention of HIV transmission. J Int AIDS Soc 2010;13:1.

[2] Setse RW, Cutts F, Monze M, Ryon JJ, Quinn TC, Griffin DE, et al. HIV-1 infection as a risk factor for incomplete childhood immunization in Zambia. J Trop Pediatr 2006;52(October (5)):324–8.

[3] Bekker V, Scherpbier H, Pajkrt D, Jurriaans S, Zaaijer H, Kuijpers TW. Persistent humoral immune defect in highly active antiretroviral therapy-treated children with HIV-1 infection: loss of specific antibodies against attenuated vaccine strains and natural viral infection. Pediatrics 2006;118(August (2)):e315–22.

[4] Pancharoen C, Ananworanich J, Thisyakorn U. Immunization for persons infected with human immunodeficiency virus. Curr HIV Res 2004;2(October (4)):293–9.

[5] Tejiokem MC, Gouandjika I, Beniguel L, Zanga MC, Tene G, Gody JC, et al. HIV-infected children living in Central Africa have low persistence of antibodies to vaccines used in the expanded program on immunization. PLoS ONE 2007;2(12):e1260.

[6] Abzug MJ, Warshaw M, Rosenblatt HM, Levin MJ, Nachman SA, Pelton SI, et al. Immunogenicity and immunologic memory after hepatitis B virus booster vaccination in HIV-infected children receiving highly active antiretroviral therapy. J Infect Dis 2009;200(September (6)):935–46.

[7] Tejiokem MC, Njamkepo E, Gouandjika I, Rousset D, Beniguel L, Bilong C, et al. Whole-cell pertussis vaccine induces low antibody levels in human immunodeficiency virus-infected children living in sub-Saharan Africa. Clin Vaccine Immunol 2009;16(April (4)):479–83.

[8] Bates M, Monze M, Bima H, Kapambwe M, Kasolo FC, Gompels UA. High human cytomegalovirus loads and diverse linked variable genotypes in both HIV-1 infected and exposed, but uninfected, children in Africa. Virology 2008;382(December (1)):28–36.

[9] Kovacs A, Schluchter M, Easley K, Demmler G, Shearer W, La Russa P, et al. Cytomegalovirus infection and HIV-1 disease progression in infants born to HIV-1-infected women. Pediatric pulmonary and cardiovascular complications of vertically transmitted HIV infection study group. N Engl J Med 1999;341(July (2)):77–84.

[10] Pawelec G, Derhovanessian E, Larbi A, Strindhall J, Wikby A. Cytomegalovirus and human immunosenescence. Rev Med Virol 2009;19(January (1)):47–56.

[11] Ben-Smith A, Gorak-Stolinska P, Floyd S, Weir RE, Lalor MK, Mvula H, et al. Differences between naive and memory T cell phenotype in Malawian and UK adolescents: a role for cytomegalovirus? BMC Infect Dis 2008;8:139.

[12] Miles DJ, van der Sande M, Jeffries D, Kaye S, Ismaili J, Ojuola O, et al. Cytomegalovirus infection in Gambian infants leads to profound CD8 T-cell differentiation. J Virol 2007;81(June (11)):5766–76.

[13] Pass RF, Stagno S, Britt WJ, Alford CA. Specific cell-mediated immunity and the natural history of congenital infection with cytomegalovirus. J Infect Dis 1983;148(December (6)):953–61.

[14] Bunders M, Thorne C, Newell ML. Maternal and infant factors and lymphocyte, CD4 and CD8 cell counts in uninfected children of HIV-1-infected mothers. AIDS 2005;19(July (10)):1071–9.

[15] Clerici M, Saresella M, Colombo F, Fossati S, Sala N, Bricalli D, et al. T-lymphocyte maturation abnormalities in uninfected newborns and children with vertical exposure to HIV. Blood 2000;96(December (12)):3866–71.

[16] Miyamoto M, Pessoa SD, Ono E, Machado DM, Salomao R, Succi RC, et al. Low CD4+ T-cell levels and B-cell apoptosis in vertically HIV-exposed noninfected children and adolescents. J Trop Pediatr 2010;56(December (6)):427–32.

[17] Abramczuk BM, Mazzola TN, Moreno YM, Zorzeto TQ, Quintilio W, Wolf PS, et al. Impaired humoral response to vaccines among HIV-exposed uninfected infants. Clin Vaccine Immunol 2011;18(September (9)):1406–9.

[18] Jones CE, Naidoo S, De Beer C, Esser M, Kampmann B, Hesseling AC. Maternal HIV infection and antibody responses against vaccine-preventable diseases in uninfected infants. JAMA 2011;305(February (6)):576–84.

[19] 2010 [cited; Available from: http://apps.who.int/immunization_monitoring/en/globalsummary/countryprofileselect.cfm

[20] Minor PD. The polio-eradication programme and issues of the end game. J Gen Virol 2012;93(March (Pt 3)):457–74.

[21] Okayasu H, Sutter RW, Czerkinsky C, Ogra PL. Mucosal immunity and poliovirus vaccines: impact on wild poliovirus infection and transmission. Vaccine 2011;29(October (46)):8205–14.

[22] Polio vaccines and polio immunization in the pre-eradication era: WHO position paper. Wkly Epdemiol Rec;85:213–28.

[23] Ryder RW, Oxtoby MJ, Mvula M, Batter V, Baende E, Nsa W, et al. Safety and immunogenicity of bacille Calmette-Guerin, diphtheria-tetanus-pertussis, and oral polio vaccines in newborn children in Zaire infected with human immunodeficiency virus type 1. J Pediatr 1993;122(May (5 Pt 1)):697–702.

[24] Stringer EM, Chintu NT, Levy JW, Sinkala M, Chi BH, Muyanga J, et al. Declining HIV prevalence among young pregnant women in Lusaka, Zambia. Bull World Health Organ 2008;86(September (9)):697–702.

[25] Stringer EM, Ekouevi DK, Coetzee D, Tih PM, Creek TL, Stinson K, et al. Coverage of nevirapine-based services to prevent mother-to-child HIV transmission in 4 African countries. JAMA 2010;304(July (3)):293–302.

[26] Team CS. Micronutrient fortification to improve growth and health of maternally HIV-unexposed and exposed Zambian infants: a randomised controlled trial. PLoS ONE 2010;5(6):e11165.

[27] Herremans MM, Reimerink JH, Ras A, Van Der Avoort HG, Kimman TG, Van Loon AM, et al. Evaluation of a poliovirus-binding inhibition assay as an alternative to the virus neutralization test. Clin Diagn Lab Immunol 1997;4(November (6)):659–64.

[28] Chi BH, Sinkala M, Mbewe F, Cantrell RA, Kruse G, Chintu N, et al. Single-dose tenofovir and emtricitabine for reduction of viral resistance to non-nucleoside reverse transcriptase inhibitor drugs in women given intrapartum nevirapine for perinatal HIV prevention: an open-label randomised trial. Lancet 2007;370(November (9600)):1698–705.

[29] Gompels UA, Larke N, Sanz-Ramos M, Bates M, Musonda K, Manno D, et al. Human cytomegalovirus infant infection adversely affects growth and development in maternally HIV-exposed and unexposed infants in Zambia. Clin Infect Dis 2012;54(February (3)):434–42.

