A GENOMIC AND IMMUNOLOGICAL INVESTIGATION OF DENGUE VIRUS TRANSMISSION

James Ashall

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy
of the
University of London

December 2022

Department of Infection Biology

Faculty of Infectious Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the Medical Research Council

Research group affiliation: Professor Martin Hibberd

Abstract

Flaviviruses are single-stranded positive sense RNA viruses that include Dengue virus (DENV). Composed of 4 serotypes that are genetically and antigenically distinct thus allowing up to 4 infections in a lifetime, DENV is transmitted by *Aedes* species of mosquitos, which limits its presence to tropical and sub-tropical regions of the world. Over half of the world's population are now at risk of dengue, and this is forecasted to increase with continued urbanisation and climate change. DENV shows strong evidence of seasonality, peaking in case numbers when weather conditions are favourable to the propagation and survival of the insect vector. Yet, outside of these transmission seasons, reporting case numbers can drop to very low levels.

This thesis investigated the immunological and viral risk factors for dengue outbreaks. Using patient data collected as part of an index case led household study in Vietnam that enrolled 1,650 participants with 838 dengue positive index cases and their household contacts (HHCs). Initially, DENV whole genome sequencing approaches were developed and contrasted. This allowed sample appropriate techniques to be deployed on patient and HHC samples, which resulted in 511 whole genomes from 701 qPCR positive subjects. Predictors of transmissibility were then considered with little difference observed between serotypes and prior immune status (although without serotype specific information). Analysis of serological and viral marker data collected from enrolled participants revealed that sub-clinical disease from HHCs had similar viral RNA blood levels to index cases. Analysis of the sequenced DENV whole genomes showed long-term viral clade persistence in the study area despite fluctuations in reported case numbers. 90 transmission clusters were identified using genomic, epidemiological and serological data, revealing that within household transmissions were relatively rare.

Together, the findings indicate the importance of underreported sub-clinical disease in maintaining disease transmission and that human movement is an important driver of outbreaks.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the contributions sections within each chapter and/or as specified in the text. It is not being concurrently submitted for a degree or diploma or other qualification at any other University or similar institution.

COVID impact statement

The lockdown and subsequent inability to enter the school for laboratory work due to COVID caused a significant delay to the timeline of my PhD. Samples for analysis arrived at the end of February but unfortunately this work could not be finished until August. A significant chapter towards my thesis involved undertaking lab work with a collaborator but unfortunately this could not happen due to COVID restrictions on international travel. Both of these meant that alternatives had to be considered which involved making the project more data analysis and bioinformatics focussed. At the time the large amount of data which had previously been generated meant that the project could be changed to this new timeline with little difficulty.

This pivot to compensate for delays in lab work required me to learn a significant number of new skills which I had no/minimal prior experience in, learning these skills took up more time than was anticipated as there were no dedicated training courses or bioinformatic supervision available in my direct group, so I had to be self-taught.

COVID has additionally affected delivery times for several vital consumables especially qPCR reagents. The delay to many of these reagents caused knock-on effects with down-stream analysis.

Acknowledgments

I would like to thank my supervisors Prof. Martin Hibberd, Dr. Julius Hafalla and Dr. Stéphane Hué for their support and guidance throughout my PhD. I am additionally thankful for all the work done in collaboration with the team based at Nha Trang, especially Dr. Lay-Myint Yoshida.

Thanks to Dr. Sonal Shah who supervised me at the start in the laboratory and kept her unwavering support throughout despite having long left the group, she made my time so much more enjoyable, and I learnt so much from her. To Drs Jody Phelan and Neneh Salah thanks for their support and guidance with the bioinformatics component of the project I finish with significantly more skills than when I started.

I'm grateful to my friends and flatmates who have supported me throughout, especially Dr Giulia Beraud who reminded me there's life outside of work and provided me with a home during these final stages. Thanks to Dr Josie Williams who was so amazing to live with for 2 years, glad I had someone to share the ups and downs with.

Table of Contents

CHAPTER 1 INTRODUCTION 14

1.1 The increasing significance of Dengue as a global disease 14
1.2 Dengue virus infection and the complicated human immune response 26
1.3 The challenges associated with diagnosing current and historical Dengue infections 32
1.4 Current Dengue control measures and how their success can be determined 38
1.5 Challenges associated with generating sequencing data from viral infections 39
1.6 How viral sequencing has the potential to enhance our understanding of transmission 42
1.7 Thesis Aims 45

CHAPTER 2 DEVELOPMENT AND REVIEW OF NEXT-GENERATION RNA VIRAL SEQUENCING METHODS 48

2.1 Introduction 48
2.1.1 Chapter Aims 52
2.2 Methods of Sequencing Approaches Investigated 53
2.2.1 qPCR Diagnosis of DENV 53
2.2.2 Large Amplicon sequencing approach 56
2.2.3 Small Overlapping Amplicons 59
2.2.4 Target Enrichment 62
2.2.5 Modified RNA-seq technique using the Peregrine system 66
2.3 Illumina sequencing and resulting consensus generation pipeline 72
2.4 Results 73
2.4.1 qPCR results 73
2.4.2 Comparison of Coverage and sequencing depth 76
2.4.3 Comparison of Viral reads 79
2.4.4 Successful library preparation 80
2.5 Peregrine library sequencing analysis 81
2.6 Discussion 83

CHAPTER 3 AN INVESTIGATION OF THE DIFFERENT VIROLOGICAL AND SEROLOGICAL MARKERS OF INFECTION IN A DENV ENDEMIC POPULATION SHOWS DIFFICULTIES IN IDENTIFYING THOSE AT RISK OF A SEVERE INFECTION 85

3.1 Introduction 85
3.1.1 Chapter Aims 88
3.2 Methods 89
3.2.1 Study Population and Sample Collection 89

3.2.2 Serological analysis 89
3.3 RESULTS 91
3.3.1 Prevalence and distribution of DENV serotypes during the study 91
3.3.2 Observed viraemic differences between serotypes 94
3.3.3 Antibody profiles of reporting population 96
3.3.4 Effect of primary/post-primary status of viraemia 98
3.3.5 Symptomatic outcomes of reporting population 102
3.3.6 Differences in viraemia between sub-clinical and reporting populations 105
3.4 DISCUSSION 107
CHAPTER 4 A PHYLOGENETIC STUDY OF DENGUE VIRUS IN URBAN VIETNAM SHOWS LONG-TERM PERSISTENCE OF ENDEMIC STRAINS 111
4.1 ABSTRACT 116
4.1.1 Background 116
4.1.2 Methods 116
4.1.3 Results 116
4.1.4 Conclusions 117
4.2 INTRODUCTION 118
4.3 METHODS 121
4.3.1 Study Population 121
4.3.2 DENV RNA detection 121
4.3.3 Sequence Generation 122
4.3.4 Phylogenetic analysis 123
4.3.5 Selection analysis 124
4.4 RESULTS 125
4.4.1 Genetic diversity of DENV1-4 genomes isolated in Nha Trang, Vietnam, between 2016-2019 125
4.4.2 Evidence for Dengue virus persistence in the study area 133
4.5 DISCUSSION 138
4.5.1 Conclusions 140
CHAPTER 5 A MIXED METHODS STUDY OF DENGUE VIRUS TRANSMISSION IN URBAN VIETNAM 146
5.1 ABSTRACT 146
5.2 INTRODUCTION 148
5.2.1 Chapter Aims 154
5.4 METHODS 155
5.4.1 Study cohort 155
5.4.2 Phylogenetic analysis 155

5.4.3 Identifying transmission pairs 156
5.4.4 Identifying exposure to DENV among household contacts..... 157
5.5 RESULTS 159
5.5.1 Recruitment 159
5.5.2 Transmission chain analysis 162
5.5.3 Serological evidence for transmission 172
5.5.4 Combining serological and genomic data reduces the number of transmissions consistent with intra-household events. 172
5.6 DISCUSSION 178
CHAPTER 6 CONCLUSIONS AND FUTURE WORK 182
6.1 IMMUNOLOGICAL AND VIRAL RISK FACTORS FOR DENV DISEASE OUTCOME 184
6.2 IDENTIFYING PATHOGEN PERSISTENCE IN POPULATIONS USING PHYLOGENETIC TECHNIQUES 186
6.3 USING NGS TO INVESTIGATE PATHOGEN DYNAMICS DURING AN OUTBREAK 187
CHAPTER 7 REFERENCES 190

List of Figures

Figure 1.1: Global distribution of DENV cases.. 15
Figure 1.2: Effects of various meteorological measurements on DENV incidence in Thailand. 17
Figure 1.3: Diagram depicting the organisation of *flaviviral* genomes and their subsequent processing into mature viral proteins. 19
Figure 1.4: Phylogenetic tree showing genetic relatedness of the 4 DENV serotypes and their sub-order genotypes 20
Figure 1.5: Replication cycle of DENV in animal cells.. 22
Figure 1.6: Map of Vietnam showing the location of Nha Trang city within Khanh Hoa Province 25
Figure 1.7: Laboratory diagnostic options based on viraemia and immune responses. 34
Figure 1.8: Example of 500 basepair (bp) overlapping primer scheme across a DENV genome. 41
Figure 2.1: Overview of different sequencing approaches employed during study. 51
Figure 2.2: Overview of Peregrine library preparation process 68
Figure 2.3: Box-and-whisker plot showing qPCR Ct result of Pan-DENV assay combined with serotyping qPCR result 74
Figure 2.4: Logistic regression of probability of positive serotyping qPCR result based on the Pan-DENV qPCR value 75
Figure 2.5: Distribution of sequencing reads and coverage for a control DENV sample across sequencing methods employed 78
Figure 2.6: Sequencing success by library preparation strategy 81
Figure 2.7: Read depth and coverage plots of Peregrine library preparation sequencing approach 82
Figure 3.1 Distribution of dengue serotypes in the study area.. 92
Figure 3.2 Map of study location within Vietnam, with location of serotyped subjects home address location plotted.. 93
Figure 3.3 Viraemia differences between serotypes. 95

Figure 3.4 Inter-individual variability of serological responses to DENV of Patients and associated household members at their first visit and with complete data..... 97
Figure 3.5 Effect of prior exposure to DENV on Viraemia. 100
Figure 3.6 Investigation of predictors of severe disease. 104
Figure 3.7: Comparison of viraemia between reporting Index cases and their associated sub-clinical Household contacts..... 106
Figure 4.1: Stacked bar chart of number of full genome sequences generated during the study and their breakdown by serotype and genotype. 125
Figure 4.2: Midpoint rooted maximum-likelihood trees of 2446 publicly available full genome length DENV1 sequences..... 128
Figure 4.3: Midpoint rooted maximum-likelihood trees of 1891 publicly available full genome length DENV2 sequences..... 130
Figure 4.4: Midpoint rooted maximum-likelihood trees of 407 publicly available full length DENV4 sequences..... 132
Figure 4.5: Number of sequences in each clade sampled each month from October 2016 to April 2019 with total number of NS1 positive individuals recorded during the same time period..... 134
Figure 4.6: Dated phylogenies of 5 DENV clades sampled in Nha Trang between 2016 and 2019..... 136
Figure 4.7: Timescale of sampled and predicted historical samples based on linear regression analysis of phylogenetic diversity overlayed on historical and recently reported DENV cases in Nha Trang..... 137
Figure 5.1: Comparison of contact tracing maps between a respiratory and vector-borne disease when an infected individual walks between 2 locations..... 149
Figure 5.2: Estimates of human and mosquito infection periods are used to distinguish different infection profiles of sequential and parallel infections involving human hosts and the mosquito vector..... 151
Figure 5.3: Workflow for identifying putative transmission pairs from whole genome viral sequences 158
Figure 5.4: Optimisation of SNP threshold for determining clusters..... 163
Figure 5.5: Key findings from analysis of identified transmission pairs and distances between them 167

Figure 5.6: Distance analysis of transmission pairs identified 169
Figure 5.7: Map showing sampled pairs household and given daytime locations. 171
Figure 5.8: Midpoint rooted maximum-likelihood phylogenetic trees for full genome length sequences generated during study for serotype D1, D2 and D4..... 174
Figure 5.9: Representation of transmission pairs by the different experimental techniques employed 177

List of Tables

Table 1.1: The Flaviviridae family 18
Table 1.2: Summary of DENV structural and Non-structural proteins..... 24
Table 2.1: PCR cycling conditions for NEB polymerase kit..... 53
Table 2.2: Quantifast multiplex qPCR kit run conditions 55
Table 2.3: List of primers used in large amplicon sequencing protocol..... 57
Table 2.4: PCR cycling conditions for large amplicon generation..... 57
Table 2.5: PCR reaction conditions for amplicon generation 60
Table 2.6: PCR cycling conditions for amplicon generation..... 60
Table 2.7: Primer sequences used in Peregrine library preparation technique 70
Table 2.8: Reaction mixture for Peregrine PCR..... 71
Table 2.9: Peregrine PCR protocol 71
Table 2.10: Breakdown of qPCR results for Pan-DENV and Serotyping assays..... 74
Table 2.11: Breakdown of target viral reads by sequencing strategy..... 79
Table 3.1: Force of infection for serotypes D1, D2 and all patients across the final year of the study period (April 2018 – April 2019) with 95% confidence intervals. 101
Table 5.1: Characteristics of the study participants 160
Table 5.2: The number of household members that experienced dengue infections in the households of reported NS1+ index cases. 161
Table 5.3: Distribution of transmission pairs identified characterised by time interval and optimal distance between cases..... 166
Table 5.4: Breakdown of serotype seroconversions observed among HHCs of index cases with a positive qPCR result..... 176
Table 5.5: Parallel exposures between participants inferred by various detection techniques. 176

List of Abbreviations

ADE	Antibody-dependent enhancement
bp/kb	Nucleotide basepairs/thousand basepairs
CDC	Centers for Disease Control
D1	Dengue Virus – Serotype 1
D2	Dengue Virus – Serotype 1
D3	Dengue Virus – Serotype 1
D4	Dengue Virus – Serotype 1
DENV	Dengue Virus
DF	Dengue fever
DHF	Dengue Haemorrhagic fever
DNA	Deoxyribonucleic acid
DSS	Dengue shock syndrome
E	Viral envelope protein
ELISA	Enzyme-linked immunosorbent assay
FcγR	Fc receptor gamma
GIT	Gastro-intestinal tract
HCV	Hepatitis virus C
HIV	Human Immunodeficiency virus
IFN-γ	Interferon gamma
IgG	Immunoglobulin M
IgM	Immunoglobulin G
IL	Interleukin
NGS	Next Generation Sequencing
NHP	Non-human primate
NS	Non-structural viral proteins
PCR	Polymerase Chain Reaction
prM	Viral precursor membrane protein
PRNT	Plaque-reduction neutralisation test
QC	Quality control
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic acid
RNA-Seq	mRNA sequencing
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
SARS-COV2	Severe acute respiratory syndrome coronavirus 2
TGF-β	transforming growth factor beta
TNF-α	Tumour Necrosis Factor alpha
WGS	Whole Genome Sequencing
WHO	World Health Organisation
ZIKV	Zika virus

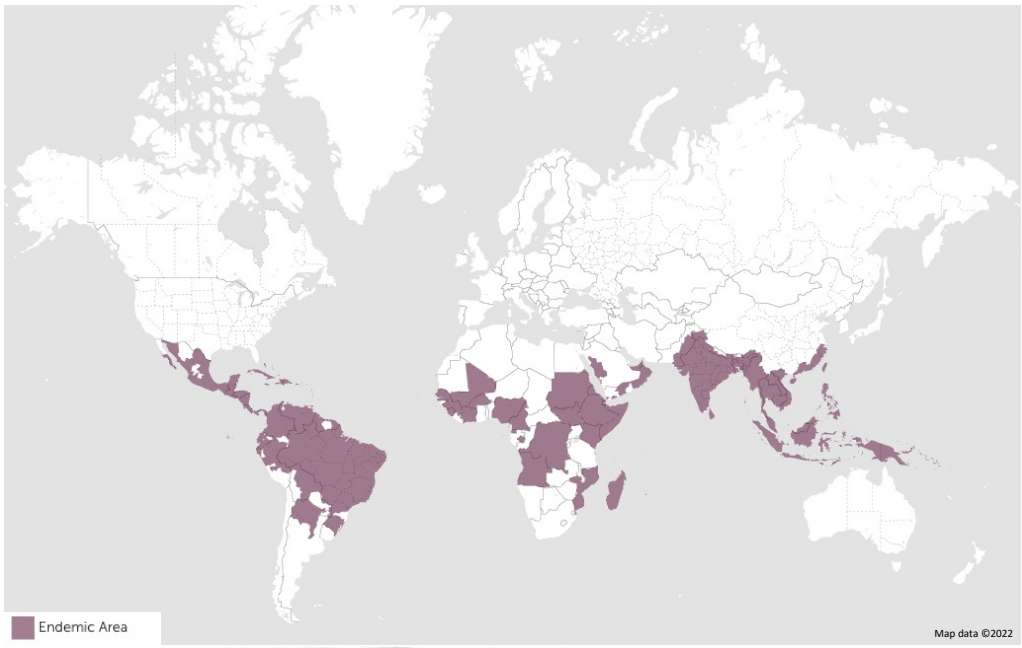
Chapter 1 Introduction

1.1 The increasing significance of Dengue as a global disease

The World Health Organisation (WHO) considers dengue as a major global public health challenge throughout tropical and sub-tropical regions of the world. DENV saw a 30-fold upsurge worldwide between 1960 and 2010, due to increased human population growth rate, global warming, rapid urbanisation, unsatisfactory vector control and lack of suitable health care facilities[1,2]. Due to population growth in endemic areas, around 2.5 billion people are now at risk of infection[1] and there are estimated to be 400 million cases every year, although fortunately mortality is generally very low, less than 1%[3]. Southeast Asia is the region where around two-thirds of the global endemic DENV population live and within this area it is a leading cause of death, hospitalization and economic burden[4]. With increasing urbanisation and climate change the burden of DENV is modelled to grow in these areas and expand to surrounding areas which were previously unsuitable for the mosquito vector[5] and there are frequently cases being reported in previously non-endemic areas such as Europe and the United States of America[6] as the climate becomes more viable for the vector. Countries which are reported as being endemic for DENV are shown in Figure 1.1A and burden of disease in 2021 is shown in Figure 1.1B.