[30] Chisenga M, Siame J, Baisley K, Kasonka L, Filteau S. Determinants of infant feeding choices by Zamian mothers: a mixed quantitative and qualitative study. Mater Child Nutr 2010;7(2):148–59.

[31] Gibson RS, Kafwembe E, Mwanza S, Gosset L, Bailey KB, Mullen A, et al. A micronutrient-fortified food enhances iron and selenium status of Zambian infants but has limited efficacy on zinc. J Nutr 2011;141(May (5)):935–43.

[32] Hahn-Zoric M, Fulconis F, Minoli I, Moro G, Carlsson B, Bottiger M, et al. Antibody responses to parenteral and oral vaccines are impaired by conventional and low protein formulas as compared to breast-feeding. Acta Paediatr Scand 1990;79(December (12)):1137–42.

[33] Pabst HF, Spady DW. Effect of breast-feeding on antibody response to conjugate vaccine. Lancet 1990;336(August (8710)):269–70.

[34] Factors affecting the immunogenicity of oral poliovirus vaccine: a prospective evaluation in Brazil and the Gambia. World Health Organization Collaborative Study Group on Oral Poliovirus Vaccine. J Infect Dis 1995 May;171(5):1097–106.

[35] Palmer AC. Nutritionally mediated programming of the developing immune system. Adv Nutr 2011;2(September (5)):377–95.

[36] Rakoto-Andrianarivelo M, Gumede N, Jegouic S, Balanant J, Andriamamonjy SN, Rabemanantsoa S, et al. Reemergence of recombinant vaccine-derived poliovirus outbreak in Madagascar. J Infect Dis 2008;197(May (10)):1427–35.

APPENDIX 16: MANUSCRIPT 3

Original Article

Increased Cytomegalovirus Secretion and Risks of Infant Infection by Breastfeeding Duration From Maternal Human Immunodeficiency Virus Positive Compared to Negative Mothers in Sub-Saharan Africa

Kunda G. Musonda,[1,3] **Mary Nyonda,**[1] **Suzanne Filteau,**[2] **Lackson Kasonka,**[4] **Mwaka Monze,**[3] **and Ursula A. Gompels**[1]

[1]Faculties of Infectious and Tropical Diseases, and [2]Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, University of London, United Kingdom; [3]Virology, and [4]Obstetrics Unit, University Teaching Hospital, Lusaka, Zambia

Corresponding Author: Ursula A. Gompels, Department of Infectious & Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, Keppel St, London WC1E 7HT, UK. E-mail: ursula.gompels@lshtm.ac.uk.

Received June 30, 2015; accepted March 4, 2016; electronically published April 21, 2016.

Background. Breastfeeding imparts beneficial immune protection and nutrition to infants for healthy growth, but it is also a route for human immunodeficiency virus (HIV) and human cytomegalovirus (HCMV) infection. In previous studies, we showed that HCMV adversely affects infant development in Africa, particularly with maternal HIV exposure. In this study, we analyzed infants risks for acquisition of HCMV infection from breastfeeding and compared HIV-positive and HIV-negative mothers.

Methods. Two cohorts were studied in Zambia. (1) Two hundred sixty-one HIV-infected and HIV-uninfected mothers were compared for HCMV deoxyribonucleic acid (DNA) loads and genotypes (glycoprotein gO) in milk from birth to 4 months postpartum. (2) Maternally HIV-exposed and HIV-unexposed infants were compared for HCMV infection risk factors. The second cohort of 460 infants, from a trial of micronutrient-fortified complementary-food to breastfeeding, were studied between 6 and 18 months of age. Human cytomegalovirus seroprevalence was assayed, and logistic regression was used to calculate risk factors for HCMV infection, including maternal HIV exposure and breastfeeding duration.

Results. Human cytomegalovirus was detected in breast milk from 3 days to 4 months *postpartum*, with significantly raised levels in HIV-positive women and independent of genotype. In infants, HCMV antibody seroprevalence was 83% by 18 months age. Longer breastfeeding duration increased infection risk in maternally HIV-unexposed (odds ratio [OR] = 2.69 for 18 months vs <12 months; 95% confidence interval [CI], 0.84–8.59; $P = .03$) and HIV-exposed infants (OR = 20.37 for >6 months vs never; 95% CI, 3.71–111.70; $P < .001$).

Conclusions. Prolonged breastfeeding, which is common in Africa, increased risk of HCMV infection in infants. Both HIV-positive and HIV-negative women had extended milk HCMV secretion. Women who were HIV-positive secreted higher HCMV levels, and for longer duration, with their children at increased infection risk. Human cytomegalovirus control is required to maintain health benefits of breastfeeding.

Key words. breast milk; HCMV; HIV-exposed infants; infant HIV; maternal HIV.

Human cytomegalovirus (HCMV) is a cause of serious disease in infancy particularly with immunosuppression from human immunodeficiency virus (HIV) [1]. Congenital HCMV, acquired in utero, is the main infectious cause of mental retardation and neurodevelopmental impairment in neonates [1]. Postnatal infection, mainly via breast milk, can cause severe morbidity in some preterm or low birthweight babies [2–10]. Human immunodeficiency virus and HCMV coinfected children have increased neurological and respiratory disease, acquired immune deficiency syndrome (AIDS) progression, and death [11–13]. In sub-Saharan Africa where HIV is endemic, maternal

Journal of the Pediatric Infectious Diseases Society, Vol. 5, No. 2, pp. 138–46, 2016. DOI:10.1093/jpids/piw015
© The Author 2016. Published by Oxford University Press on behalf of the Pediatric Infectious Diseases Society.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

HCMV plasma deoxyribonucleic acid (DNA) was linked to increased mortality in both mother and child in Kenya [14]. Human cytomegalovirus secretion in milk is associated with infant HIV transmission in South Africa and Malawi [15, 16]. In maternally HIV-exposed Zambian children, who themselves remained HIV-negative, we showed that HCMV was associated with lower infant growth and psychomotor development [17]. In Zimbabwe, HCMV milk secretion was related to growth faltering in maternally HIV-exposed children [18]. West African children infected with HCMV already express the HCMV "aged" immune phenotype, present in older Europeans, which may alter immune responses to infections and vaccines [19, 20].

Although maternal HIV exposure has been shown to increase HCMV congenital infection prevalence (from 1% to 4% [21–25]), the effects of maternal HIV on postnatal HCMV transmission via breastfeeding, the predominant route of infection, is unknown [26]. In women who are HIV-positive, a correlation has been made among milk HCMV loads, lower CD4 counts, reduced growth, and infant transmission [18, 27]. However, because comparisons have not been made to mothers who are HIV-negative, it is not known whether maternal HIV causes greater HCMV reactivation, load, or extended secretion in milk. In addition, the effects of breastfeeding duration, which varies greatly in women who are HIV-positive, are unknown. In Europe, breastfeeding of infants up to 3 months of age and raised HCMV viral load in breast milk increased risks of HCMV infection [3, 28]. In Africa, extended breastfeeding into the second year of life is common practice. In order to apply any intervention against HCMV, it is important to determine infection risk factors and their timing in HIV-positive versus HIV-negative women, especially in endemic regions of Africa where comorbidity is increased.

In this study, we examined breast milk directly for HCMV DNA loads and genotypes. In addition, we examined the association of breastfeeding duration with HCMV infant infection risks in Zambia, where maternal HIV exposure is frequent. To our knowledge, there is no previous research comparing the effects of maternal HIV status on secretion and transmission of HCMV in breast milk. In this study, we have compared both HIV-positive and HIV-negative mothers in order to understand the effects of breastfeeding practice and HIV on secretion and transmission of HCMV.

METHODS

Ethical Approval

This study was approved by the ethics committees of both the University of Zambia and the London School of Hygiene and Tropical Medicine, University of London.

Study Population and Protocol

Studies were conducted at Chilenje clinic in Lusaka, Zambia, in 2 cohorts: in the first cohort, studies examined extended HCMV secretion in breast milk; the second cohort focused on the physiological relevance in the longer term. Both cohorts were recruited from a similar region served by the clinic, many mothers participated in both cohorts, and there were similar antiretroviral therapy (ART) treatment protocols. At the time of the studies, local care standards included single-dose of nevirapine for HIV-positive mothers and their newborns. In the second, later cohort, ART was available to mothers with CD4 count <200 cells/μL, and towards the end of the study this was changed to <350 cells/μL, in accordance with revised National HIV treatment guidelines (Ministry of Health, Zambia). Only a few mothers in the study were on ART.

Breastfeeding Cohort. The Breastfeeding and Postpartum Health study investigated *postpartum* health among 387 (198 HIV negative, 189 HIV positive) women, from 2001 to 2003 [29]. Breastfeeding was a recruitment criterion, and all women were breastfeeding exclusively or predominantly. Milk samples were collected on 11 scheduled visits during the first 16 *postpartum* weeks and stored at −80°C. Two hundred sixty-one milk samples (from 118 HIV-positive and 143 HIV-negative mothers), collected at *postpartum* week 16, were available for our study. For a subset of 40 women (20 HIV-negative and 20 HIV-positive), we also analyzed samples collected at 5 earlier time points: day 3, and weeks 2, 4, 9, and 12 *postpartum*. Maternal HIV serostatus was determined antenatally using a serial testing algorithm, per local care standards [28].

Infant Cohort. The Chilenje Infant Growth, Nutrition and Infection Study (CIGNIS) was a randomized double-blind controlled trial of micronutrient-fortified infant foods (ISRCTN37460449; www.controlled-trials.com/mrct) and was conducted from 2005 to 2009. Infants (n = 811) were enrolled at 6 months of age and observed for 12 months [30]. Socio-demographic information was obtained using a questionnaire at recruitment. At recruitment and monthly, mothers were asked whether they were still breastfeeding or when they stopped [31]. Infant venous blood was collected in plain vacutainers, serum was separated, and antibody was assayed for HIV and HCMV at study completion (18 months age). Human immunodeficiency virus serostatus was determined using the local Ministry of Health-approved serial testing algorithm as described previously [17]. The ETI-CYTOK-G PLUS ELISA Kit (DiaSorin) was used to test for HCMV immunoglobulin (Ig)G, with standard curves plotted using control sera provided and then used to interpolate each sample IgG

titer; we considered HCMV IgG positive above a cutoff of 0.4 IU/mL.

Deoxyribonucleic Acid Extraction, Qualitative and Quantitative Polymerase Chain Reaction

Deoxyribonucleic acid was extracted from 200 µL homogenized milk using the QIAamp DNA Mini kit (QIAGEN) and eluted in 50 µL H_2O. Human cytomegalovirus glycoprotein *gB* gene was used for qualitative polymerase chain reaction (PCR) screening and quantification, human *GAPDH* gene was used as internal control, and hypervariable marker HCMV glycoprotein gO (gO) was used for genotyping [17, 32, 33]. Human cytomegalovirus DNA copy numbers were computed by TaqMan real-time assay, performed in triplicate on the Applied Biosystems 7500 Fast Real-Time PCR System (Applied Biosystems Inc.) [32]. Standard curves were generated from 10-fold serial dilutions of plasmid-cloned amplification products (pGEM-T Easy Vector Systems, Promega), normalized with an internationally certified clinical reference standard (National Institute of Biological Standards and Control, UK Medicines and Healthcare products Regulatory Agency, Potters Bar, United Kingdom). Levels detected below the sensitivity of the virus standard were recorded at a value of half the limit of detection. Each quantitative PCR (qPCR) reaction had 10 µL KAPA PROBE FAST Universal qPCR Master Mix (Kapa Biosystems), 1 µL probe (5 mM), 1 µL each of both forward and reverse primer (10 mM), 0.4 µL ROX Low, 7 µL dH_2O, and 5 µL template DNA. Cycling conditions were as follows: 95°C for 10 minutes, then 45 cycles of 95°C for 15 seconds, and 60°C for 1 minute. All PCR assays included both positive and negative controls (reagents and water-only).

Genotyping

Human cytomegalovirus gO (*UL74* gene) genotyping was performed on a subset of 34 milk samples with sufficient remaining volume. Based on translated amino acid sequences, HCMV gO (*UL74* gene) has 8 distinct genotypes: gO1a, gO1b, gO1c, gO2a, gO2b, gO3, gO4, and gO5 [32, 33]. We used PCR-based genotyping with the gOup/gOlw primers, which detected all 8 gO genotypes. Nucleotide sequences were determined using Sanger methods [32, 33] and compiled using ChromasPro software. Multiple alignments used CLUSTAL and phylogenetic analysis via maximum-likelihood methods in MEGA6 [34].

Statistical Analysis

Programs Prism (version 6; GraphPad Software, Inc.) and SPSS (version 20.0; IBM Corp., Armonk, New York) were utilized to analyze milk HCMV DNA data by Student's *t* test and Mann–Whitney *U* test with 2-tailed *P* values and alpha = 0.05. Stata (version 11.1; StataCorp, College Station, Texas) was used in analyses of the infant cohort data, and odds ratios (ORs) and 95% confidence intervals (CIs) were obtained by logistic regression. We assessed associations of breastfeeding duration with HCMV antibody at 18 months, in a multivariable model. Analyses were adjusted for maternal education and socioeconomic status, as main effect modifiers as described [17], and stratified by maternal HIV status. To account for missing data, mainly from limited infant serum sample volumes, we also imputed missing HCMV results using multiple imputations with chained equations. The results imputed to account for missing data were similar to the main analyses (data not shown).