The number of DENV infections reported has been steadily increasing over the past few decades with 505,430 cases reported in the year 2000 to 5.2 million in 2019, in addition worldwide deaths from 2000 to 2015 increased from 960 to 4032[7,8]. It is important to note that a significant increase in this reporting is likely due to better testing and reporting as many of these countries have significantly improved their healthcare infrastructure over the same time period as they have become wealthier.

A.



B.

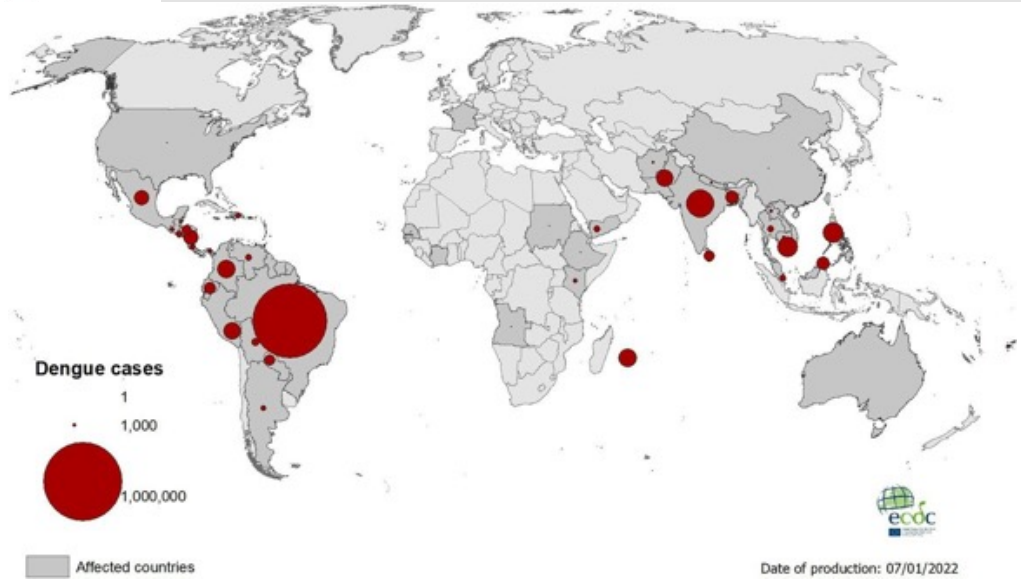


Figure 1.1: Global distribution of DENV cases. A. World map showing countries/regions which are regarded as endemic (as of 14.03.2022). Generated using the CDC HealthMap tool: https://www.healthmap.org/dengue/en/. B. Number of DENV cases reported worldwide during 2021 (from the European Centre for Disease Prevention and Control[9]).

Although present in tropical and sub-tropical regions of the world, DENV can show evidence of significant periodicity from year-to-year, with Southeast Asia showing a significant difference between years and generally showing peaks in case numbers every three to five years[10]. In addition to varying by year, case numbers show seasonality and are highly correlated with temperature and periods of high rainfall both of which are associated with vector density[11] (for example Northern Thailand in **Figure 1.2**).

The mechanisms behind this periodicity are not fully understood but are thought to be the interplay between meteorological phenomena such as the El Niño weather pattern and changes in immunity in the endemic population as protection wanes or susceptible individuals enter (or are born into) the area[12].

Cyclical cases strongly correlate with weather conditions indeed in many areas the virus can appear to disappear and in several countries, especially China and Taiwan, there is evidence that DENV is a seasonal disease which is introduced from abroad most years[13]. In tropical countries however, the virus can persist all-year despite periods of very low case numbers[14].

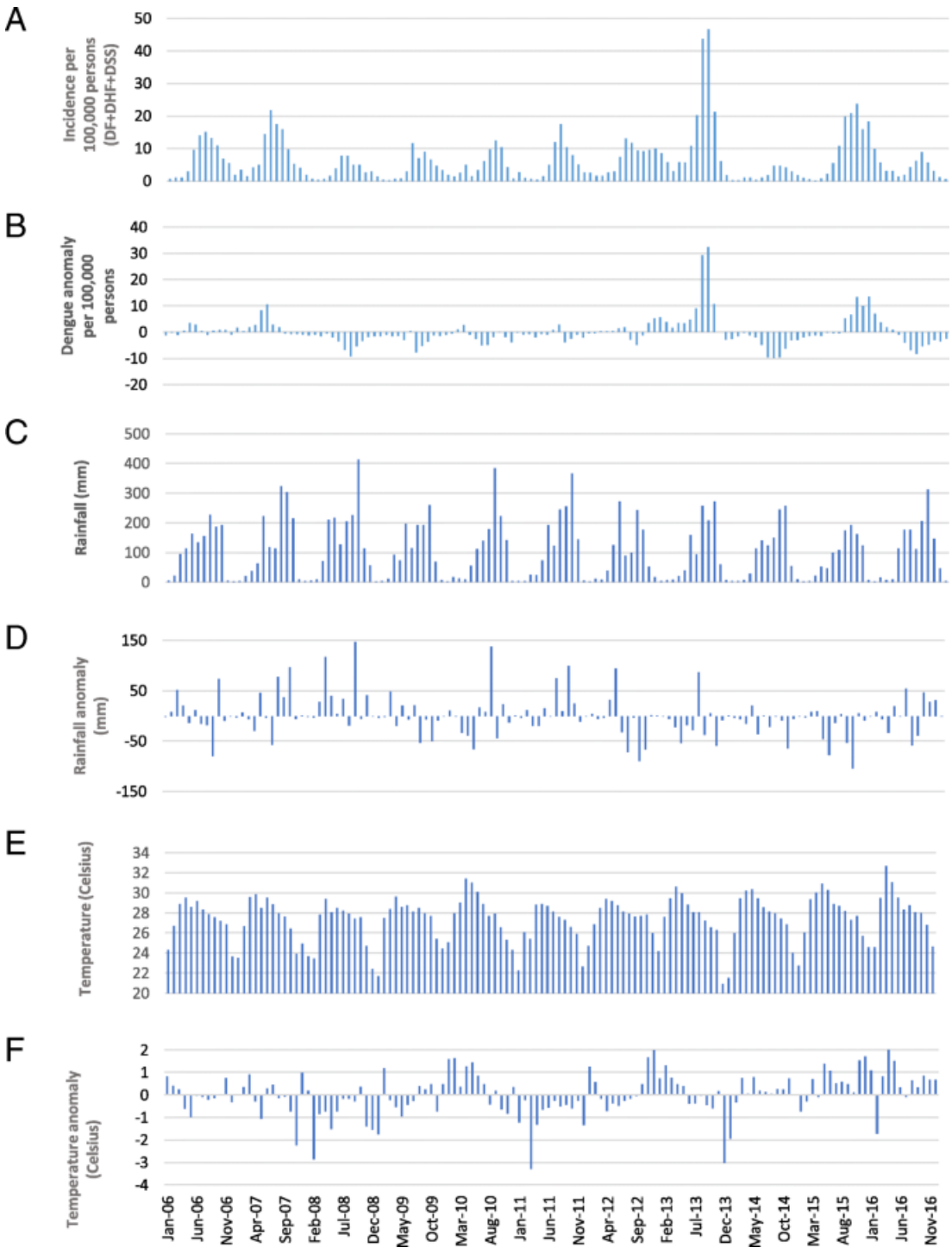


Figure 1.2: Effects of various meteorological measurements on DENV incidence in Thailand. A. DENV incidence, B. DENV incidence difference from mean, C. Rainfall, D. Rainfall difference from mean, E. temperature, F. Temperature difference from mean. From Phanitchat *et al.* 2019[12].

Virology of DENV

DENV along with other flaviviruses (Table 1.1) contains a positive-sense single-stranded RNA genome of around 10,500 nucleotides in length that encodes a precursor polyprotein which is around 3,500 amino acids long[15]. This polyprotein is subsequently cleaved into ten polypeptides: three structural proteins (the capsid, precursor membrane (prM) and envelope glycoprotein (E)) that form the virion and seven non-structural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) that are essential for viral replication.

Table 1.1: The Flaviviridae family

Genus	Hosts	Transmission	Notable members
Flavivirus	Humans, mammals, mosquitos, ticks	Zoonosis, arthropod bite	Yellow fever, Dengue and Zika viruses
Hepacivirus	Humans	Sexual, Blood-borne	Hepatitis C virus
Pegivirus	Mammals	Unknown	Pegivirus A, B,C
Pestivirus	Mammals	Vertical	Bovine viral diarrhoea virus, classical swine fever virus

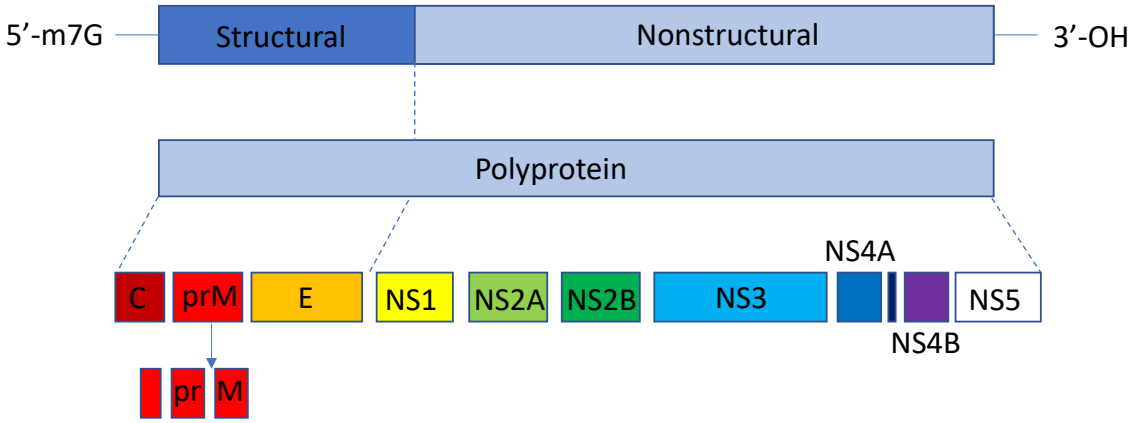


Figure 1.3: Diagram depicting the organisation of *flaviviral* genomes and their subsequent processing into mature viral proteins. Adapted from Campos et al. 2018[15]

The four different DENV serotypes can be further broken down into different genotypes, which have traditionally denoted where subtypes have circulated, however as there has been more transmission of DENV between geographical locations as globalisation and international travel increases, these genotype denominations are being made more redundant[16], the phylogenetic relationships between the serotypes and genotypes are depicted in Figure 1.4.

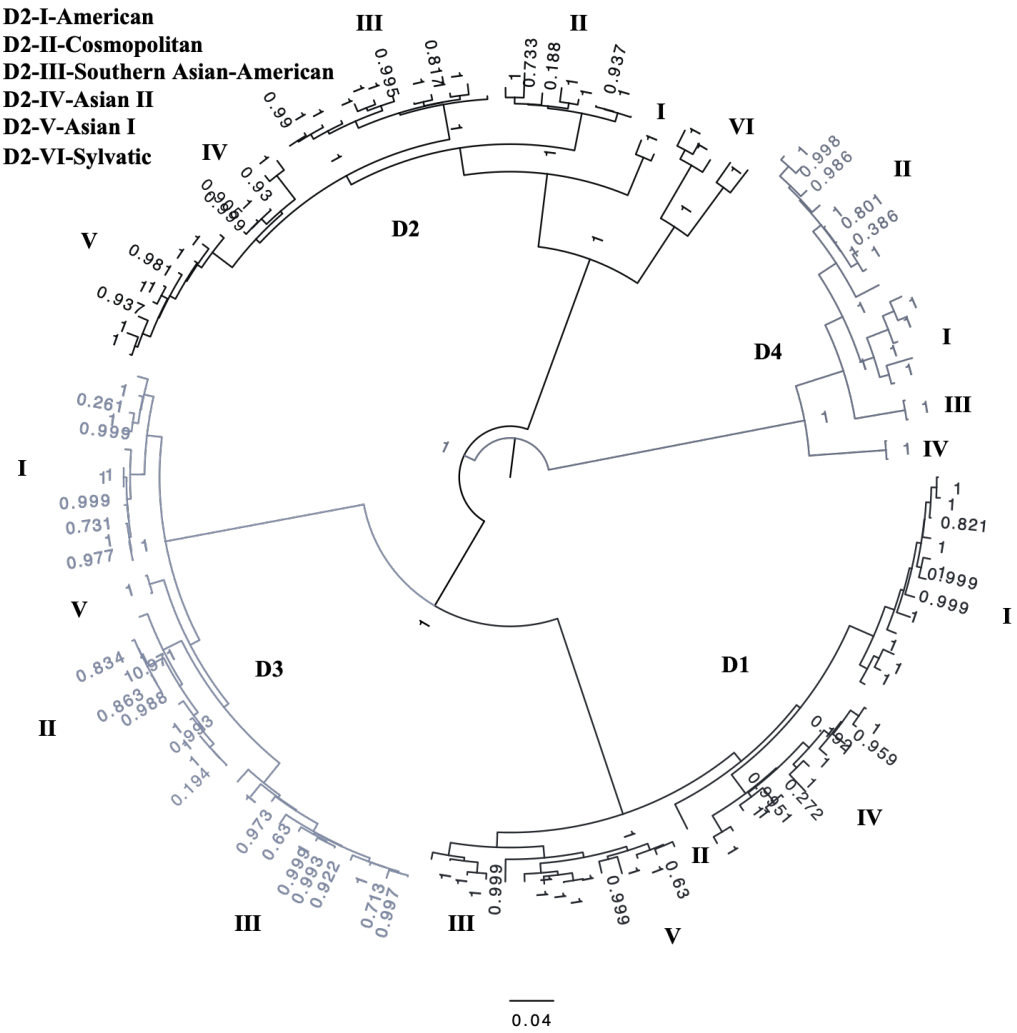


Figure 1.4: Phylogenetic tree showing genetic relatedness of the 4 DENV serotypes and their sub-order genotypes shown on the outside with bootstrap supports given. Generated with type sequences from Dengue genome detective typing tool[17].

Viral proteins and their use in entry, replication and immune evasion

DENV must be able to infect both mammalian and insect cells to enable its continued transmission cycle. Following feeding on an infected host, the mosquito takes up infectious virus, a number of mosquito proteins have been suggested as attachment targets for virus entry with heat shock protein 70, R80, R67 and an additional 45 kDa protein all being shown to interact with DENV[18,19]. The virus fully disseminates throughout the mosquito and eventually infects the salivary glands where it can be inoculated into a subsequent human host. The following human cellular protein have been found as attachment target for viral proteins: Heparin sulphate Hsp 90, CD14, a 78 kDa glucose-regulated protein (GRP78/BiP), a 37/67 kDa high-affinity laminin receptor and C-type lectin receptors[19].

An additional potential entrance mechanism for DENV is through the binding of non-neutralising antibodies and subsequent phagocytosis by immune cells. This is termed antibody-dependent enhancement and is one of the hypothetical mechanisms behind the severe haemorrhagic form of DENV infection that can be seen[20].

Regardless of host infected, the process of viral replication is the same following viral entry (Figure 1.5). Viral entry itself starts with a conformational change of the E protein which is triggered by a low pH (6.5), this is subsequently followed by clathrin-mediated endocytosis. At this stage the capsid protein releases the viral genome into the cell cytoplasm and subsequently translates it into a single polyprotein in association with endoplasmic reticulum-derived membranes. This polyprotein is processed further into the structural and non-structural viral proteins by the action of different cellular and virus-derived proteases[21].

Following translation and post-translational processing, viral proteins are ready for assembly and trafficking out of the cell. The viral genome is replicated using the non-structural proteins, which initiate the replication process itself by using a negative strand intermediate as a template and through circularization of the genome multiple copies of positive-strand viral genomic RNA are produced[22]. Afterwards, this newly

formed RNA is packaged by the C protein to form a nucleocapsid and, meanwhile, a heterodimer complex formed by prM and E proteins leaked into the ER lumen is thought to be involved in the virion budding process (Kuhn et al., 2002). The immature virus particles mature by using the secretory pathway, and the trans-Golgi network triggers dissociation of the prM/E heterodimers in an acidic environment.

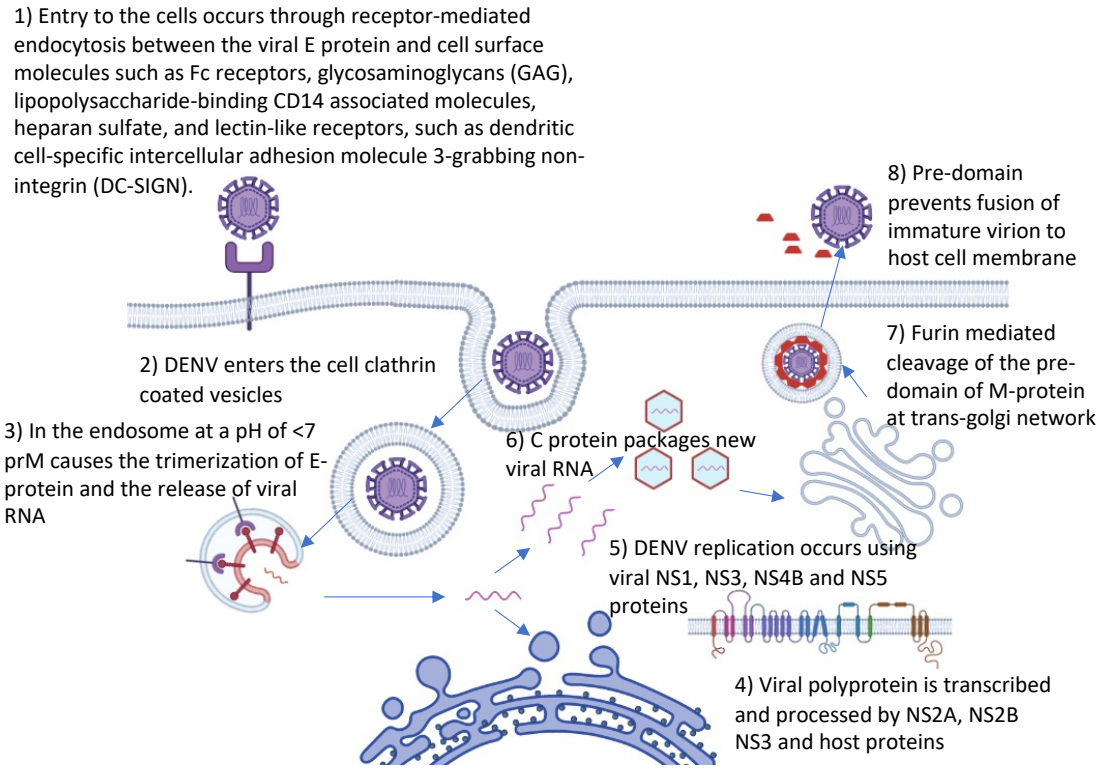


Figure 1.5: Replication cycle of DENV in animal cells. Adapted from Nasar *et al.* 2019[23].

The proteins encoded by the DENV genome are used by the virus for replication within a host cell and to evade host immune responses. The viral proteins NS1 and Envelope are most relevant to this thesis however others are summarized in Table 1.2.

NS1 is found in all flaviviruses and is approximately 1056 nucleotides in length, the resulting synthesised protein contains 352 amino acids and has a molecular weight of 46-55 kDa that is dependent on glycosylation status. For its size NS1 shows significant difference between the flaviviral species with phylogenetic analysis between the species showing conservation that ranged from 50% to ~80%[24].

In DENV infections NS1 is secreted in high levels into the extracellular environment as a barrel-shaped soluble hexamer[25]. NS1 is therefore a common target for antigen based diagnostics as it can persist in serum for longer than RNA[26], assays for its detection can be cheaper and easier to run than RNA based detection and as NS1 has high observed diversity between flaviviral species there can be greater specificity than with other potential viral target proteins.

Table 1.2: Summary of DENV structural and Non-structural proteins

Structural proteins	**Protein functions**
C	Capsid protein, viral RNA packaging
prM	Prevents pre-mature fusion to host cell membrane
E	Envelope glycoprotein, receptor binding, and entry into host cell
Non-structural proteins	**Protein functions**
NS1	Viral replication and immune pathogenicity Acts as a viral toxin, inducing endothelial hyperpermeability Secreted form acts as an immune agonist, inducing a pro-inflammatory response and endothelial cell.
NS2A	Interacts with 3'-UTR, critical for RNA synthesis Virion assembly Role in virus-induced cytopathic effect
NS2B	Co-factor regulates NS3 protease activity Immune evasion, targets cGAS DNA sensor for lysosomal degradation, inhibiting IFN-I response in infected cells
NS3	Serine-protease RNA helicase RTPase/NTPase
NS4A	Membrane remodelling Autophagy induction Antagonising host cell IFN response Contributes to viral replication
NS4B	Forms Virus Replication Complex Immune evasion
NS5	Methyltransferase, required for 5'-RNA capping, aiding viral immune evasion RdRp

DENV in Vietnam and Nha Trang City

Vietnam is one the countries with one of the most consistently high case numbers every year with a modelled yearly burden of 2 million infections and an average of 95,000 cases reported annually[27]. The disease was first reported in 1959 and has since become endemic with higher and consistent incidence rate in southern region than the north of the country[28]. DENV was regularly introduced into the Central and Northern Vietnam from the endemic southern regions through frequent human migration, resulting in increased dengue cases throughout the country[29].

Within Vietnam is Nha Trang, a coastal city situated in the South-Central region (Figure 1.6), the population in 2021 was reported as 332,410 and has a mean growth rate of 1.19% per year[30]. Nha Trang City reported seasonal epidemic peaks in dengue cases between 2006 and 2016 during the wet season (July-August or November-January), with all four DENV serotypes co-circulated among hospitalised patients within a single season; it is also been reported that the highest dengue incidence was seen in the central urban wards of the city[31].

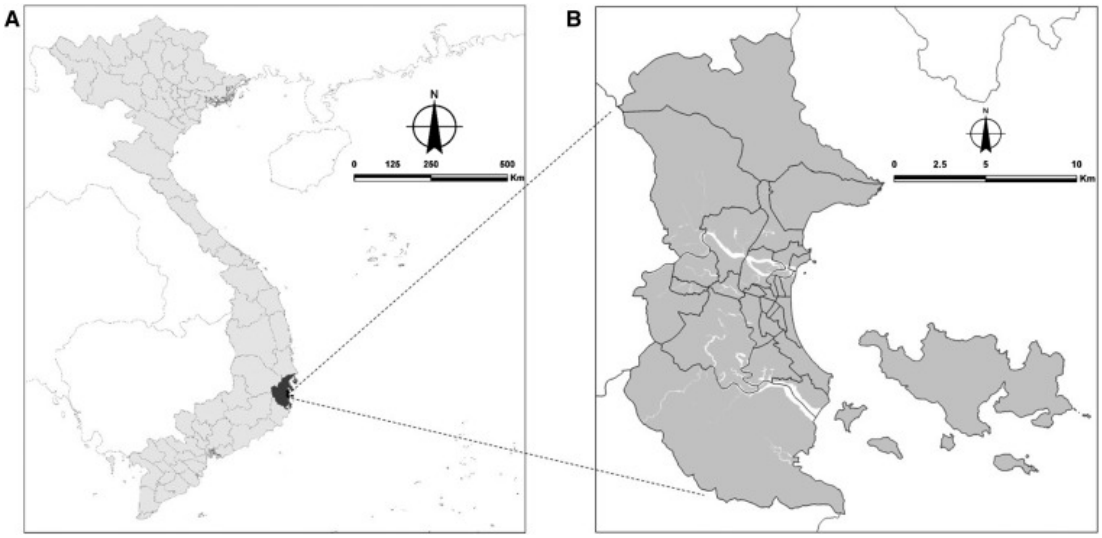


Figure 1.6: Map of Vietnam showing the location of Nha Trang city within Khanh Hoa Province (A) and the administrative boundaries (B). From Quyen *et al.* 2018[31]

1.2 Dengue virus infection and the complicated human immune response

Following infection with DENV in a susceptible individual, there is a 4-10 day incubation period after which time symptoms can vary from individual to individual depending on factors such as prior flaviviral exposure (other DENV serotype and Zika virus)[32,33], however frequently reported symptoms resemble a severe flu-like illness[3]. Infection rarely results in death and is generally self-limiting. Symptomatic DENV infection has been divided into two categories based on how presenting symptoms can indicate risk of developing a more severe infection.

The WHO lists certain criteria which should cause clinicians in endemic areas to suspect DENV infection; these are when a high fever (>40°C) is accompanied by any 2 of the following symptoms: severe headache, pain behind the eyes, muscle and joint pains, nausea, vomiting, swollen glands or rash[2]. These are not sufficiently specific for diagnosis alone as many other infections can present with these symptoms, including other vector-borne infections such as Zika, Yellow fever and Japanese encephalitis viruses.

Research into the pathogenesis of the severe haemorrhagic form of DENV infection is complicated by the lack of a suitable animal model which replicates the symptoms suffered by humans[34]. Severe DENV is defined by an acute high-grade fever which lasts from 2 days to 1 week with accompanied haemorrhagic episodes (which include according to the WHO: “Petechiae, purpura, ecchymosis, epistaxis, gingival and mucosal bleeding, GIT or injection site, hematemesis and/or malena”[3]).

The WHO has also produced a list of warning signs which may indicate a patient is at risk of developing the severe form of DENV infection, these include any two of the following: severe abdominal pain, persistent vomiting, rapid breathing, bleeding gums, fatigue, restlessness or blood in vomit[2,8].

Treatment options

There are currently no specific drugs or treatments licensed for DENV infection, however the vast majority of DENV infections are self-limiting. For severe DENV infection cases, medical care which prioritises maintaining fluid levels can improve prognosis significantly[2]. Paracetamol is commonly given in many countries to help with fever reduction, however as hepatocytes are targeted during DENV infection there are increased risks of liver damage with no measurable benefit to fever or pain reduction[35,36]

As with other acute viral infections the usefulness of targeted anti-virals is often limited, as by the time a patient presents with severe symptoms, they are often only viraemic for 1 or 2 more days and viraemia is generally already waning[37]. For example, anti-influenza drug therapeutics are only effective at reducing symptoms by 0.5-1.5 days when given within 48 hours of symptom onset[38], but have very high efficacy when given prophylactically[39]. As with dengue disease it is likely an over-reactive immune response rather than runaway viral infection that leads to severe DENV infections, this limited effect is likely to be the same with potential DENV therapeutics. Despite this, there is some evidence that reducing viraemia may lead to less hospitalisation and improved recovery in DENV infections[40]. There have therefore been many candidate anti-DENV drugs over the years, some of which have made it to clinical trials[41]. One such recent example is an anti-viral which targets the interaction between the NS3 and NS4B which has made it through safety and efficacy testing and has been shown to be effective at reducing viraemia *in vivo* mouse models for up to 4 days after infection[42]. This anti-DENV drug, as with anti-influenza drugs, has also been shown to be most effective as a prophylactic in limiting infection. Unlike influenza however where effective ring prophylaxis can limit the spread of a pathogen from an infected individual as it is transmitted through relatively close contact[39], DENV is transmitted by a vector which can add an unknown dimension to the viral transmission area, therefore identifying and giving prophylaxis to those most at risk is difficult. One potential use could be to limit exposure to tourists to endemic areas, so therapeutics could be taken in the same way anti-malarial drugs are used

by visitors to endemic countries[43] or as a preventative measure during peak seasons. There are also studies ongoing and previously publish to determine how DENV transmits around a population and who should be targeted with treatment to limit its spread.

Vaccines for DENV

DENV remains a good candidate for a vaccine as it causes a significant amount of morbidity in countries where it is found and tourists to the area are often infected, therefore local and international vaccines could be economically viable. Additionally, humans are both able to naturally clear infection and develop immunity to reinfection of homotypic serotypes. Despite this, the increased pathology associated from secondary infections, which is attributed to the cross-reactive sub-neutralising immune response from an initial infection, has so far made designing a vaccine challenging.

An additional difficulty associated with DENV vaccine development has been a lack of a suitable animal model other than non-human primates (NHP). Since mice are naturally resistant to DENV infection, human cell chimeric mice and immunodeficient mice sensitive to DENV infection, as well as mouse-adapted virus[44] have been established to be used as animal models, these often do not reflect the complicated immune responses that are seen in human and NHP infections however[45].

Despite the difficulties associated with producing a vaccine there is a licensed vaccine available and there have been attempts to produce one using strategies including: live attenuated, inactivated, recombinant subunit, viral vectored, and DNA mediated[46]. The only currently licensed vaccine against DENV is called Dengvaxia® (Sanofi, Paris, France), it has been licensed for use in children over the age of 9, in 20 countries in South America and Asia, however, uptake has been very low.

The vaccine itself is a live attenuated vaccine that, through the use of recombinant DNA technology, contains the premembrane (prM) and envelope genes from the wildtype DENV Thailand PUO-359/TVP-1140 (Serotype 1), Thailand PUO-218

(Serotype 2), Thailand PaH881/88 (Serotype 3), and Indonesia 1228 (TVP-980) (Serotype 4) that are embedded in the yellow fever nonstructural genes[47]. Although originally recommended for everyone over 9 years of age, subsequent long-term safety data showed that the risk of hospitalisation and severe disease was greater among those who had received the vaccine without a prior dengue infection, when compared with a placebo control group[48]. As a result of these findings, in April 2018, the WHO Strategic Advisory Panel recommended that DENV serostatus should be factored in to ensure that only DENV-seropositive individuals received the vaccine[49].

Host immune response to Infection

Following a single DENV infection and recovery, there are three immune outcomes: 1. Long lasting and strong immunity against an infection of the same DENV strain, 2. Transient immunity towards an infection of different serotypes of DENV for up to 6 months, and 3. Insufficient protection against a subsequent infection from a different DENV serotype after 6 months that can result in more severe disease[50].

The immune response to DENV infection affects the clinical outcome with a notable difference in symptoms and severity observed between a primary infection, the second and then subsequent infections[51]. Several studies have investigated the role of time intervals between successive DENV infections on the disease outcome with two such studies showing that the average time for asymptomatic secondary infection was significantly shorter than for symptomatic or severe infection [52,53]. This supports the leading hypothesis that waning or non-neutralising antibody titres from prior infections increase the infectivity of DENV in subsequent infections[54], a phenomenon termed antibody dependent enhancement (ADE).

During a DENV infection a typical anti-viral adaptive immune response is produced with anti-viral IgM antibodies later being replaced with IgG. Antibodies are produced against both the structural and non-structural viral proteins with anti-structural antibodies associated with opsonisation and neutralisation, whilst the roles of all non-structural antibodies, apart from anti-NS1 which can induce apoptosis in

infected endothelial cells though a caspase dependent route[55], are not fully understood.

During this initial exposure to DENV there is significant evidence that cross-reactive and weakly neutralising antibodies to other serotypes are produced which are weaker, less protective and shorter lasting[56]. It is hypothesised that these antibodies can form immune complexes with DENV and can increase infectivity by activation of the FcγR[20,57,58]. Indeed monoclonal antibodies produced against DENV E protein have been shown to enhance infection when below the threshold for neutralisation[59].