RESULTS

Increased Loads and Duration of Human Cytomegalovirus Milk Excretion in Human Immunodeficiency Virus-Positive Women

Human cytomegalovirus DNA measurements were made in 405 breast milk samples from 261 mothers at week 16 *postpartum* and a subset of 40 mothers at multiple time points. In the longitudinal subset (Figure 1), all mothers screened were HCMV positive by qualitative assay from day 3, and the median HCMV DNA in day 3 milk (colostrum or transitional milk) was not significantly different in HIV-infected compared with HIV-uninfected mothers (3.9 and 4.1 $\log_{10}$ copies/mL, respectively; $P = .90$). Deoxyribonucleic acid lactia then increased in both groups, reaching peak levels at week 4. In the HIV-negative

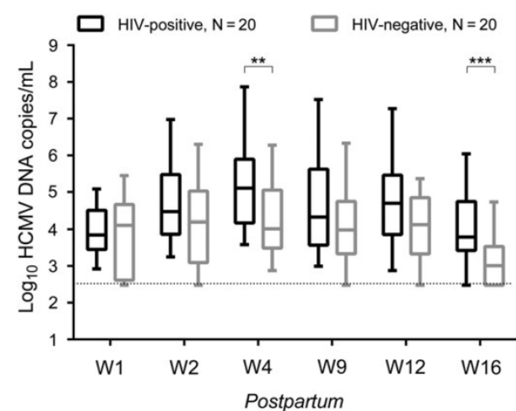


Figure 1. Human cytomegalovirus (HCMV) shedding kinetics in milk of human immunodeficiency virus (HIV)-positive and HIV-negative women. Comparison of milk HCMV deoxyribonucleic acid (DNA) kinetics between HIV-positive (black lines) and negative women (gray lines) over the first 16 *postpartum* weeks (n = 40, 20 in each group). Human cytomegalovirus DNA load, log copies/mL milk, increased from comparable levels in the 2 groups from day 3 (week 1 [W1]) to peak levels by week 4 (W4) postpartum. Sensitivity cut-off is indicated by the lower dotted line. Milk DNA loads from HIV-positive women were raised compared with HIV-negative women from W4. Box plots show the median and interquartile range using a Mann–Whitney *U* test, $^{**}P = .026$ and $^{***}P < .001$ indicate significant differences from W4 and week 16 (W16), respectively.

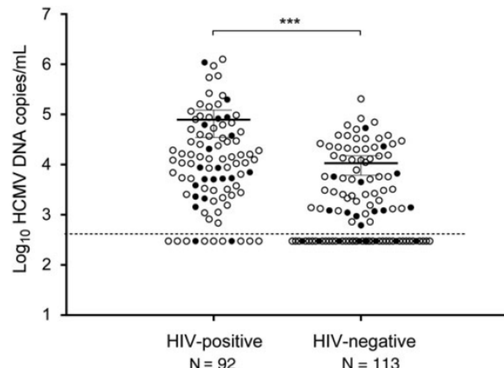


Figure 2. Human cytomegalovirus (HCMV) deoxyribonucleic acid (DNA) loads in week 16 milk samples, stratified by human immunodeficiency virus (HIV) serostatus. Scatter plot showing HCMV DNA levels in all available milk samples, at week 16 (W16) (92 HIV-positive and 113 HIV-negative). Human immunodeficiency virus-positive women had a significantly higher mean milk HCMV DNA load, approaching 1 log higher, compared with their HIV-negative counterparts. Furthermore, a higher proportion of milk samples from the HIV-positive group remained with detectable HCMV at this late time point (88.0% [81 of 92] vs 59.3% [67 of 113]; $P < .001$), with correspondingly decreased proportions below the limit of detection, indicated by the dashed line, in the HIV-positive compared with HIV-negative group (12% [11 of 92] vs 40.7% [46 of 113]). Means were 7.9×10^4 and 1.1×10^4 copies/mL in HIV-positive and HIV-negative women, respectively ($P < .001$, 2 sample, 2 tailed Student's *t* test with 95% confidence interval indicated by error bars), with similar differences in the W16 values for the subset with the complete kinetics from Figure 1 as indicated by black circles, mean 8.2×10^4 and 0.5×10^4 copies/mL for HIV-positive compared with HIV-negative women.

mothers, these loads declined gradually by week 16 to below day 3 levels, whereas in HIV-infected mothers, the DNA loads remained elevated. From week 4 to week 16, the median DNA loads had sustained increases in HIV-infected compared with the HIV-uninfected women ($P < .001$ by week 16), with over a 10-fold difference at peak levels recorded at week 4 ($P = .026$).

At week 16, HCMV was detected by the *gB* screening assay in 83.9% (99 of 118) of milk from HIV-infected women, versus 63.6% (91 of 143) in HIV-negative women ($P < .001$). Human cytomegalovirus DNA loads were quantified by real-time qPCR in 205 samples, all of which contained sufficient sample (with no demographic differences). Of these, there was a significantly higher proportion of HIV-infected women with DNA levels above the assay detection limit (88.0% [81 of 92] of HIV-infected versus 59.3% [67 of 113] among HIV-uninfected; $P < .001$). The mean DNA load was significantly higher in the HIV-infected group (7.9 × 104 copies/mL; 95% CI, 3.5×10^4 to 1.2×10^5) compared with the HIV-uninfected group (1.1×10^4 copies/mL; 95% CI, 0.6×10^4 to 1.5×10^4; $P < .001$) (Figure 2).

Human Cytomegalovirus Glycoprotein O Genotypes in Breast Milk

Human cytomegalovirus gO (*UL74* gene) genotyping was performed in a subset of W4 (peak HCMV DNA loads)

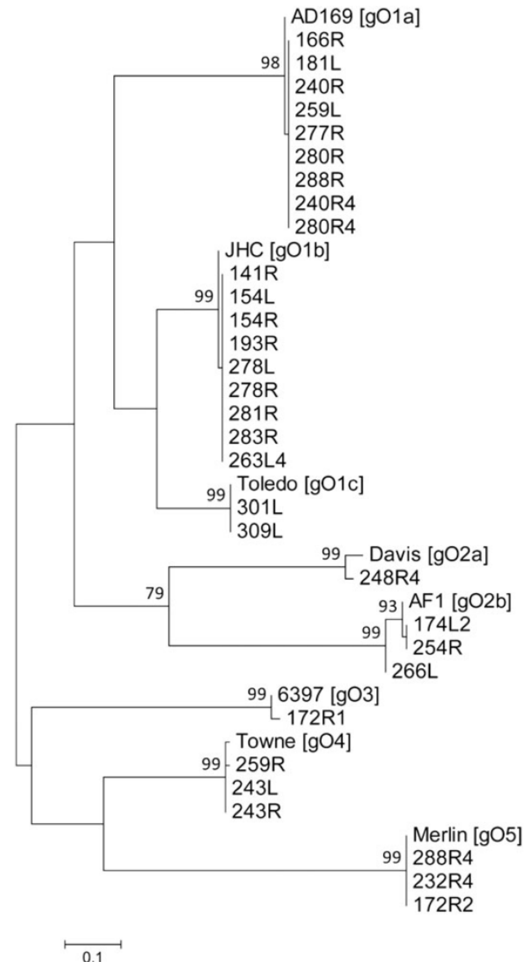


Figure 3. Phylogenetic analyses glycoprotein O (gO) genotypes. Representatives of genotype groups defined here were analyzed in comparison to reference strains, as described previously [32, 33]. Multiple alignments were performed using CLUSTAL in MEGA6 [34], followed by phylogenetic constructions inferred using the Maximum Likelihood method based on the JTT matrix-based model. The analysis involved 39 amino acid sequences and 154 positions in the final dataset. Reference strains for gO genotypes are indicated. Bootstrapping analyses indicate that major nodes are well supported.