The cellular response to DENV infection is similarly important in producing protective and pathogenic effects. T cells generated in initial DENV exposure can become predominant during a subsequent infection, resulting in an expansion of low-affinity memory T cells, which provide ineffective viral control and can contribute to immunopathology and severe dengue disease by the production of excessive amount of inflammatory cytokines[60].

Through immunisation studies of monovalent vaccines of all four serotypes performed on healthy volunteers it was found that CD8+ T-cell responses are directed against a variety of viral proteins, with participant immune cells either recognising the NS3 or NS1.2a protein[61], it was found that donors who had a response to the NS3 protein had more cross-reactive T-cells to other serotypes.

CD4+ T-cells can produce interferon gamma (IFN-γ), tumour necrosis factor-α (TNF-α) and TNF-β in response to DENV antigens, the production of these cytokines has been implicated in contributing to the pathogenesis sometimes seen in secondary infections[62]. IFN-γ plays a key role in DENV infection of monocytes, especially during a secondary infection, as it upregulates Fc gamma rectors which is the primary receptor through which they enter in ADE situations[63]. Additionally, DENV infection-associated liver pathology is likely a result of CD4+ T-cells destroying non-antigen presenting target cells such as hepatocytes[35].

Epithelial and immune cells including monocytes, B-cells and mast cells can all be infected by DENV and subsequently produce cytokines in response to this infection. One study demonstrated that serum concentrations of TNF-α, interleukin (IL)-2, IL-6 and IFN-γ were most elevated within the first 3 days of illness whilst IL-10, IL-5 and IL-4 increased afterwards[64]. Due to the immune effects of these cytokine groups, it is therefore implied that Th1 responses are seen during the first 3 days with Th2 responses occurring later.

Differences in cytokine levels have been observed between severe and mild DENV infections with increased levels of IL-13 and IL-18 correlated with more severe forms of disease in infected individuals[65]. This correlation has also been found with transforming growth factor-β (TGF-β) which is an inhibitor of Th1 and enhancer of Th2 responses, whilst the inverse relationship appears to happen with IL-12[65]. Together these findings suggest that a dominant Th2 response occurs in DHF/DSS whilst dominant Th1 responses are more associated with protective responses.

Antibody dependant enhancement

Early on it was noticed that DENV infection can present in different ways and that the more severe haemorrhagic manifestation of the disease occurred more frequently in individuals experiencing a secondary infection[66]. It was additionally noted that individuals who received the DENV vaccine without prior exposure were significantly more likely to develop severe DENV infections in a subsequent infection[48].

During a DENV infection cross-reactive neutralising antibodies are produced which are weaker, less protective and shorter lasting. It is hypothesised that these antibodies can form immune complexes with DENV and can increase infectivity by activation of the FcγR [20,57]. Indeed monoclonal antibodies produced against DENV E protein have been shown to enhance infection when below the threshold for neutralisation[59].

A study into the cellular mechanisms and interactions behind ADE showed that antibody opsonised DENV has been shown to have unknown surface domains which co-ligate the inhibitory receptor LILRB1 when engaged by FcyR[58]. This leads to inhibition of FcyR signalling including production of interferon gamma. The presence of these proteins are essential for the occurrence of DENV ADE in some cell models[58,67]. ADE can therefore be replicated *in vitro* relatively easily with cell lines and non-neutralising titres or monoclonal antibodies. How well these replicate what is seen *in vivo* has been difficult to establish.

The lack of a suitable animal model, such as mouse, has made research into the *in vivo* mechanisms behind ADE in subsequent DENV infections difficult to study[45]. Although non-human primates can serve as natural hosts in a sylvatic setting and have a very similar immune response they fail to show clinical manifestations[68]. The lack of a licensed DENV treatment means that DENV human infection models have too great a risk for the participants to develop severe untreatable infection[69].

ADE is the proposed mechanism behind the failure of many candidate DENV vaccines and with the previously approved live attenuated vaccine Dengvaxia where in previously approved countries it is no longer recommended without initial screening for previous exposure to DENV[70].

1.3 The challenges associated with diagnosing current and historical Dengue infections

In order for DENV to be accurately diagnosed the patient must seek medical attention once clinical presentation of the disease has started. DENV infection which is severe enough to require such medical attention is rare, with 70-80% of infections being mild or asymptomatic[71]. There are a wide variety of symptoms for DENV which make accurate diagnosis based on symptomatic presentation difficult, with symptoms of infection manifesting from a mild generic fever to Dengue Fever (DF); an acute infection presenting clinically 4-10 days following the bite of a mosquito, to the severe illnesses of dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS)[2].

Laboratory diagnosis can be undertaken using several approaches, these target various biomarkers of infection and include aspects of the virus itself or the host immune response to viral infection.

Diagnosis of the viral component of infection is performed by isolating the virus in cell culture, RT-qPCR of viral RNA and antigen recognition assays usually focussed on conserved or genetically distinct sites of the virus such as the NS1 protein[72]. Serological diagnosis looks for a human immune response by detecting specific antibodies using ELISA-based techniques or in rare cases neutralising antibodies by plaque-reduction neutralisation tests (PRNT).

ELISA diagnostic tests can be used to infer whether a presenting case has a primary infection or a post-primary infection, based on the ratios of IgM to IgG antibodies. While relatively higher IgM levels indicate a recent primary infection compared to a post-primary infection, the level ratios change quickly during the course of the infection. This can be useful in explaining the reasons for symptoms observed with severe cases more often seen in secondary infections[73,74].

The various diagnostic techniques have their benefits and drawbacks depending on the time into infection when the assay is performed, and the number of times the individual has been infected, both affecting which test would work best (Figure 1.7).

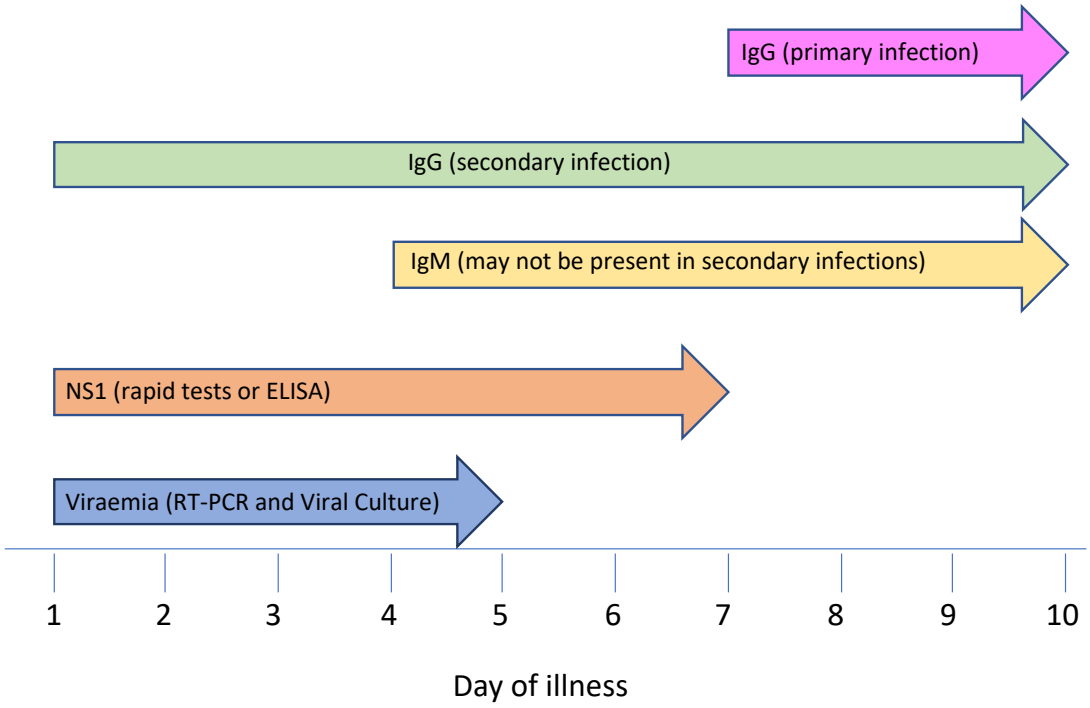


Figure 1.7: Laboratory diagnostic options based on viraemia and immune responses. Adapted from Simmons *et al.* 2012[34].

Antigen-based

Testing for viral antigens, especially those found in mucosal secretions, has become one of the most frequent techniques for confirming viral diagnosis due to the low skill requirement, low relative cost, and high speed of obtaining a result[75]. With viral infections including DENV, where viral loads are most easily detected in the circulatory system, instead of for example mucosal secretions, antigen tests are still used relatively rarely compared to antibody markers and nucleic acid techniques. This is likely due to the need for a phlebotomy sample reducing the benefit from unskilled testing, as well as a historical lack of accurate antigen tests available, with common antigens of DENV being highly cross-reactive to other flaviviruses. Despite this, there are several new antigen tests available which target the more viral specific NS1 protein and are increasingly being used to rapidly triage patients[76].

Serology

Determining DENV infection status using serology is made difficult by the high-degree of cross-reactivity between not only the 4 DENV serotypes but also other flaviviruses especially in areas where Zika virus is also found. Serology markers can be used to diagnose active DENV infection through the detection of IgM as well as historical infection through IgG detection. Before the wide-scale adoption of qPCR as a diagnostic tool plaque-reduction neutralisation assays (PRNT) were the only method of determining which serotypes an individual had been exposed to. These are still used but their reliability is highly dependent on convalescence time as well as number of previous DENV exposures.

Other serum and blood markers of infections which are used, especially to determine the risk of a patient developing a severe infection, are haematocrit levels and platelet counts. For example new guidelines issued by the Singaporean Ministry of Health employ an admission criteria in parallel with the WHO guidelines: these suggest clinicians should look for when a suspect DENV case has a rise in haematocrit ≥20% of the baseline levels or a drop in their platelet threshold to <50,000/mm^3, these

measures have helped focus medical attention on those most at risk especially during peak DENV seasons[4]

ELISA

The detection of serological markers of infection in a diagnostic environment are generally performed with ELISA based technologies. These platforms are easily scalable, can be automated and the data generated by them is reproduceable.

The high degree of cross-reactivity between *flaviviral* species can make accurate diagnosis with ELISAs difficult. To date no ELISA exists which can reliably differentiate between DENV serotype infection and other techniques must be used. Many commercial ELISAs have struggled to differentiate between DENV and ZIKV infections although some developed following the ZIKV epidemic of 2015 have shown good specificity[77,78].

ELISAs can be used to detect historical infection as well as active depending on the markers investigated. Antibody profiles in DENV infections are different depending on the infection number, with primary infections showing a classical pathogen immune response of IgM which is followed and eventually replaced by an IgG response.

IgM or antigen-detecting ELISAs are used in a clinical setting to detect active infection, however due to the delay in the production of antibodies, IgM assays are less than 50% sensitive within the first 4 days of a primary infection[79]. In post-primary infections IgM titres are significantly lower and can be undetectable in around 20% of cases[80]. The ratio of IgM to IgG can therefore be used to determine whether a patient is experiencing a primary or post-primary infection, with improved accuracy if the ratio cut-offs are changed by reported day of infection[73].

Plaque-reduction neutralisation test (PRNT)

PRNT assays used to be the preferred test for determining historical exposure to DENV and serotyping the causative virus[76]. They cannot be used in viraemic individuals and compared to newer assays they are significantly more expensive, take longer to yield a result and require relatively highly skilled operators to perform the labour-intensive protocols. Although they remain the only assay which can determine historical exposures to DENV serotypes, their reliability for determining this and subsequent inferred immunity to reinfection of this serotype is questionable[81], qPCR is significantly more reliable for use in active infections and ELISAs are now preferred for determining historical exposure.

Genomic

Reverse transcription quantitative polymerase chain reactions (RT-qPCR) have been developed for uses in diagnosing DENV, these can have a very high sensitivity and a specificity of 100% in acute patients who are still viraemic[76].

RT-qPCR can take advantage of multiple probe fluorophores to generate a multiplexed assay which can detect each serotype. The CDC have developed validated multiplex assays which can detect commonly circulating arboviruses[82] (DENV, ZIKV and Chikungunya virus) or individual DENV serotypes[83]. Although these assays are commonly used, particularly in the Americas, there is no global consensus on a suitable assay, additionally there are no WHO reference materials to standardise RT-qPCR assays between labs. An external quality assessment in 2009 of RT-qPCR assays in 37 laboratories across 27 countries showed that fewer than half were fit for diagnostic purposes with many missing moderate-to-low viraemic patient samples[84].

1.4 Current Dengue control measures and how their success can be determined

Currently, the primary control measures for DENV are aimed at reducing or removing insect vector populations. Countries vary in their approaches to tackling the vector, but generally focus on removing the breeding habitats, still bodies of fresh water, and spraying insecticide in response to high case numbers. The day biting habits of the primary vector reduces the effectiveness of insecticide-treated bed nets even if used by children napping during the day[85] and studies have shown little evidence of their effectiveness for use in control[86,87].

There are several biological control techniques which are being used to control DENV in endemic populations. These include "Sterile Insect Technique" which refers to the release of male mosquitoes, which have been previously sterilised through irradiation, this reduces the reproduction rate of female mosquitoes, as no viable offspring are produced, thereby lowering the vector density and subsequently reducing the transmission of associated vector-borne diseases including DENV.

There is additionally the relatively novel strategy of infecting *A. aegypti* with *Wolbachia pipientis* (subsequently referred to as Wolbachia). Wolbachia is a naturally occurring intracellular bacterium which infects many species of insect but has not been found to occur naturally in *A. aegypti*[88]. Experimentation has demonstrated that transinfection of *A. aegypti* with Wolbachia confers resistance to disseminated infection by DENV and other arboviruses[89,90]. Indeed, a field trial has shown that the release of Wolbachia infected mosquitoes into endemic DENV populations has shown a significant effect in decreasing the amount of virologically confirmed DENV presenting at hospitals, with the protective efficacy of the intervention being 77.1% (95% CI, 65.3 to 84.9)[91]. These studies have occurred relatively recently so their success at reducing DENV infection in the future still remains to be seen.

This highlights the difficulties of assessing the long-term impacts of DENV control strategies, especially as most studies are from relatively short-term trials and we can see huge natural fluctuations in yearly case numbers independently of any control

measures being in place. It is therefore possible that the effect of an intervention may overestimate the effectiveness of the same intervention applied in a higher transmission intensity context or underestimate the effects of the intervention in a lower transmission intensity setting. Studies often use patients presenting at care facilities as measures of case numbers and therefore any changes as a result of the intervention however, as expressed earlier, a significant number of DENV is asymptomatic and sub-clinical with most individuals not going to a health facility[92]. Additionally, there may be some differences in severity between DENV serotypes and sub-genotypes which may have an effect on the reporting population[93,94]. Although randomly sampling the whole population for those positive for the virus may assist with identifying asymptomatic infections, this is time-consuming, costly and may not yield many viraemic individuals[95]. A more reliable method may be serological testing a proportion before and after a study to look for changes in serostatus or evidence of recent infection by serological markers[96].

Techniques that directly or indirectly look for evidence of viral exposure can be reasonably effective at measuring how effective a control strategy has been in reducing the presence of the virus in a population, however, if the virus reappears in the next year or later with the same serotype, it cannot identify whether it was a reintroduction or a failure of the control strategy to eliminate the virus from circulating within the population. Through genome sequencing of pathogens of interest it is possible to reconstruct their ancestral relationship and therefore determine how likely it is that they are persisting in a population or being reintroduced[97].

1.5 Challenges associated with generating sequencing data from viral infections

WGS at its most basic involves generating the full genetic code of an organism. Produced genetic information can then inform us of key protein-coding genes as well as the amount of diversity that is available within a population. Traditional techniques such as Sanger sequencing are regarded as highly accurate but have several shortfalls

in their application to sequencing pathogen genomes especially during an outbreak situation.

Pathogen genomes were sequenced by generating ~2kb amplicons of the genome by PCR from isolated patient samples which may have to have been grown in culture to increase the pathogen load to a required amount. These were then subsequently sequenced by Sanger sequencing methods[98]. Although sufficiently accurate for pathogen identification and basic analysis the technology is unable to provide the deep coverage required for variant detection. Additionally, identifying novel pathogen genomes is incredibly difficult as prior knowledge is required to design sufficiently specific amplifying primers. The primers themselves and amplifying techniques can further introduce bias and make metagenomic analysis difficult.

Next-generation sequencing platforms resolve many of the issues of this older technology[99]. They can combine short reads with extremely deep coverage in order to determine low-frequency within host variants[100,101] as well as capturing fragmented or partially degraded samples. The high throughput additionally allows large numbers of viral sequences to be analysed in a relatively short time for a lower cost.

Preparing viral genomes for whole-genome sequencing is primarily performed using three methods: PCR amplicon sequencing, metagenomic sequencing and target enrichment[102]. Whereas metagenomic approaches sequence all nucleic acid within a sample, they require high very viral titres to provide sufficient coverage and will generally yield very high background sequence data (which may or may not be useful). The main goal of the other two methods is to produce or purify enough viral nucleic acid to be distinct from the background contamination such as from human serum in blood-borne viral isolates. PCR amplicon sequencing is similar to older Sanger techniques, however, running the generated amplicons on next-generation sequencing instruments generates considerably more data and enables variant detection. In this study one of the techniques used was a modified version of the PCR

amplicon techniques whereby multiple 400-500bp overlapping primers were used as demonstrated in Figure 1.8.

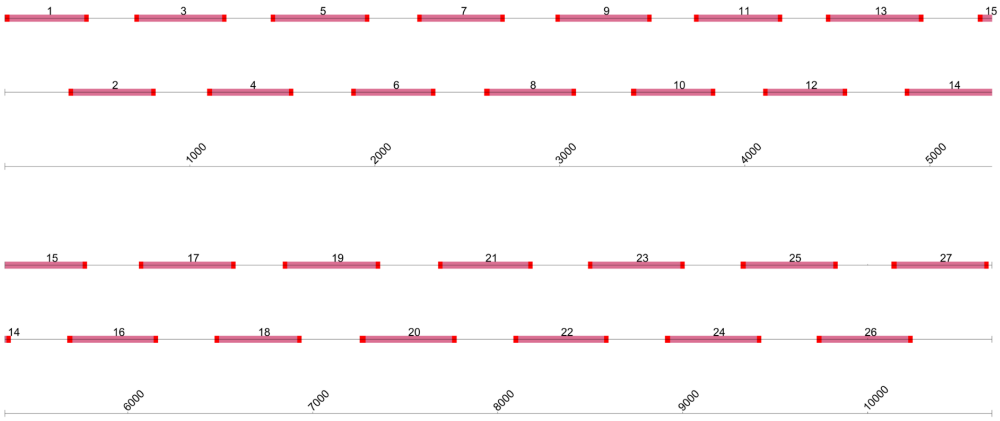


Figure 1.8: Example of 500 basepair (bp) overlapping primer scheme across a DENV genome.

These overlapping primers mean lower titres of virus can be sequenced and primer binding sites (which introduce bias) can be removed computationally after sequencing. Additionally, whereas each amplicon would have to be read separately with Sanger techniques, next-generation sequencing platforms allow for multiplexing therefore amplicons can be pooled together.

Target enrichment techniques can sequence viral genomes without the need for culture or a large number of PCR cycles. This method utilises small RNA or DNA probes that have been designed to target the virus of interest. An advantage of this method is that many target viruses (or other suspect pathogens) can be included in the probes which can enable sequencing of pathogens which may be unable to be accurately diagnosed; co-infections can additionally be detected if probes have been designed accordingly. The hybridization reaction utilises probes which are bound to a solid phase (generally streptavidin-labelled magnetic beads) and then are used to purify complementary DNA from the target sample.

There have been very few studies comparing the results from different viral sequencing approaches in detail. A paper sequencing Hepatitis C virus with the three different methods described previously confirmed that PCR methods generated

greater read depth than the other two but resulted in more incomplete genomes for highly variable strains. Metagenomic and target enrichment approaches meanwhile were extremely limited by viral copy number with coverage directly proportional[103].

1.6 How viral sequencing has the potential to enhance our understanding of transmission

Whole genome sequencing of viruses from clinical samples is vital for gaining knowledge of the phylogenetics, transmission and evolution of a virus [104,105]. Epidemiological modelling is used to understand the spread and mechanism of transmission of a viral outbreak. These methods have historically had to rely on contact tracing and mathematical modelling techniques to inform researchers on how a virus spreads[106,107].

Phylogenetics is the study of evolutionary relationships between biological entities. Traditionally these relationships were generated using morphological similarities between organisms however with the development of Sanger sequencing and PCR we were able to use genetic information and subsequent similarities to develop phylogenetic analysis[108].

Recent advances in sequencing technologies and reductions in costs have enabled almost real-time monitoring of viral genomes from infected patients as demonstrated during the Ebola outbreak in 2015[109] as well in vector-borne diseases demonstrated by a Zika virus outbreak in Singapore[110]. Such techniques have enabled the identification of drug-resistance strains and virulence factors[111]. Viral sequencing has been especially useful during influenza outbreaks for determining likely outbreak source, for example during the 1997 H1N1 influenza outbreak in Hong Kong, phylogenetic analysis showed that the virus most likely arose from a reassortments between an H5N1 virus in poultry and a similar strain found in Quail[112]. This subsequently led to changes in legislation in the sale of live poultry.

There is a wide range of methods for phylogenetic reconstructions, the most popular (and accurate) of which are maximum likelihood or Bayesian inference[113,114]. These methods use aligned consensus sequence information from genomes of interest and apply a nucleotide substitution model to determine the evolutionary distance between them. Start date and cause of an outbreak can be determined using a phylogenetic tree with sufficient and appropriate sampling date information. This enables branch lengths to be converted to time based on the assumption that nucleotide substitutions accumulate at a relatively constant rate, hence exhibiting a 'clocklike' behaviour', in reference to the molecular clock theory[115]. Sufficient sequence data obtained during a viral outbreak can therefore be used to infer the likely origin of an outbreak and track its spread amongst a population more accurately than just using contact tracing methods[116,117].

The ongoing 2019 coronavirus (SARS-COV2) outbreak is the most sequenced disease outbreak in history with over 10 million genomes having been sequenced and made publicly available as of April 2022[118]. This information has been invaluable in generating vaccines, therapeutics as well as identifying variants which are more virulent or resistant to available treatment options[119]. Additionally, genome sequencing has been used to help characterise transmission dynamics of the pathogen at various resolutions from across the world down to a small healthcare setting[120,121]. As well as generating relevant data from the pathogen, whole genome sequencing of patients infected with pathogens of interest has revealed host factors which may influence disease outcomes[122]. Patient total mRNA sequencing (RNA-Seq) has been undertaken on DENV infected individuals with the hope that an host marker which is different in severe DENV cases can be identified and a potential targeted treatment can be established[123,124].

Phylogenetics as a tool to study infectious diseases

A hallmark of nucleic acid-based life is that this genetic code changes with each successive generation in a process known as evolution. The rate at which this genetic code changes is determined by several internal factors such as the fidelity of the

proteins tasked with copying it or external ones generally related to physical and chemical environmental factors such as UV radiation. These mutations may cause a beneficial, neutral or negative change in fitness, which is determined by the ability of derived lineage to proliferate thereby increasing its population. Most viruses have high rates of evolution which are attributed to their large population sizes, short generation times and relatively high mutation rates. Single-stranded RNA viruses including DENV have some of the highest mutation rates of all viruses which is explained partly by their use of RNA-dependent RNA polymerase (RdRp) which has no proofreading ability[125].

The genomic relationship between lineages related by a common ancestor is termed phylogeny, with phylogenetic reconstruction being the quantification of differences accumulated between lineages. Reconstructing the phylogenetics of viruses enables us to investigate the relationships, putative origins, spread and selection pressures that are applied over time[104,108]. These have implications for understanding how the virus can be effectively controlled.

Methods to characterise transmission of a pathogen

Identifying and characterising how pathogens spread from host to host and subsequently inferring who infected whom can all help towards its eventual control by attempting to disrupt and prevent future routes of transmission. The means by which a pathogen can get from an infected host to susceptible ones generally determine how easily the transmission pathways of a pathogen can be determined. Diseases such as HIV and HCV require sexual contact or direct exchange of bodily fluids through intravenous drugs, whilst for arboviruses including DENV there is an insect vector which is needed to transmit the virus between hosts which can create an unknown transmission pathway.

Contact tracing is commonly used with many diseases to reconstruct disease transmission pathways[126]. These most frequently take the form of questionnaire-based studies whereby people note down their contacts or their locations visited

during a set period of time. These aim to infer transmission between individuals based upon previous contact which may be known or inferred from their location history. The effectiveness of such techniques are highly dependent upon the pathogen of interest with greater accuracy associated with diseases which are spread via close contact and which can persist for long periods of time, such as HIV or other sexually transmitted diseases, and less reliable with pathogens spread by aerosols such as influenza, especially when case numbers are high[107,126]. There are limited studies which have investigated the usefulness of contact tracing with vector-borne diseases, especially those transmitted by mosquitoes, owing to this added step which can significantly increase the distance which a pathogen can spread beyond just human-dependant factors[116].

Through WGS data, the phylogeny of pathogens will contain information of the historical transmission of that pathogen in a given population, which would be impossible to gain through other methods. Many studies using genomic information have focussed on the phylodynamics of a pathogen in a large population, generating information on the likely geographic origin of a newly detected strain[104,127,128], but some have been able to narrow this down to a local and even pairwise resolution[129,130]. However, a key limitation of using phylogenetic tools to infer transmission is the requirement for sufficient genetic diversity at the epidemiological timescales, which are being investigated. RNA viruses are generally strong candidates for this kind of inference owing to their high mutations rates[131]. Another limitation is the difficulty to infer transmission directionality from phylogenies, in which a cluster of sequences can indicate recent transmission but not 'who infected whom'.

1.7 Thesis Aims

In this thesis I describe 3 distinct chapters that aim to investigate the immunological and virological factors associated with DENV transmission and expand our understanding of how DENV persists in an endemic population with seasonal fluctuations and investigate whether we can utilise genomic and serological data to explore its transmission characteristics.

Objective 1. To develop an experimental protocol for generating whole-genome sequence data for DENV from isolated patient data (Chapter 2).

i) Develop a method for generating complete RNA viral genome data with sufficient coverage that is suitable for applying to large numbers of patient samples

Objective 2. To investigate the immunological and virological factors that may predispose patients to severe infection and characterise sub-clinical infections (Chapter 3).

i) Investigate correlations between virological responses and disease outcome
ii) Determine whether underlying immune status has any implications for disease outcome
iii) Investigate viral titres in sub-clinical individuals to determine their potential for transmission of the virus

Objective 3. Investigate whether genomic information can be used to determine the time period that endemic DENV strains have persisted in a population (Chapter 4).

i) Using phylogenetic reconstruction of the available viral diversity determine how long different sub-populations have persisted in an endemic population.
ii) Characterise the virus in an endemic population by its putative country of origin to determine the amount of introduction

Objective 4. Characterise the transmission of DENV within an endemic population at both a whole population scale and within households containing an identified infected individual (Chapter 5).

i) Using genomic data and available epidemiological information about enrolled participants generate putative transmission pairs.

ii) Using serological data identify potential exposures to the same virus and their likely locations

iii) Combining genomic and serological data determine the likelihood of identified transmission pairs

Chapter 2 Development and Review of Next-Generation RNA viral Sequencing Methods

2.1 Introduction

Viral whole genome sequencing (WGS) strategies are becoming more common in both an epidemic and a routine diagnostic setting as the technology becomes cheaper, faster and the data generated more useful. Positive sense RNA viruses are the causative agents behind many viral infectious diseases of significant concern around the globe including HIV and the SARS-COV2 pandemic. As DENV also has a genome with a positive sense RNA strand, sequencing strategies are applicable between viruses within Baltimore classifications IV and VI.

WGS with next-generation techniques can help improve are understanding of viral infections and has several practical used for treatment, for example with viruses for which anti-viral drugs and therapeutics exist, as even the prevalence of a small sub-population of drug-resistant viruses can have implications for the success of treatment options such as with HIV and HCV[102]. Although with DENV at least one treatment option has shown promise, at least for prophylaxis and lowering viraemia, and is currently in early clinical trial phases[42], there remain no direct anti-DENV drug options for treating an infection[132].

WGS enables researcher to investigate several other important virological factors of infection, many of these are in their relative infancy with more research required to validate the results for example it is increasingly being used to reconstruct putative transmission events and outbreak locations, which is not possible with traditional sub-genomic sequencing techniques. This is because often the variability in single viral genes is too low, even in highly variable RNA viruses, to reliably conduct phylogenetic analysis[133]. More frequently, WGS is used extensively in the to identify introduction events for example during the SARS-COV2 outbreak in 2020[119], additionally to identify novel strains which may escape vaccine protection[134] and was extensively used in mapping the spread of the Zika and Ebola outbreaks[109,135].

There are several challenges which must be overcome when performing WGS on an RNA virus, especially one isolated from a clinical sample. High-throughput sequencing instruments such as the Illumina Miseq and the Ion Torrent require DNA as an input, which necessitates the conversion of RNA genomes into DNA using a reverse-transcriptase enzyme which can have a high error rate[136]. Fortunately, these instruments are also commonly used for mRNA-seq experiments, so often these protocols can be easily applied to RNA virus sequencing. One of the main issues with RNA viral sequencing is therefore overcoming the high background RNA material, which even in a highly viraemic individual is often over 99% of the RNA reads[102]. Although metagenomic strategies can be employed which will sequence all material non-specifically, as viral reads are so proportionately low, full viral genome coverage may not be possible to achieve.

One of the strategies which are most commonly used to overcome this background RNA material is PCR amplicon sequencing. In a similar vein to Sanger sequencing strategies, large amplicons of the viral genome can be amplified by PCR using specific primers. Products are often 1-2kb in size so for a small viral genome like DENV only 5 amplicons can be needed[99]. Produced amplicons must be subsequently pooled in equimolar amounts and fragmented to smaller sizes for analysis on an Illumina instrument such as the Miseq.

A similar strategy to this uses separate overlapping pools of small amplicons which require less downstream processing. The simplicity of this strategy has enabled its use in resource poor environments such as in a mobile laboratory in the Amazon rainforest during the ZIKV outbreak. It can also be easily scaled up so is currently the most used technique in the UK for COVID-19 sequencing[137].

These PCR-based strategies can relatively quickly and cheaply enrich a sample with target viral nucleic acid; however, they have several limitations. Primers must be designed to a representative reference sequence and especially in new or diverse

outbreaks, primer drop-off may occur as viral genomes mutate. This requires primer schemes to be continuously updated or viruses to be partially genotyped before sequencing. The requirement of already knowing the viral sequence additionally limits such strategies during a new viral outbreak due to issues with primer design accuracy. Many PCR cycles on a low copy-number starting template can cause PCR-induced point mutations from lack of proofreading. This can result in an incorrect sequence being generated for a viral target.

Capture based target enrichment protocols utilise thousands of DNA or RNA probes which are complementary to the viral genome. This technique requires prior knowledge of the target genome and works best when there is a large repertoire of viral genome variants available[138]. The biotinylated probes are generally 80-120bp in length and bind across the whole length of the genome, enabling the purification of bound fragmented genomes. An overview of these sequencing approaches is given in Figure 2.1.

Figure 2.1: Overview of different sequencing approaches employed during study

2.1.1 Chapter Aims

This chapter will be a summary of the main sequencing strategies employed during the laboratory work component of this thesis and will discuss the benefits and drawbacks of each technique as applied to whole genome sequencing of non-segmented, positive sense, single-stranded RNA viral genomes.

i) Develop an in-house sequencing strategy that produces full-length viral genomes with a read depth of at least 100x across the whole genome.
ii) Develop a sequencing strategy which can enable the efficient whole genome sequencing of hundreds of clinical viral samples isolated during the study.
iii) Conduct an assessment of the various sequencing strategies

Contributions

The team based at the Institute Pasteur in Nha Trang, Vietnam coordinated sample collection as well as viral RNA extraction. RNA extracts were sent to LSHTM for qPCR and whole genome sequencing. 18 of these sequences were initially generated using a large amplicon sequencing strategy developed by Sonal Shah with assistance from the Genome Institute of Singapore. Subsequent sequencing techniques for the remaining 386 sequences were developed by me. 96 viral sequences were generated by Renee Chang, an MSc summer project student using these protocols. Data was analysed using a pipeline developed by myself with assistance from Neneh Salah and Jody Phelan.

2.2 Methods of Sequencing Approaches Investigated

2.2.1 qPCR Diagnosis of DENV

Initial optimisation steps of the various qPCRs used were performed on 4 DENV serotypes donated from cultured virus by the National university of Singapore and a combined 4-plex control from the CDC.

2.2.1.1 Pan-DENV qPCR

The Roche Tib-Molbiol Dengue Diagnostic qPCR kit was used with the NEB Luna rt-qPCR kit for identification of samples which were positive for DENV. The manufacturers protocol was followed with the run conditions for the NEB polymerase kit used according to Table 2.1.