and W16 (latest time point) milk samples. This included all samples with sufficient DNA and 7 samples paired at both time points. On the basis of encoded amino acid sequences, all 8 HCMV gO distinct genotypes—gO1a, gO1b, gO1c, gO2a, gO2b, gO3, gO4, and gO5 [32, 33]—were detectable in the milk samples. These were predominantly genotypes 1a, 1b, and 5, as represented by reference strains AD169, TR, and Merlin, respectively (Figure 3); genotype prevalence differences in HIV-positive compared with HIV-negative women did not reach significance (Supplementary Table S1). There was no evidence for higher viral load with any one genotype, although all appeared

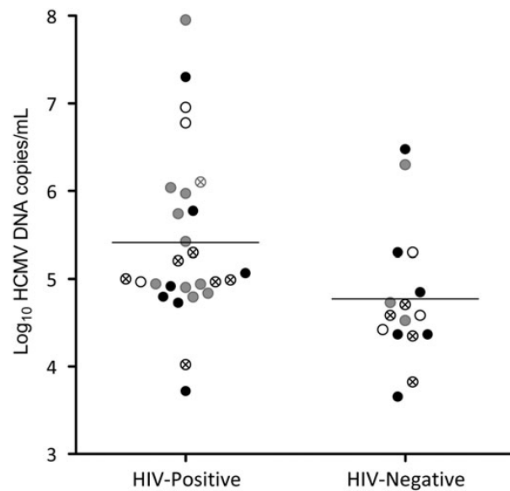


Figure 4. Genotype-independent increases in human cytomegalovirus (HCMV) load in milk from human immunodeficiency virus (HIV)-positive women. Viral loads per genotype were examined for HIV-positive and HIV-negative women. The main 3 glycoprotein O (gO) genotypes (gO1a, gO1b, and gO5) were plotted, and all the remaining genotypes were grouped together. All genotypes appear increased in HIV-positive women compared with HIV-negative women. Genotype gO1a (black circle), gO1b (gray circle), gO5 (white circle), other gO genotypes (crossed circle) are shown. Mean values are as follows: 2.6×10^5 and 5.9×10^4 for HIV-positive and HIV-negative women at both maximal and minimal secreted levels, at 4 and 16 weeks postpartum. Abbreviation: DNA, deoxyribonucleic acid.

raised in milk samples from HIV-positive women compared with HIV-negative women (Figure 4). Of the 7 paired samples, 3 had different genotypes detected at W4 and W16 (HIV-positive and HIV-negative).

Prevalence of Infant Human Cytomegalovirus Antibody at 18 Months of Age and Breastfeeding Duration

The effects of breastfeeding duration on infant HCMV infection were compared between HIV-positive and HIV-negative women in the CIGNIS infant cohort. Overall, 460 of 811 (57%) infant samples were available for HCMV antibody testing at 18 months age. Most infants were seropositive (384 of 460; 83%). As shown previously, we found no effect of micronutrient fortification (trial intervention) on HCMV antibody at 18 months, either overall or by maternal HIV status, but the prevalence of HCMV antibody significantly increased with decreasing maternal socioeconomic conditions or education and increased with longer breastfeeding duration, which were all measures adjusted in analyses of HCMV effects on growth [17]. Breastfeeding duration differed markedly between HIV-infected and HIV-uninfected women in this cohort, as shown previously [31]. Of the HCMV study subgroup analyzed here, only 3 HIV-negative women, compared with 29 HIV-positive women, never breastfed; and only an additional 6 HIV-negative women breastfed for less than 6 months. Therefore, to further analyze this

Table 1. Effects Maternal HIV and Breastfeeding Duration on HCMV Infection

HCMV Infant Infection (Antibody)			
Months Breastfeeding	Antibody N (%)	Adjusted OR[a] (95% CI)	*P* Value
HIV-Negative Mothers			
<12[b]	25/32 (78.1%)	1	
12–17	128/161 (79.5%)	0.94 (0.35–2.53)	
18+	110/119 (92.4%)	2.69 (0.84–8.59)	.03
HIV-Positive Mothers			
Never	13/26 (50.0%)	1	
<6	31/35 (88.6%)	6.83 (1.69–27.6)	
6+	42/44 (95.5%)	20.37 (3.71–111.7)	<.001

Abbreviations: CI, confidence interval; HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; OR, odds ratio.
[a] Adjusted for socioeconomic status and maternal education.
[b] Only 3 HIV-negative mothers never breastfed and only 6 for <6 months.

result, we examined the effect of breastfeeding duration on HCMV antibody stratified by maternal HIV status. Among children of HIV-negative mothers, those who were still breastfeeding at 18 months had nearly 3 times the odds of HCMV antibody as those who had breastfed for <12 months (OR = 2.69; 95% CI, 0.84–8.59; $P = .03$) (Table 1). Even though it was a relatively small group who were uninfected by 18 months age, the children of HIV-positive mothers were now at significantly greater risk of early HCMV infection, as detected by antibody at 18 months, with prolonged breastfeeding (OR for breastfeeding >6 months compared with no breastfeeding = 20.37; 95% CI, 3.71–111.70; $P < .001$) (Table 1).

DISCUSSION

The study shows widespread HCMV infections in Zambian infants, from both HIV-positive and HIV-negative women. Human cytomegalovirus seropositivity was 83% by 18 months age, which is both higher and earlier than in many regions. Although similar to some other regions in Africa, this population has the added effect of endemic HIV. In Africa, the common practice of extended breastfeeding was identified as a risk for HCMV infection, which was increased for maternally HIV-exposed children. Furthermore, HIV-positive women had strikingly higher loads of HCMV secreted in their breast milk, with extended periods of raised levels, compared with HIV-negative women. There was some overlap so other factors may also influence secretion. This is comparable to studies from East Africa that reported high HCMV loads in HIV-infected women, although comparisons to HIV-negative women were not made [27]. Our studies have now compared both HIV-positive and HIV-negative mothers as well as considered breastfeeding duration in this sub-Saharan African setting. Similar to European and South East Asian surveys, our studies show

peak HCMV DNA levels at 4 weeks *postpartum* but differences in shedding duration. More HIV-positive than HIV-negative women had detectable HCMV secretion in breast milk, with initial reactivated levels equal at day 3, then raised from 2 to 16 weeks *postpartum*. Recent studies show HCMV-susceptible $CD14^{+}$ leukocytes increasing in breast milk at this time [35]. Thus, in milk secreted from HIV-positive women, there appears to be decreased immune regulation or increased susceptible cells, possibly from immune activation or inflammation. Studies in this cohort showed increased mastitis in HIV-positive women [29], and we have demonstrated increased secretion of HCMV. We further show that, in addition to increased risk of congenital infection from intrauterine HCMV infection [21-23], infants of HIV-infected women have increased odds of HCMV infant infection from breastfeeding. This can be a confounder for determining congenital infection, because diagnostic tests by saliva in those under 2 weeks of age may detect HCMV in saliva from breast milk. Our studies of birth prevalence of HCMV in Zambia using newborn saliva show 1% (1 of 100) prevalence in normal labor ward (K.G.M. and U.A.G, unpublished), lower than studies in the neonatal intensive care unit where breast milk could also be a source of early infection [25].