Table 2.1: PCR cycling conditions for NEB polymerase kit

Temperature	**Time**
55°C	10 minutes
95°C	1 minute
95°C	10 seconds
60°C	30 seconds
Steps 3 & 4 repeated for a total of 45 **cycles**	
4°C	∞

2.2.1.2 qPCR Serotyping of DENV

The CDC real-time RT-PCR assay was used to determine the serotype of DENV in the infected individuals. The protocol used was based on Santiago 2013[83] with minor adjustments. Primers and probes used are listed in the reference protocol. The Quantifast multiplex qPCR kit (QIAGEN) was substituted for the recommended kit with the run conditions according to Table 2.2.

Table 2.2: Quantifast multiplex qPCR kit run conditions

Temperature	**Time**
50°C	20 minutes
95°C	5 minutes
95°C	15 seconds
60°C	30 seconds
Steps 3 & 4 repeated for a total of 45 **cycles**	
4°C	∞

2.2.2 Large Amplicon sequencing approach

Background

16 DENV-1 consensus sequences had been generated previously in the group using a large amplicon approach. The technique used for this was based on Aw 2014[99]. I generated 3 sequences using this approach as part of a laboratory induction. Sequences for this technique had previously been serotyped using the CDC serotyping qPCR assay.

cDNA preparation and PCR

Serotyped D1 samples were first converted to cDNA using the Maxima H Minus First Strand Synthesis Kit (Thermo Scientific) according to the manufacturer's protocol and using the final D1 reverse primer at the 3' end of the viral genome from Aw *et al.* 2014[99].

Prepared cDNA was then run in 7 PCR reactions to generate amplicons which spanned the whole DENV genome (Table 2.3). PfuUltra II Fusion HS DNA Polymerase was used (Agilent) with the manufacturer's protocol followed and the PCR cycling conditions used in Table 2.4.

Table 2.3: List of primers used in large amplicon sequencing protocol

Primer Name	Sequence
D1f1F	GTTAGTCTACGTGGACCGAC
D1f1R	CATCGTGATAGGAGCAGGTG
D1f3F4	TCACAAGAAGGAGCAATGCACA
D1f2R2	AAGAAGAACTTCTCTGGATGTTA
D1f5F	ACCAATGTTTGCTGTAGGGC
D1f5R	TATTCCCCGTCTATTGCTGC
D1f7F	CAGAGCAACGCAGTTATCCA§
D1f7R	CAATTTAGCGGTTCCTCTCG
D1f9F	TCACAGATCCTCTTGATGCG
D1f9R	CATGGCACCACTATTTCCCT
D1f10F	ATGGCTCACAGGAAACCAAC
D1f10R	TGCCTGGAATGATGCTGTAG

Table 2.4: PCR cycling conditions for large amplicon generation

Activation	2 min at 92°C
40 cycles	10s 92°C
	20s at 55°C
	1.5 mins at 68°C
Final extension	5 mins at 68°C

Library Preparation for sequencing

Individual amplicons from each PCR were processed to remove unincorporated primers and nucleotides using Agencourt AMPure XP (Cat # A63881, BeckmanCoulter). The DNA was purified with the ratio of 1.8 ratio of beads to DNA volume. The amplicon concentration was determined by measuring its 260nm absorbance with a Nanodrop ND-1000 spectrophotometer (Nanodrop, Wilmington, DE). The required volume of each amplicon from one sample was calculated to match the lowest concentration amplicon and amplicons from one sample were pooled in a single tube. The DNA concentration of amplicon pool was 500ng/µl and the volume of amplicon pool for library preparation was 35µl.

Amplicons were sheared enzymatically first into smaller fragments using the QIAseq FX DNA library (cat #180479, Qiagen). The fragmentation incubation time for 2kb and 1kb amplicons were 3 and 2.5 minutes, respectively. The end-repaired and an 'A' adding to the 3' ends of the fragmented DNA were performed immediately. The DNA fragments ready for adapter ligation was made. Each amplicon pool was labelled by dual-barcoded by combining one D5 barcode and one D7 barcode in each ready-to-use adapter. After library amplification, the reaction clean-up and removal of adapter-dimers were achieved by using Agencourt AMPure XP (Cat # A63881, Beckman Coulter). The library size was assessed using High Sensitivity DNA Kit (Cat # 5067-4626, Agilent) and the concentration of each library was measured using the Qubit dsDNA High Sensitivity Assay Kit (Cat # Q32851, ThermoFisher). The libraries were adjusted to 4nM and pooled into 1 tube. Denaturation of the 4nM library pool was performed using fresh NaOH 1N 0.2N and the library was diluted to 10pM with HT1 buffer (Miseq Reagent kit v2, Illumina technologies). Phix control was spiked into the final denatured 8pM library at a final concentration of 5% prior to loading on the Illumina Miseq instrument.

2.2.3 Small Overlapping Amplicons

Background

For samples which were positive by the serotyping assay (serotype 1 and 2), an amplicon sequencing approach was used based on the protocol developed by Josh Quick for Zika virus[139]. This approach is also commonly used for other viruses such as SARS-COV19, HIV and Influenza with resulting amplicons suitable for both Illumina and Oxford Nanopore technologies[138].

Primer Design

Multiple 400bp amplicons were designed using the primal scheme software (available at http://primal.zibraproject.org/) which is based on primal3. 10 whole genome sequences from temporally and spatially close outbreaks were used to design the sequencing primers. To ensure the maximum amount of diversity within the viruses identified as being within the population was captured for primer design, identified sequences were selected for using a 10% redundancy threshold on JalView[140].

cDNA synthesis and PCR

Isolated RNA, which was positive by serotyping qPCR, was converted into cDNA using the SSIV VILO Master Mix (ThermoFisher) according to the manufacturers protocol. DENV cDNA was then run in the amplicon generation PCR using the two pools of primers and Q5 High-Fidelity DNA polymerase (New England Biolabs) which produced 400bp overlapping amplicons of the whole genome according to the conditions listed in Table 2.5 Table 2.6.

Component	Volume in 25 µL reaction
Q5 2x Master Mix	12.5 µL
Primer pool (#1 or #2)	1 µL
Nuclease-free water	9 µL
cDNA	2.5 µL

Table 2.5: PCR reaction conditions for amplicon generation

Temperature	Time
98°C	30 seconds
95°C	15 seconds
65°C	5 minutes
Steps 2 & 3 repeated for a total of 35 cycles	
4°C	∞

Table 2.6: PCR cycling conditions for amplicon generation

Following completion of the amplicon PCR the resulting products were purified using a 1.8:1 bead to sample ratio of KAPA Pure beads (Roche – KK8000). Purified PCR product was quantified using the HS DNA Qubit assay (ThermoFisher) and a correct size was confirmed by agarose gel.

50ng of each pool were combined together for library preparation using the KAPA Hyper prep kit according to the manufacturers protocol but with ¼ of the recommended reagents. Following library preparation product was quantified again using the Qubit assay and size distribution was confirmed by Bioanalyzer assay. A maximum of 48 samples were pooled together for sequencing by Illumina Miseq.

2.2.4 Target Enrichment

Background

The relatively small number of available D4 sequences on GenBank, especially from a similar time and space as the study location, and the poor sensitivity of the serotyping qPCR assay, meant that a large number of qPCR positive samples were not able to be sequenced using either of the previous techniques which had been implemented. For this reason, I explored using a target enrichment approach which would not require knowing the serotype and could make use of all the available sequences for designing capture probes thus minimising the chances of primers not recognising the target sequences.

RNA bait design and synthesis

The candidate capture bait sets were designed by *CATCH* (compact aggregation of targets for comprehensive hybridisation, v1.3.0[141]) to generate 120-mer RNA baits that cover the entire DENV genome and encompass the genetic diversity of DENV1 (n=2446), DENV2 (n=1891), DENV3 (n=1121) and DENV4 (n=339) on the National Centre for Biotechnology Information (NCBI) GenBank [142]. The input sequences were aligned by multiple sequence alignment program *MAFFT* (multiple alignment using fast fourier transform, v7.480[143]); the custom parameters were used to allow the baits to tolerate up to 2nt mismatches within a bait and at an interval of 60nt between baits in the viral genome. The designed biotinylated oligonucleotides (baits) for DENV1 (n=416), DENV2 (n=736), DENV3 (n=648) and DENV4 (n=339) were synthesised by Agilent Technologies for hybridisation and capture of DENV in samples.

cDNA synthesis and Library preparation

Depleted RNA samples underwent fragmentation, first- and second-strand cDNA synthesis using NEBNext® Ultra™ II Non-Directional RNA First Strand Synthesis Module (New England Biolabs, Ipswich, MA, USA) and NEBNext® Ultra™ II Non-Directional RNA Second Strand Synthesis Module (New England Biolabs) in a Veriti™

96-Well Thermal Cycler (Thermal Fisher Scientific, Waltham, MA, USA) according to manufacturer's instructions for each. Briefly, 5µL of RNA was first fragmented and primed in 4µL of first-strand reaction buffer and 1µL of random primer at 94°C for 10 minutes, followed by first strand cDNA synthesis reaction in 8µL of nuclease-free water and 2µL of first-strand enzyme mix and incubated at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes. Second-strand cDNA synthesis was performed by incubating 20µL of first-strand cDNA with 48µL of nuclease-free water, 8µL of second-strand buffer and 4µL of second-strand enzyme mix at 16°C for 1 hour.

The purification of double stranded cDNA was performed using 1.8:1 ratio of Agencourt AMPure XP beads (Beckman Coulter, Inc., Brea, CA, USA) to volume of sample, followed by two washes of beads with 80% EtOH and eluted in 53µL of nuclease-free water. The quantitation of purified double-stranded cDNA was carried out using Qubit™ dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific) prior to library preparation.

For sequencing library preparation, the SureSelect^{XT} Low Input Library Preparation Kit (Agilent Technologies, Santa Clara, CA, USA) was used. Briefly, the cDNA samples underwent end-repair, dA-tailing, ligation of unique barcoded adaptors, 14 cycles PCR amplification of adaptor-ligated library and purification of amplified library according to SureSelect^{XT} Low Input Target Enrichment System for Illumina Multiplexed Sequencing Platforms protocol (Version E1, April 2021); the modifications to the existing protocols were based on the application note by Williams et al. 2019 [144]. Purification step was performed with AMPure beads after each step to isolate libraries and remove short fragments. The quantity and validation for appropriate size of purified library were performed on a Qubit™ fluorometer (Thermo Fisher Scientific) prior to target enrichment.

Hybridisation, capture and post-capture amplification of DENV sequencing libraries

The hybridisation and enrichment of DENV in samples were performed using SureSelect^{XT} Low Input Target Enrichment Kit ILM Hyb Module (Agilent Tech.) according to SureSelect^{XT} Low Input Target Enrichment System for Illumina Multiplexed Sequencing Platforms protocol (Version E1, April 2021) with modifications recommended for pathogen sequencing[144].

The prepared indexed and adaptor-ligated libraries were hybridised with the custom designed bait set. Briefly, DNA libraries were incubated with 5µL of Low Input Blocker Mix at 95°C for 5 minutes followed by 65°C for 10 minutes. The DNA + Blocker sample was mixed with 25% RNase Block solution, 1:10 diluted capture bait library, hybridisation buffer and nuclease-free water to make up to a 30µL reaction. The assembled hybridisation reactions were then incubated as followed: 60 cycles of 65°C for 1 minute and 37°C for 3 seconds followed by overnight incubation at 21°C.

For capture procedure per each 30µL hybridisation reaction, the hybridised DNA libraries were bound to washed and re-suspended Dynabeads^{TM} MyOne^{TM} Streptavidin T1 magnetic beads (Thermo Fisher Scientific) in SureSelect Binding Buffer (Agilent Tech.) and mixed at 1400rpm for 30 minutes at room temperature. After binding, the beads were washed one time at room temperature with Wash Buffer 1 (Agilent Tech.) and six times at 70°C with Wash Buffer 2 (Agilent Tech.). After washing, the beads-bounded DNA-libraries were resuspended in 50µL of nuclease-free water and subjected to post-capture PCR amplifications. The cycling conditions were as followed: 98°C for 1 minute; followed by 22 cycles of 98°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute; and a final extension at 72°C for 5 minutes. The streptavidin beads were removed after the PCR and PCR products were purified with AMPure XP beads and eluted in nuclease-free water. The fragment size distribution of post-capture libraries was assessed by 2100 Bioanalyzer System with High Sensitivity DNA Kit (Agilent Technologies) and concentrations were quantified using NEBNext Library Quant Kit for Illumina (E7630, New England Biolabs) according to manufacturer's instructions. Each sample library was normalised to 4nM in TE

buffer (Invitrogen) for pooling and the concentration of the final pooled library sets were determined by an additional qPCR reaction prior to sequencing.

2.2.5 Modified RNA-seq technique using the Peregrine system

Background

The Peregrine method was established to overcome costs and time pressures associated with these previous techniques, I used a technique developed for mRNA-seq[145] but modified it for RNA viral sequencing whilst updating it with newer materials and the option of target enrichment for purifying viral sequences from background contamination.

The technique itself relies on a property of the reverse transcriptase enzyme MMLV to add three untemplated cytosines (C) to the end of the template RNA. These unmatched nucleotides can subsequently act as a template for second strand synthesis in the presence of a complementary sequence with the addition of primers containing tags for downstream processing (

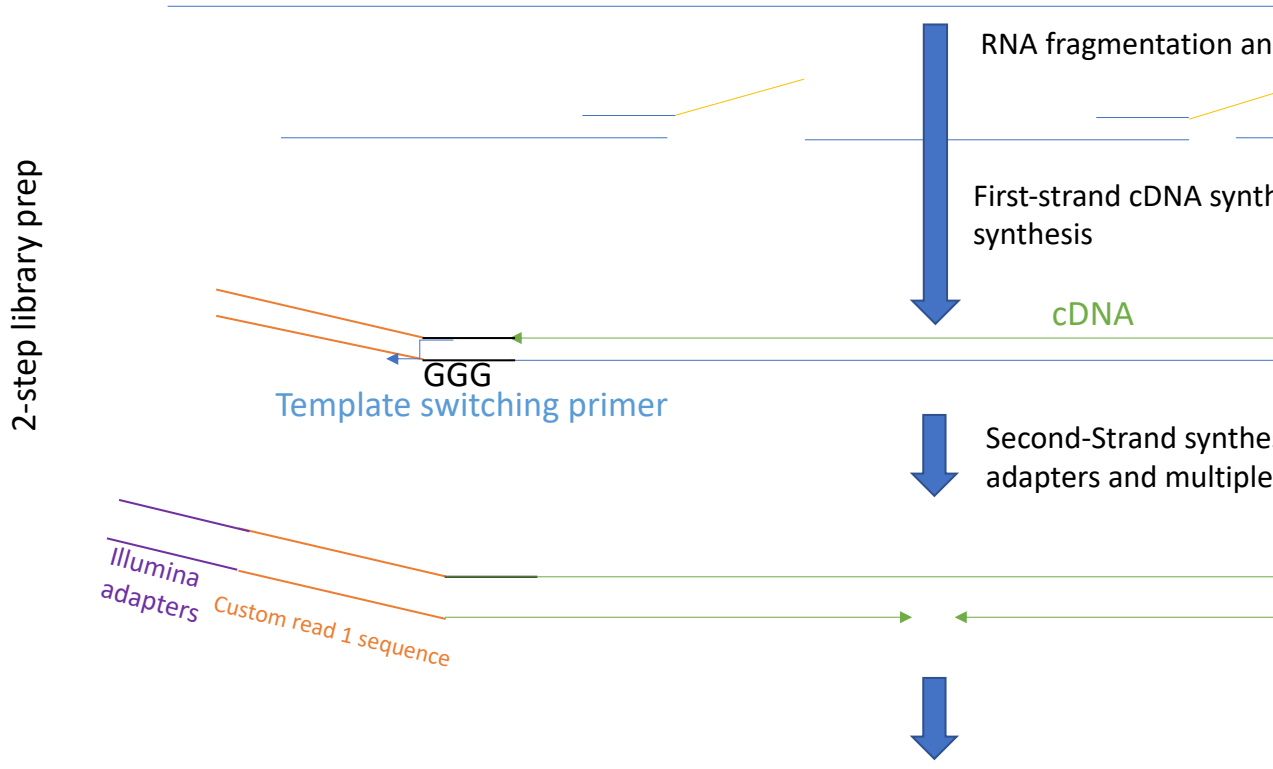


Figure 2.2).

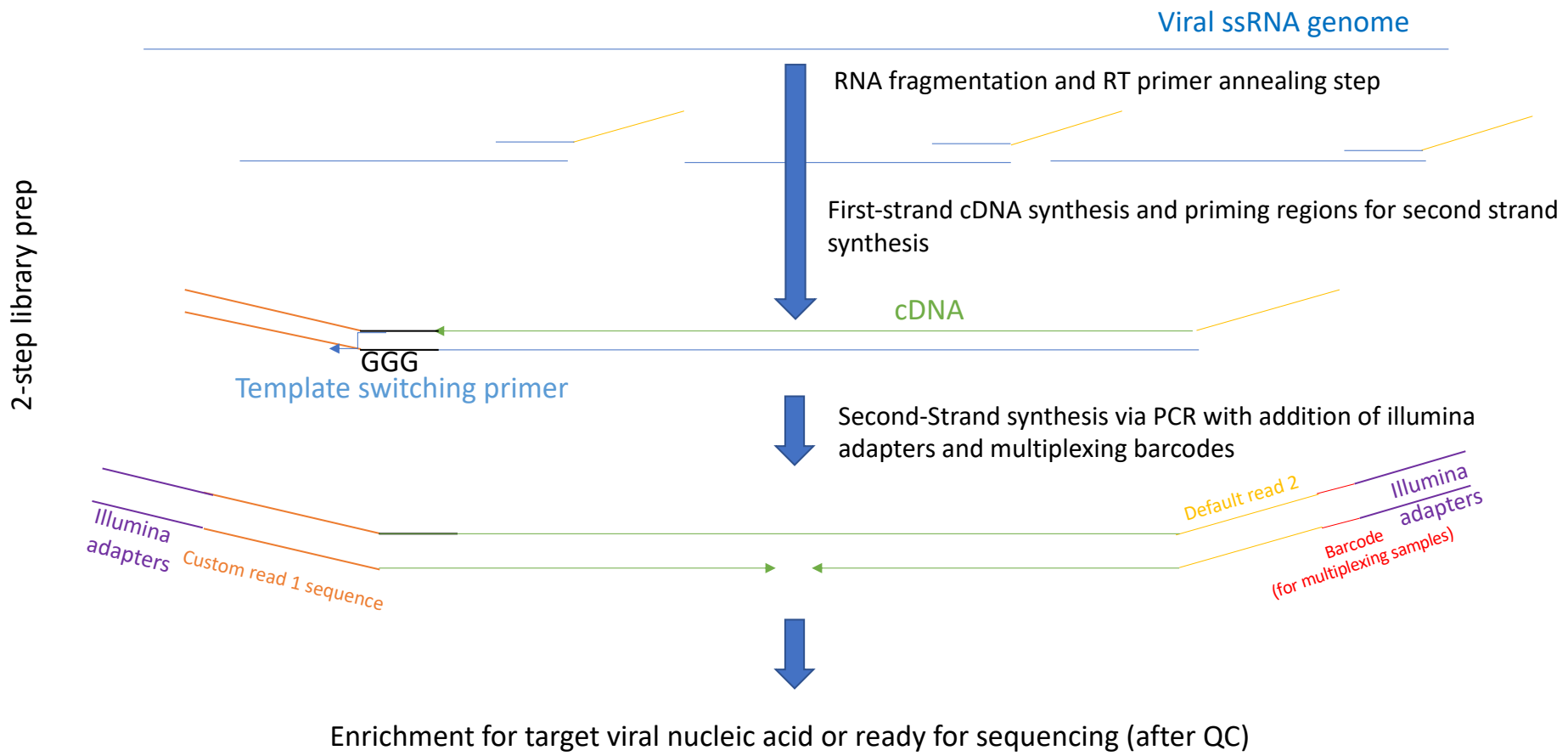


Figure 2.2: Overview of Peregrine library preparation process

cDNA synthesis, fragmentation and initial tag addition

First strand synthesis of target RNA samples was done using the NEBNext® Ultra™ II Non-Directional RNA First Strand Synthesis Module (New England Biolabs, Ipswich, MA, USA). 5µL of RNA was first fragmented and primed in 4µL of first-strand reaction buffer and 1µL of 25µM reverse transcription primer (Table 2.7) at 94°C for 10 minutes to obtain average fragments of 400bp. 10µl of fragmented and primed RNA from the previous step was then added to a reaction mixture containing 8µl water, 2µl NEBNext First Strand Synthesis Enzyme and 1µl of 12µM Template switching primer (Table 2.7). This was then incubated at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes. A purification step was then performed as before using a 1.8:1 ratio of AMPure beads.

Purified reverse-transcribed product was subsequently amplified using the PCR binding sites added during the reverse-transcription step. This was performed using the QC forward and reverse primers in Table 2.7 and according to Table 2.8.

Target enrichment of Viral Sequences

Target enrichment protocol used in 2.2.4 was followed.

Addition of indexing and multiplex barcodes by PCR

Following target enrichment protocol, a final PCR was run to add indexing barcodes and Illumina adapters necessary for sequencing. Custom universal forward and Barcode reverse primers from Table 2.7 were made into a master mix with polymerase according to Table 2.8 with the protocol from Table 2.9 followed.

Resulting PCR product was purified following the standard AMPure beads purification protocol.

Table 2.7: Primer sequences used in Peregrine library preparation technique

Primer name	Sequence
Reverse transcription primer	CAGACGTGTGCTCTTCCGATCTNNNNNN
Template switching primer	CAGGACGCTGTTCCGTTCTauggg (lower-case letters indicating ribonucleotides)
Custom universal forward	AATGATACGGCGACCACCGAGATCTACACTTCGCTACAGGACGCTGTTCCGTTCTATGGG
Barcoding reverse	CAAGCAGAAGACGGCATACGAGAT-*BARCODE-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (*unique 6nt barcode sequence)
QC forward	CAGGACGCTGTTCCGTTCTATGGG
QC reverse	CAGACGTGTGCTCTTCCGATCT

Table 2.8: Reaction mixture for Peregrine PCR

Reagent	**Volume**
Q5® High-Fidelity 2X Master Mix	20 µl
Custom forward indexing primer (10µM)	1µl
Individual index/barcoding reverse primer (10µM)	1µl
Water	8µl
cDNA from rt reaction	10µl

Table 2.9: Peregrine PCR protocol

Step	**Temperature**	**Time**
Initial denaturation	98C	30 seconds
qPCR result Cycles (10-25)	98C	10 secs
	65C	75 secs
Final extension	65C	5 minutes
Hold	4-10C	

2.3 Illumina sequencing and resulting consensus generation pipeline

Libraries prepared using all methods were normalised to 4nM in TE buffer then run on the Illumina Miseq instrument at 10pM final concentration following denaturation. Libraries were run using the MiSeq Reagent kits v2 with 500 cycles used for the amplicon-based techniques and 300 or 500 used for the target enrichment strategies. Demultiplexing and base calling was performed using the onboard Illumina software with resulting Fastq files analysed using a developed bioinformatics pipeline.

Consensus DENV full genomes of the target viruses were generated from the FASTQ files produced by Illumina MiSeq sequencing. Raw files were screened for quality and trimmed to remove primer sequence bias using Fastp[146]. Cleaned FASTQ files were then aligned to a serotype-specific reference genome using the Burrow-Wheeler Alignment for short-reads (BWA-MEM)[147] along with the SAMtools view and sort packages[148]. Reference DENV genomes were extracted from GenBank: DENV1: JQ045626, DENV2: GU131898 and DENV4:NC_002640. Aligned reads were called using SAMtools mpileup and iVar consensus[149] programs with a resulting consensus genome sequence in FASTA format.

2.4 Results

2.4.1 qPCR results

Extracted RNA samples which had been diagnosed as positive by the antigen detecting NS1 rapid test assay, were tested by both pan-DENV and serotype specific qPCR assays **(**Table 2.10**)**. The results of this testing showed that the pan-DENV assay had significantly greater sensitivity**,** with positivity dropping from 83.9% compared to 65.9% with the serotyping assay **($p < 0.005$)**. This is shown in Figure 2.3 where we can see qPCR results which were negative on the serotyping assay, were still positive by the pan-DENV assay.

Overall D2 was the most frequent serotype observed with 247 identified, followed by D1 and D4 with 205 and 95 respectively, 4 D3 samples were additionally detected throughout the whole time of the study, indicating that all 4 serotypes were in circulation at the study location.

Further investigation of the required viraemia for a positive serotyping assay result showed that successfully serotyped samples had a significantly lower pan-DENV ct (indicating higher viraemia) than those which were not able to be serotyped. Logistic regression comparing the ct of the pan assay with the ability of it to be serotyped showed that any sample with a ct greater than 29 had a less than 80% chance of being successfully serotyped (Figure 2.4).

Table 2.10: Breakdown of qPCR results for Pan-DENV and Serotyping assays

	Pan-DENV qPCR assay	Serotyping Assay
D1		205
D2		247
D3		4
D4		95
Positive	701	551
Negative	135	285

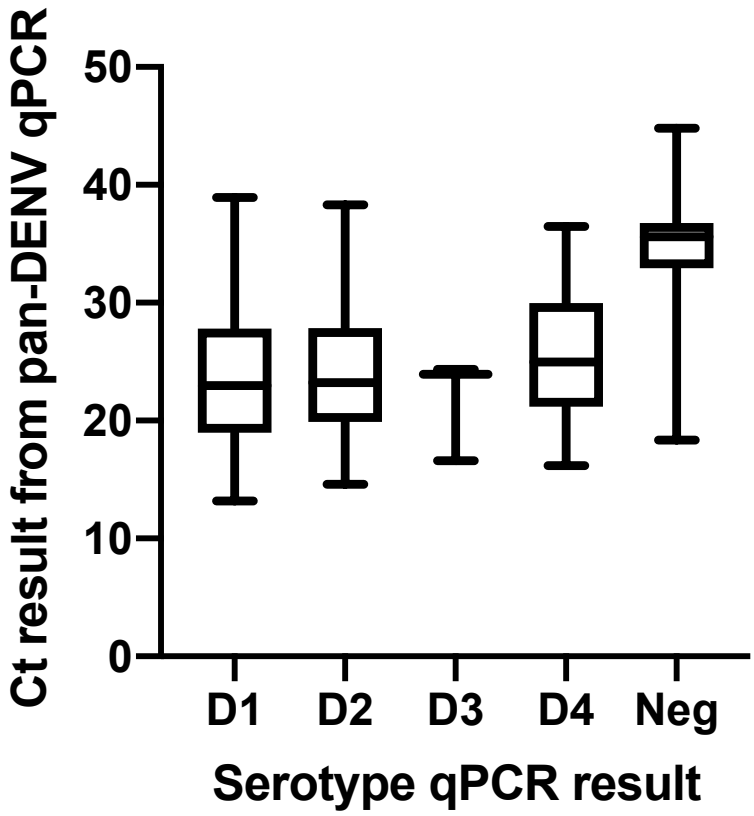


Figure 2.3: Box-and-whisker plot showing qPCR Ct result of Pan-DENV assay combined with serotyping qPCR result

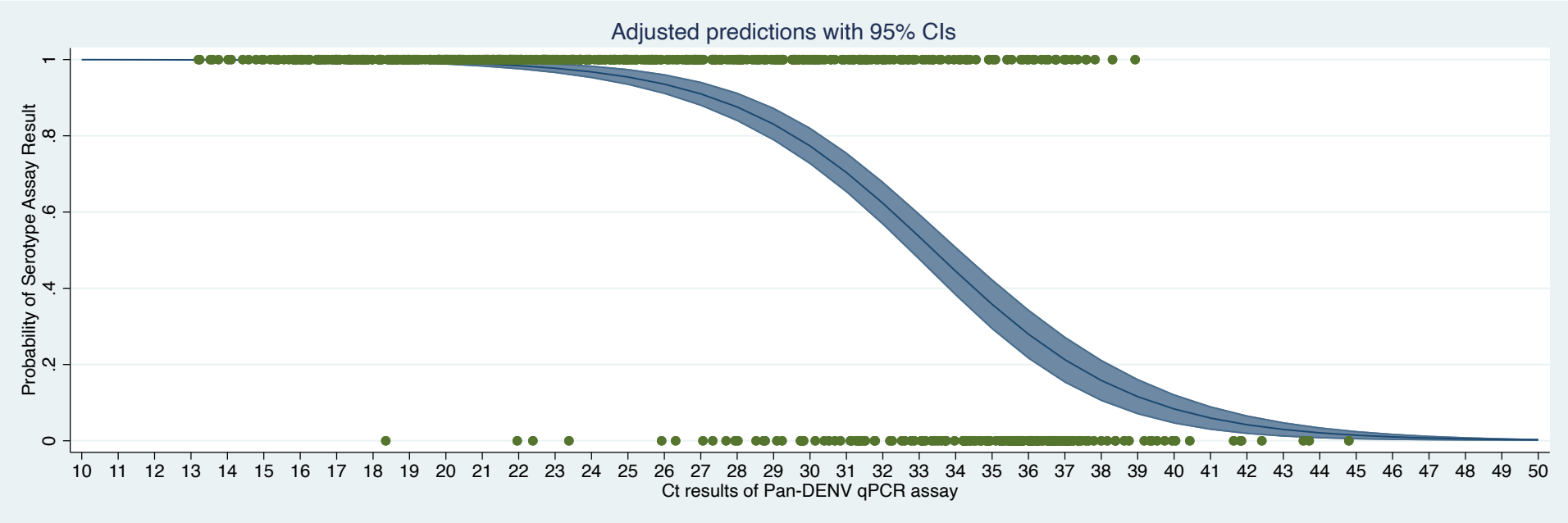


Figure 2.4: Logistic regression of probability of positive serotyping qPCR result based on the Pan-DENV qPCR value

2.4.2 Comparison of Coverage and sequencing depth

Across the 3 library preparation methods primarily employed to generate whole genome DENV sequences, there was an observable difference between the coverages and read depths obtained, a representative sample ran using the large and small amplicon as well as the target enrichment approaches is shown in Figure 2.5. Peaks in read depth can be observed where the primers are overlapping in Figure 2.5 A and B, with distinct peaks also visible in example C where the probes should be enriching for target sequences at 50nt intervals.

Genome coverage was generally sufficient to yield a consensus sequence for most of the sequences prepared, however, occasionally when using the amplicon approaches there were primer drop offs which necessitated repeat sequencing with revised primers or gap filling using sanger sequencing. This is apparent in Figure 2.5 example B where I saw regions with very low coverage for several of the primer pairs. This occurred more frequently with D2 which was likely due to 2 genotypes being present in the population.

Evenness of coverage is different between all the strategies employed with the large amplicon strategy, amplicons are manually equilibrated according to the molarity of the purified PCR reaction and subsequent fragmentation, in Figure 2.5 example A we can see clear differences between the amplicons which were pooled together indicating that they were not effectively balanced. Additionally with this technique, especially with the final 2 amplicons, I saw that the read depth dropped off as the read length increases, these were the 2 largest amplicons that were produced by PCR and fragmented so there may have been an issue with the fragmentation reaction time or PCR efficiency. The small amplicon technique had the most observable difference across the genome, as mentioned earlier there are both regions of high and very low coverage. The probes used in the target enrichment strategy were designed using a bioinformatics program that factored in regions with high diversity. I saw even coverage across the whole genome with all regions within the required number to call a base (x100).

Although the example given in Figure 2.5 shows the same D1 sequence run in all three methods, comparisons between them are made difficult by differences in coverage between runs as a result it being impossible to achieve the exact same read number in repeat runs on an Illumina instrument especially with different libraries. Unfortunately, as it is uneconomical to run the Illumina sequencer individually for the same sequence every time, comparisons between read depth are based solely on observable differences between these 3 runs. As the read depths between these runs were generally sufficient to call a consensus sequence, I consequently can make some conclusions on the efficacy in achieving a whole genome sequence out of each method. There are additionally practicalities of applying each method to large numbers of samples which as can be seen from Figure 2.1 are generally based on the number of steps required in their preparation and will be discussed further.