Low socioeconomic status and level of education were associated with HCMV seroprevalence in analyses of risk factors, similar to those reported elsewhere [17, 36]. With adjustment for these risk factors, breastfeeding up to 18 months remained significantly associated with HCMV infection. Studies in other continents show that almost all HCMV-positive women excrete HCMV in breast milk from local tissue reactivation, which is distinct from detection in plasma. Even with breast milk secretion in HCMV-positive mothers, the transmission rate varies, and 58%–80% infants were found to be seropositive by 1 year of age [26]. In Europe and Asia, studies have shown breast milk HCMV secretion up to 2–3 months *postpartum*, with higher milk viral loads and prolonged secretion linked to infant transmission [3, 28, 37]. In Zambia, we showed sustained breast milk HCMV secretion for over 4 months. Furthermore, increased breastfeeding over 6 months among HIV-positive women, or over 18 months among HIV-negative women, increased risk of infant HCMV infection. This shows that HCMV is secreted (or possibly more transmissible in breast milk) for longer duration than that reported in Europe or Asia. In HIV-negative women, longer duration of lactation may increase HCMV local reactivation and secretion. In HIV-positive women, secreted levels may be further raised through both HIV and HCMV immune dysregulation and amplified with breastfeeding duration.

Increased transmission during breastfeeding also allows for reinfections with multiple strains and widens the total population exposure to HCMV infection. We previously showed complex mixtures of HCMV strains, a potential factor for severe HCMV disease (currently under assessment) and a marker for burden of infection in this population, from blood or lung samples of HIV-positive infants [32]. We used gO for genotyping, one of the most variable genes in HCMV. Our genotype analyses showed that all 8 gO genotypes were detected in breast milk, demonstrating that genotypes were not constrained in this tissue compartment. There was also evidence for mixed-infections indicative of reinfection, which was confirmed using gN genotyping (data not shown). In HIV-positive women, viral load seemed to be raised independent of the genotypes secreted. The main milk gO genotypes 1a, 1b, and 5 were similar to prevalences we previously described in blood and respiratory compartments of primarily HIV-infected children in the same region and in other tissues from different global sources; differences in minor genotypes did not reach significance, although gO3 was greater in other tissues (Supplementary Table 1 and Supplementary Figure S1) [32, 33]. The analyses were limited by the small proportion of samples available for genotype analyses. These genotype ratios require further study, because the gO trimeric or alternate pentameric complexes with gH/gL glycoproteins affect host transmission and candidate vaccines [38, 39].

In the infant cohort, almost all HIV-negative women breastfed their infants, whereas a quarter of HIV-positive women never breastfed their infants, and overall HIV-positive women had significantly shorter breastfeeding durations than HIV-negative women (median 6 months vs 15 months, $P < .01$). Various reasons, including trying to limit infant HIV, were presented by the mothers [31]. This may have provided some protection from HCMV infection to children of HIV-positive mothers because those who never breastfed were only 50% HCMV positive, whereas breastfeeding at least to 6 months increased the prevalence of HCMV-positive children to 88.6% and over 6 months up to 95.5%. These differences in breastfeeding behavior between HIV-positive and HIV-negative women also meant that comparisons could only be made using different infant breastfeeding duration categories and are a limitation of the study. The 6-fold and 20-fold increase in odds of infection in adjusted analyses with breastfeeding to 6 months and over 6 months age, respectively, is in agreement with the viral loads measured in breast milk. Although limited from 2 different cohort studies, these were from the same residential area and study clinic. A study in Kenya showed that HCMV transmitters had a median of

5.4 compared with 4.5 logs copies/mL in milk for nontransmitters at 2 weeks *postpartum* [27]. Although it is difficult to apply exact thresholds, at similar times *postpartum*, in our Zambian cohort, HIV-positive compared with negative women had a median of 5.1 versus 4.0 logs copies/mL in breast milk. Furthermore, between day 3 and week 16, HIV-positive women had 23 measurements above 5.5 logs copies/mL compared with only 7 for HIV-negative women. This would indicate clinical relevance for the log differences in HCMV milk secretion of increased risk for transmission and at earlier times in the HIV-positive group.

In a separate analysis, we showed that HCMV adversely affected this infant cohort's growth and psychomotor development, particularly in maternally HIV-exposed children [17]. In Kenya, HCMV DNA in maternal plasma is a predictor of mortality in HIV-infected women and their infants [14], and lower CD4 levels were related to lower levels of milk HCMV required for transmission [27]. Although it was not measured, HCMV could be a factor in studies in Zambia where women with advanced HIV/AIDS who breastfed less than 4 months had improved infant survival [42]. In Zimbabwe, HIV-positive women with HCMV co-infection excreted both pathogens in breast milk, and HCMV infection correlated with higher levels of HIV ribonucleic acid [43]. Further analyses show HCMV milk secretion in HIV-positive mothers correlates with infant growth-faltering [18]. In United States, studies assessing earlier ART during pregnancy in HIV-positive women showed reductions in peri- or postnatal HCMV infection [44]. Although this was a non-breastfeeding population, the results support use of interventions to improve maternal health, such as earlier use of ART in HIV-positive women, which is now being applied in Zambia. However, recent studies in African breastfed populations showed that the use of antiretrovirals in HIV-positive mothers did not restrict HCMV transmission to their infants and therefore remained a risk for increasing numbers of HIV-exposed uninfected children [16, 18].

There are several limitations to our studies. The HCMV serostatus of the mothers was not known, although breast milk DNA PCR showed that all mothers were HCMV positive. We did not have CD4 levels, and these can affect milk secretion of CMV, as shown in studies of only HIV-positive mothers in Kenya [27], and may have contributed to varying HCMV levels in the HIV-positive mothers. We did not include day care status as a possible source for infant transmission via saliva [40]; however, in Zambia, infant day care is minimal to nonexistent. Other HCMV secretory routes, particularly saliva and also urinary [41] from siblings, can affect transmission, and 50% of infants became HCMV seropositive from HIV-positive women who did not breastfeed. However, we could not compare this result to HIV-negative mothers because almost all of them breastfed. Furthermore, most of our cohort mothers had children under age 5 who could be secreting HCMV; therefore, in general, background exposure was equal. We did not screen for prenatal HCMV, but this result would be only approximately 1% based on previous studies on newborns. The strengths of this study are the size of the cohorts and the comparison between both HIV-negative and HIV-positive groups, which clearly show raised levels of HCMV milk secretion and duration, with increased risks for infant infection in HIV-positive compared with HIV-negative women.

The World Health Organization recommends 6 months exclusive breastfeeding for ideal infant nutrition and immune protection. In Zambia, Demographic Health Survey data show that 73% of infants are exclusively breastfed for 6 months, and, overall, 98% of children breastfed with a median duration of 20.1 months [45]. Possible interventions against HCMV need to retain breastfeeding benefits. Direct HCMV inactivation in milk using ultrashort heat treatment has been evaluated in Europe, particularly for at-risk, premature, and underweight infants, and was found to be effective while preserving the nutritional and immunological qualities of breast milk [46]. This intervention warrants assessment in the sub-Saharan Africa setting, where it could be useful for concurrent inactivation of HCMV and HIV in breast milk, for at risk groups. Administration of anti-HCMV drugs to mothers to lower milk HCMV load is another potential intervention, but this requires efficacy and safety trials; no anti-HCMV drug to date is licenced for use during lactation. Improved hygiene could lower complementary routes of transmission, including saliva or urine [47]. Furthermore, there are several promising vaccines currently in development [48, 49].

CONCLUSIONS

We conclude that longer breastfeeding duration over 6 to 18 months increases HCMV infant infection. We also showed that HIV-positive compared with HIV-negative women had both raised breast milk secretion and likelihood for infant infection. Interventions to reduce HCMV infection could be considered, particularly in countries with high HCMV and HIV prevalence, because breastfeeding remains critical for infant health.

Acknowledgments

We thank all of the mothers and children who had participated in these studies as well as the Lusaka District Health staff who gave their support. We are also grateful for the contributions of the

Breastfeeding and Postpartum Health Study and the Chilenje Infant Growth, Nutrition and Infection Study team members, particularly Kathy Baisley and Andrea Rehman for statistical advice on breastfeeding duration effects and Molly Chisenga for milk sample collections.

Principal Investigator: Suzanne Filteau, London School of Hygiene and Tropical Medicine (LSHTM); Zambian Lead Investigator: Lackson Kasonka, University Teaching Hospital (UTH), Lusaka; Senior Investigators: Rosalind Gibson, University of Otago, New Zealand; Ursula A. Gompels, LSHTM; Shabbar Jaffar, LSHTM; Emmanuel Kafwembe, Tropical Diseases Research Centre, Ndola; Mwaka Monze, UTH; Moses Sinkala, Catholic Relief Services, Zambia; Andrew Tomkins, Institute of Child Health, University College, London; Rodah Zulu, National Institute of Science and Industrial Research, Zambia; Clinic Coordinator: Molly Chisenga; Clinical Officer: Joshua Siame; Data Manager: Hildah Banda Mabuda; Statisticians: Kathy Baisley, Helen Dale, Natasha Larke, Daniela Manno, Andrea Rehman; Research Fellows: Matthew Bates, Anne Mullen, Kunda Musonda, Marta Sanz-Ramos; Clinic Staff: Hellen Kangwa Bwalya, Margaret Chileshe, Priscilla Kangwa Kowa, Mabvuto Kumwenda, Munalula Likando, Sydney Mambwe, Mutinta Muzyamba, Anne Mwale, Lungowe Nyaywa; Laboratory Staff: Humphrey Bima, Julia Chibumbya, Laura Gosset, Louise Hackett, Abigail Jackson, Mirriam Kapambwe, Mazyanga Liewe, Sydney Mwanza, Ida Ndumba, Eric Njunju; Data Entry: Concillia Kabanga, Natalia Shampwaya; Drivers and Cleaners: John Chobo, Winford Kapumba, Charity Musonda, Philip Soko.

Financial support. This work was funded by the Bill and Melinda Gates Foundation (Grant ID 37253) and the Commonwealth Scholarship Commission (reference number ZMCS-2012-643).

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

Supplementary Data

Supplementary materials are available at the *Journal of The Pediatric Infectious Diseases Society* online (http://jpids.oxfordjournals.org).

References

1. Boppana SB, Fowler KB. Persistence in the population: epidemiology and transmission. In: Arvin A, Campadelli-Fiume G, Mocarski ES Jr, et al. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Chapter 44. Cambridge: Cambridge University Press, 2007.
2. Capretti MG, Lanari M, Lazzarotto T, et al. Very low birth weight infants born to cytomegalovirus-seropositive mothers fed with their mother's milk: a prospective study. J Pediatr 2009; 154:842–8.
3. Jim WT, Shu CH, Chiu NC, et al. High cytomegalovirus load and prolonged virus excretion in breast milk increase risk for viral acquisition by very low birth weight infants. Pediatr Infect Dis J 2009; 28:891–4.
4. Kurath S, Halwachs-Baumann G, Muller W, Resch B. Transmission of cytomegalovirus via breast milk to the premature born infant: a systematic review. Clin Microbiol Infect 2010; 16: 1172–8.
5. Meier J, Lienicke U, Tschirch E, et al. Human cytomegalovirus reactivation during lactation and mother-to-child transmission in preterm infants. J Clin Microbiol 2005; 43:1318–24.
6. Minamishima I, Ueda K, Minematsu T, et al. Role of breast milk in acquisition of cytomegalovirus infection. Microbiol Immunol 1994; 38:549–52.
7. Hamele M, Flanagan R, Loomis CA, et al. Severe morbidity and mortality with breast milk associated cytomegalovirus infection. Pediatr Infect Dis J 2010; 29:84–6.
8. Novakova V, Hamprecht K, Muller AM, et al. Severe postnatal CMV colitis with an extensive colonic stenosis in a 2-month-old male immunocompetent term infant infected via breast milk. J Clin Virol 2014; 59:259–63.
9. Tengsupakul S, Birge ND, Bendel CM, et al. Asymptomatic DNAemia heralds CMV-associated NEC: case report, review, and rationale for preemption. Pediatrics 2013; 132:e1428–34.
10. Lanzieri TM, Dollard SC, Josephson CD, et al. Breast milk-acquired cytomegalovirus infection and disease in VLBW and premature infants. Pediatrics 2013; 131:e1937–45.
11. Kovacs A, Schluchter M, Easley K, et al. Cytomegalovirus infection and HIV-1 disease progression in infants born to HIV-1-infected women. Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted HIV Infection Study Group. N Engl J Med 1999; 341:77–84.
12. Doyle M, Atkins JT, Rivera-Matos IR. Congenital cytomegalovirus infection in infants infected with human immunodeficiency virus type 1. Pediatr Infect Dis J 1996; 15:1102–6.
13. Nigro G, Krzysztofiak A, Gattinara GC, et al. Rapid progression of HIV disease in children with cytomegalovirus DNAemia. AIDS 1996; 10:1127–33.
14. Slyker JA, Lohman-Payne BL, Rowland-Jones SL, et al. The detection of cytomegalovirus DNA in maternal plasma is associated with mortality in HIV-1-infected women and their infants. AIDS 2009; 23:117–24.
15. Viljoen J, Tuaillon E, Nagot N, et al. Cytomegalovirus, and possibly Epstein-Barr virus, shedding in breast milk is associated with HIV-1 transmission by breastfeeding. AIDS 2015; 29: 145–53.
16. Chang TS, Wiener J, Dollard SC, et al. Effect of cytomegalovirus infection on breastfeedin transmission of HIV and on the helath of infants born to HIV-infected mothers. AIDS 2015; 29:831–6.
17. Gompels UA, Larke N, Sanz-Ramos M, et al. Human cytomegalovirus infant infection adversely affects growth and development in maternally HIV-exposed and unexposed infants in Zambia. Clin Infect Dis 2012; 54:434–42.
18. Meyer SA, Westreich DJ, Patel E, et al. Postnatal cytomegalovirus exposure in infants of antiretroviral-treated and untreated HIV-infected mothers. Infectious diseases in obstetrics and gynecology 2014; 2014:989721.
19. Ben-Smith A, Gorak-Stolinska P, Floyd S, et al. Differences between naive and memory T cell phenotype in Malawian and UK adolescents: a role for cytomegalovirus? BMC Infect Dis 2008; 8:139.
20. Miles DJ, van der Sande M, Jeffries D, et al. Cytomegalovirus infection in Gambian infants leads to profound CD8 T-cell differentiation. J Virol 2007; 81:5766–76.
21. Kaye S, Miles D, Antoine P, et al. Virological and immunological correlates of mother-to-child transmission of cytomegalovirus in The Gambia. J Infect Dis 2008; 197:1307–14.
22. Guibert G, Warszawski J, Le Chenadec J, et al. Decreased risk of congenital cytomegalovirus infection in children born to HIV-1-infected mothers in the era of highly active antiretroviral therapy. Clin Infect Dis 2009; 48:1516–25.
23. Duryea EL, Sanchez PJ, Sheffield JS, et al. Maternal human immunodeficiency virus infection and congenital transmission of cytomegalovirus. Pediatr Infect Dis J 2010; 29:915–8.
24. Manicklal S, van Niekerk AM, Kroon SM, et al. Birth prevalence of congenital cytomegalovirus among infants of HIV-infected women on prenatal antiretroviral prophylaxis in South Africa. Clin Infect Dis 2014; 58:1467–72.
25. Mwaanza N, Chilukutu L, Tembo J, et al. High rates of congenital cytomegalovirus infection linked with maternal HIV infection

among neonatal admissions at a large referral center in sub-Saharan Africa. Clin Infect Dis 2014; 58:728–35.
26. Hamprecht K, Maschmann J, Jahn G, et al. Cytomegalovirus transmission to preterm infants during lactation. J Clin Virol 2008; 41:198–205.
27. Slyker J, Farquhar C, Atkinson C, et al. Compartmentalized cytomegalovirus replication and transmission in the setting of maternal HIV-1 infection. Clin Infect Dis 2014; 58:564–72.
28. Hamprecht K, Witzel S, Maschmann J, et al. Rapid detection and quantification of cell free cytomegalovirus by a high-speed centrifugation-based microculture assay: comparison to longitudinally analyzed viral DNA load and pp67 late transcript during lactation. J Clin Virol 2003; 28:303–16.
29. Kasonka L, Makasa M, Marshall T, et al. Risk factors for subclinical mastitis among HIV-infected and uninfected women in Lusaka, Zambia. Paediatr Perinat Epidemiol 2006; 20:379–91.
30. Filteau S, Baisley K, Chisenga M, et al. Provision of micronutrient-fortified food from 6 months of age does not permit HIV-exposed, uninfected Zambian children to catch up in growth to HIV-unexposed children: a randomised controlled trial. J Acquir Immune Defic Syndr 2011; 56:166–75.
31. Chisenga M, Siame J, Baisley K, et al. Determinants of infant feeding choices by Zambian mothers: a mixed quantitative and qualitative study. Matern Child Nutr 2011; 7:148–59.
32. Bates M, Monze M, Bima H, et al. High human cytomegalovirus loads and diverse linked variable genotypes in both HIV-1 infected and exposed, but uninfected, children in Africa. Virology 2008; 382:28–36.
33. Mattick C, Dewin D, Polley S, et al. Linkage of human cytomegalovirus glycoprotein gO variant groups identified from worldwide clinical isolates with gN genotypes, implications for disease associations and evidence for N-terminal sites of positive selection. Virology 2004; 318:582–97.
34. Tamura K, Stecher G, Peterson D, et al. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013; 30:2725–9.
35. Maschmann J, Goelz R, Witzel S, et al. Characterization of human breast milk leukocytes and their potential role in cytomegalovirus transmission to newborns. Neonatology 2015; 107:213–9.
36. Bate SL, Dollard SC, Cannon MJ. Cytomegalovirus seroprevalence in the United States: the national health and nutrition examination surveys, 1988–2004. Clin Infect Dis 2010; 50: 1439–47.
37. Yasuda A, Kimura H, Hayakawa M, et al. Evaluation of cytomegalovirus infections transmitted via breast milk in preterm infants with a real-time polymerase chain reaction assay. Pediatrics 2003; 111(6 Pt 1):1333–6.
38. Lemmermann NA, Krmpotic A, Podlech J, et al. Non-redundant and redundant roles of cytomegalovirus gH/gL complexes in host organ entry and intra-tissue spread. PLoS Pathog 2015; 11: e1004640.
39. Ciferri C, Chandramouli S, Donnarumma D, et al. Structural and biochemical studies of HCMV gH/gL/gO and Pentamer reveal mutually exclusive cell entry complexes. Proc Natl Acad Sci U S A 2015; 112:1767–72.
40. Grosjean J, Trapes L, Hantz S, et al. Human cytomegalovirus quantification in toddlers saliva from day care centers and emergency unit: a feasibility study. J Clin Virol 2014; 61:371–7.
41. Cannon MJ, Stowell JD, Clark R, et al. Repeated measures study of weekly and daily cytomegalovirus shedding patterns in saliva and urine of healthy cytomegalovirus-seropositive children. BMC Infect Dis 2014; 14:569.
42. Kuhn L, Aldrovandi GM, Sinkala M, et al. Differential effects of early weaning for HIV-free survival of children born to HIV-infected mothers by severity of maternal disease. PLoS One 2009; 4:e6059.
43. Gantt S, Carlsson J, Shetty AK, et al. Cytomegalovirus and Epstein-Barr virus in breast milk are associated with HIV-1 shedding but not with mastitis. AIDS 2008; 22:1453–60.
44. Frederick T, Homans J, Spencer L, et al. The effect of prenatal highly active antiretroviral therapy on the transmission of congenital and perinatal/early postnatal cytomegalovirus among HIV-infected and HIV-exposed infants. Clin Infect Dis 2012; 55:877–84.
45. Zambia Central Statistics Office. Zambia Demographic and Health Survey 2013–14. Ministry of Health Zambia and ICF International. Rockville, Maryland. 2014.
46. Goelz R, Hihn E, Hamprecht K, et al. Effects of different CMV-heat-inactivation-methods on growth factors in human breast milk. Pediatr Res 2009; 65:458–61.
47. Stowell JD, Forlin-Passoni D, Radford K, et al. Cytomegalovirus survival and transferability and the effectiveness of common hand-washing agents against cytomegalovirus on live human hands. Appl Environ Microbiol 2014; 80:455–61.
48. Boppana SB, Britt WJ. Recent approaches and strategies in the generation of antihuman cytomegalovirus vaccines. Methods Mol Biol 2014; 1119:311–48.
49. Fu TM, An Z, Wang D. Progress on pursuit of human cytomegalovirus vaccines for prevention of congenital infection and disease. Vaccine 2014; 32:2525–33.