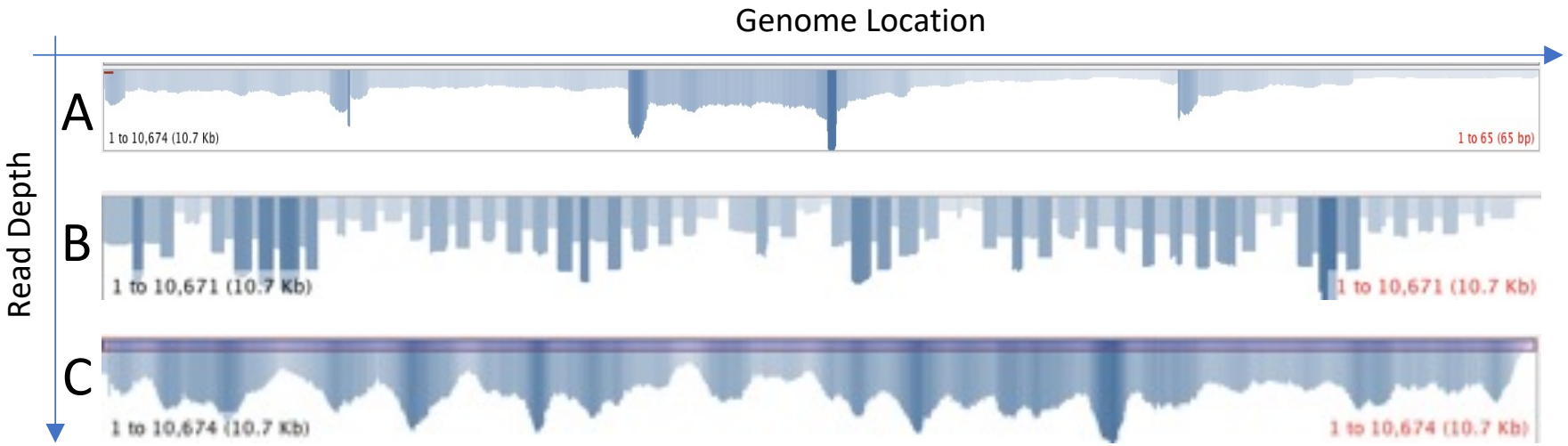


Figure 2.5: Distribution of sequencing reads and coverage for a control DENV sample across sequencing methods employed. A. Coverage distribution for Large Amplicon Technique. B. Coverage distribution for Small Amplicon Technique. C. Coverage for Target Enrichment technique.

2.4.3 Comparison of Viral reads

A subset of DENV positive samples prepared using each library preparation technique and one with only the rt step then adapter addition as a no-purification control were assessed for percentage of viral reads using the Genome Detective online program[17]. As viral RNA is normally only a very small percentage of final reads this measure informed how successful the different strategies were in isolated viral RNA from the background reads.

The results shown in Table 2.11 demonstrate that all the strategies were able to enrich the samples for their viral RNA contents with the amplicon-based PCR strategies proving to be more reliable at removing background reads compared to the target enrichment techniques. The occasionally lower reads in the target enrichment approach may be the result of poor quality or small fragments of viral nucleic acid being carried through to sequencing.

Table 2.11: Breakdown of target viral reads by sequencing strategy. Viral read mapping was calculated using the genome detective software[17] with 5 representative whole genome fastq files assessed in the software and the resulting viral reads given below.

Sequencing Strategy	% Target Viral Reads
No enrichment	<1%
Long amplicon	>99%
Short amplicon	>99%
Target Capture	80-99%

2.4.4 Successful library preparation

The majority of DENV sequences I generated were using the small amplicon and target enrichment techniques. All D4 and D3 sequences were generated using the target enrichment approach with the remaining serotypes prepared using all techniques. 31 samples for which a full-length sequence was not produced following amplicon sequencing strategies were repeated on the target enrichment approach.

In total 511 full length genome sequences were generated from 702 qPCR positive individuals. Library preparation was undertaken on 598 samples which were deemed to be suitably viraemic by qPCR (at least a ct of 30) and for the amplicon approaches it was required that they were successfully qPCR serotyped as well.

The success rate of each sequencing approach, excluding failed runs, is given in Figure 2.6. Here the small amplicon and target enrichment approaches had similar success rates at achieving full-length viral genomes with both having a success rate of 87%, however with the large amplicon this dropped to 48%. This is likely because larger amplicons require higher template ct/copy numbers for successful amplicon generation.

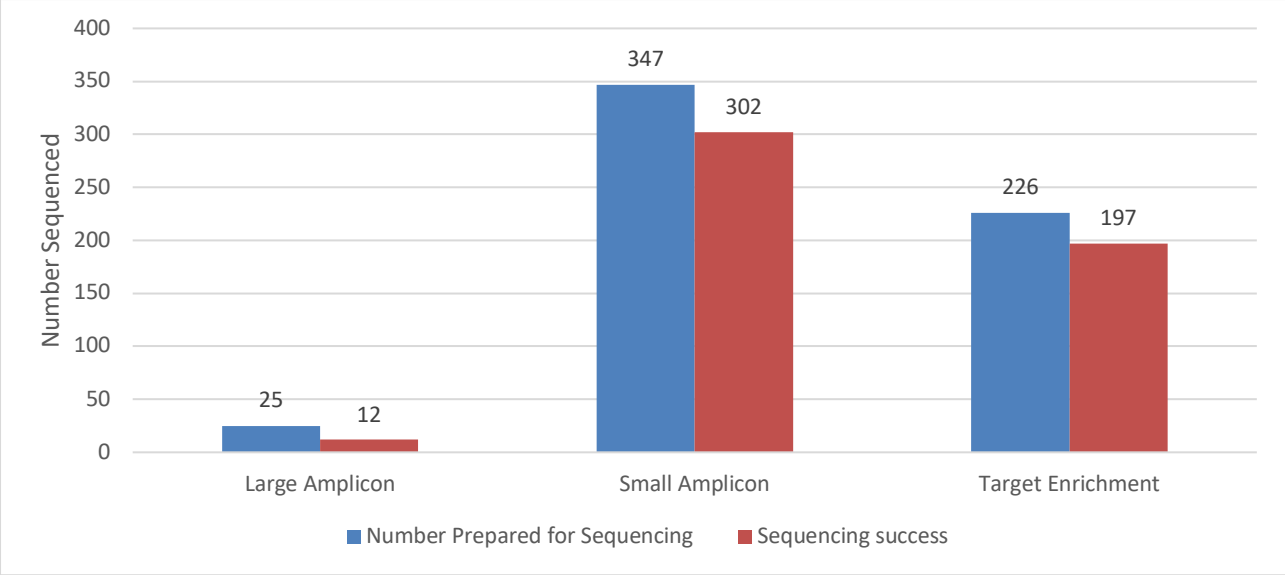


Figure 2.6: Sequencing success by library preparation strategy

2.5 Peregrine library sequencing analysis

An experiment was conducted to determine the success and potential of the Peregrine library preparation protocol followed. Tested samples were prepared with and without target enrichment from qPCR positive patient RNA extracts with ct values lower than 25.

Full length genome coverages were acquired for DENV with library preparation protocols with and without target enrichment (Figure 2.7). In this experiment samples were equilibrated to the same loading concentration on the flow cell, maximum read depths were significantly higher for the samples which was prepared with target enrichment (50 without target enrichment against 5371 for with), indicating a large number of wasted reads on non-viral DNA (99% viral vs <1% when analysed on genome detective). Interestingly the coverage maps overlap despite the added target enrichment step in one of the protocols, there is likely an issue with the peregrine step itself in the protocol in producing cDNA equally across the whole genome that requires further optimisation.

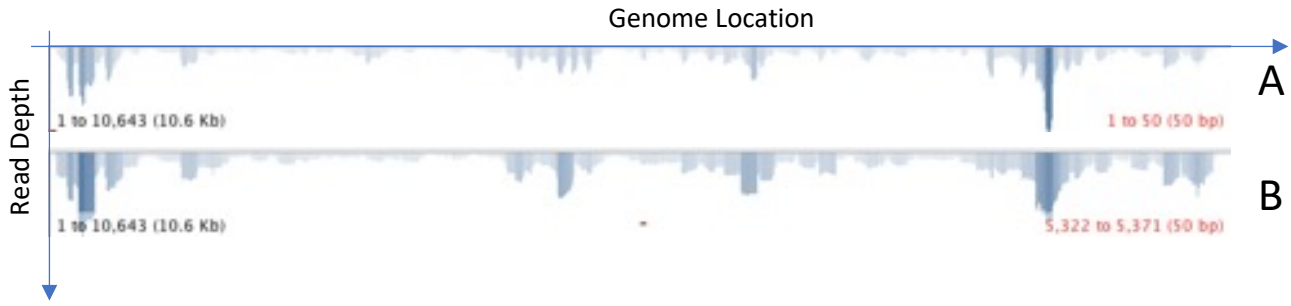


Figure 2.7: Read depth and coverage plots of Peregrine library preparation sequencing approach. A. Metagenomic approach, B. With target enrichment protocol.

2.6 Discussion

I successfully developed and optimised several diagnostic qPCR and sequencing library preparation approaches to generate both a way of comparing viraemia in samples and producing whole genome viral sequences. Previous sequencing strategies in my group have depended on working out the serotype of the target virus however this required using a serotyping qPCR which proved to have poor sensitivity when compared with the pan-DENV qPCR assay.

The serotyping assay used was based on the widely implemented CDC assay[83]. Assessment of the primer and probe schemes with genomes generated in this study showed strong similarity with all serotypes except for D4, where some genomes contained a mismatched base on the probe. However, despite this, a redesigned probe factoring in this difference had no difference in sensitivity or specificity in the assay. Future work could go into comparing the polymerases available as the one recommended by the CDC is too expensive for large-scale diagnostic purposes.

Comparisons between the success rates of the sequencing strategies showed similar efficacies between the short amplicon and target enrichment approaches with both having an over 87% success rate. This is likely due to more difficult samples being sequenced with the target enrichment approach however, as early on I found that full-length consensus sequences could be generated even from samples which had very high ct values (>30), additionally samples that had failed small amplicon sequencing were often repeated on the target enrichment approach. The failure of the small amplicon strategy to achieve full genome coverage with some D2 samples is likely due to the presence of multiple genotypes present in the population. Genotypes are sub-populations within serotypes that genomically are similar to each other but genetically distant from other genotypes[150]. Other sequencing studies using this technique have shown that primers schemes must be continuously updated as an outbreak progresses[151] or pathogens must be first sub-typed to determine their suitability for use with a specific primer scheme[152]. These problems are similar with the large amplicon approach, however that technique also generally required higher levels of viraemia in order for successful amplicon generation (ct<25).

The practical applications of applying these techniques to large numbers of clinical samples were evaluated by the ease of scaling up their protocols, with those with fewer steps and the ability to perform them in larger formats such as in 96-well plate instead of individual reaction tubes being preferential. The large amplicon protocol required additional quality control steps in order to ensure the amplicons were balanced and each amplicon had to be individually purified from its own PCR reaction. This made it difficult to scale up for the numbers which were needed in this study. The target enrichment protocol contained the greatest number of steps within its protocol, however they could all be accomplished with a multichannel pipette, the target enrichment part of the protocol itself was very labour intensive and the time limits of each of the steps meant that only 24 samples at a time could be run. The small amplicon protocol was the simplest to scale up to the large numbers that were involved in this study with multiple plates of samples able to be quickly prepared in a day, but unlike the target enrichment protocol, required a positive serotyping result and suffered from uneven coverage.

The small amplicon and target enrichment approaches could be optimised in further experiments to improve on their sequencing coverage and potentially improve sensitivity. This could be performed by redesigning the enrichment baits using the DENV genomes generated in this study, especially with D4 where enrichment probes were designed without a strong repertoire of locally circulating sequences. The primer pairs used in the small amplicon approach can be optimised to improve average coverage distribution, as has been shown with SARS-COVID19 strategies which utilise the technique[151], but this is often impractical to achieve with multiple genotypes and serotypes in circulation.

The modified Peregrine sequencing approach showed promise for a cheaper and faster method for generating libraries for RNA viral sequencing, however, further optimisation steps are required, especially to investigate the more significant variation in read depths across the whole genome that were observed.

Chapter 3 An investigation of the different virological and serological markers of infection in a DENV endemic population shows difficulties in identifying those at risk of a severe infection

3.1 Introduction

DENV is able to cause a wide variety of symptoms following infection of a susceptible host, with most clinical infections resulting in a self-limiting febrile illness. The severe disease phenotype, characterised by increased capillary permeability which can lead to dengue shock syndrome from plasma leakage, can have a mortality rate of 20% if untreated[34]. These symptoms are thought to be the interplay between the immune response of the host as well as age, genetic and virological factors such as increased pathogenicity of certain strains[153].

There are conflicting reports on differences between the 4 serotypes in their virulence and abilities to cause severe disease during a primary infection. A study in Singapore demonstrated that infection with D1 carried a higher risk of severe DENV when compared with D2[154]. However, a conflicting report in Brazil observed more severe cases among those infected with D2 than D1[155]. The difference in the human populations between these two studies may also suggest a role of how different human populations interact with DENV infections, studies have shown that the presence of different HLA domains may play a protective role against developing more severe forms of DENV infection[156,157].

The risk of developing DHF or DSS is greater in endemic regions where multiple serotypes are circulating[158] and was observed in naïve individuals vaccinated during clinical trials of the licensed Dengvaxia vaccine who were subsequently infected later, which led the WHO to recommended that it is only used in individuals who have experienced a prior DENV infection[159]. Both of these observations suggest that infection profiles and immune responses are different in primary and post-primary exposures.

ADE is the leading candidate explaining the proposed mechanism by which different DENV serotypes are able in infect individuals and also cause more severe disease[66]. During ADE DENV is able to infect immune cells with Fc receptors, despite this potential source of increased viral titre, there is evidence that viraemia is greater among individuals experiencing a primary infection and less during a post-primary infection[160,161] although this may be due to differences in the speed of the disease progression. There could additionally be differences in the time course an infection takes, with primary infections potentially lasting longer[161]. The amount of time an individual is viraemic and indeed their viral load, which is very difficult to estimate, are all factors which affect the probability of the disease being passed onto a vector and therefore transmitting to another host[162].

Although all 4 serotypes can circulate in an area, generally there are 1 or 2 dominant serotypes in circulation, with serotypes circulating in and out of the population every few years. This has led to speculation that serotype specific immunity may be a driver of specific viruses waxing and waning in a population[163]. However, there are a number of confounding factors with models demonstrating that disease fluctuations associated with ADE and antibody waning alone do not fully account for the temporal case fluctuations seen in case numbers[164]. DENV infection with one serotype appears to only confer relatively short-term antibody protection from reinfection from other serotypes, and while most are protected against the infecting serotype for at least 60 years, there have been cases of people being reinfected within 15 years [165], Although these are believed to be a- or sub-symptomatic, they are rarely captured in studies which investigate patients reporting at medical clinics. Despite this, there have been studies which have shown that asymptomatic individuals are able to transmit the virus to mosquitos[166] so homotypic reinfected individuals could be a source of the continued presence of serotypes seen in endemic populations.

DENV is generally associated with a very low mortality rate, which seldom goes above 1% of reporting cases in most populations, especially when suitable supportive care is performed[3,7]. There are documented differences between the pathogenicity of the serotypes and genotypes with the Southeast Asian genotype of serotype 2 being documented as causing significantly worse symptoms than the South American genotype in naïve individuals[167]. It is however difficult to suitably compare differences in viral pathogenesis as there are no

suitable laboratory or animal models and human populations typically have complex immunological memory that may impact on disease presentation.

DENV has been observed to cause different symptoms depending on the number of times infected with different serotypes with different outcomes based on order of DENV serotype infections[168]. The most severe form of infection, when the disease manifests with haemorrhagic symptoms which can lead to death if not managed effectively, are most frequently found in individuals who have previously experienced one previous DENV infection of another serotype and infants born to DENV immune mothers[66,169].

There have been several longitudinal studies which have investigated the differences in outcomes between which serotype caused the primary infection and the severity of symptoms following a secondary infection with a different serotype. A Cuban study showed that DENV-naïve individuals infected with D1 in 1977 experienced severe DENV infections 4 years later in 1981 and 20 years later in 1997 due to D2 introductions to the island[170]. These studies were increasingly performed following the Zika virus outbreak in the Americas in 2015/16 as outbreaks with more severe neuropathology were often associated with regions of high DENV endemicity[32]. Experimentation *in vitro* showed that ADE of human phagocytic cell lines occurred with antibodies raised against both ZIKV and DENV, indicating that enhancement was potentially possible in individuals initially infected with either virus[171].

As mentioned previously there are conflicting reports on the ability of asymptomatic infections in their ability to spread the virus to a mosquito host and whether there is suitable viraemia to enable onward transmission[92,95,172]. This is generally believed to be because viraemia has been observed to be significantly lower in these individuals [37,162]. The high number of asymptomatic individuals and the ability for humans to travel great distances could provide a significant reservoir for the virus to spread further than just through mosquitos alone. Capturing virus from asymptomatic individuals remains a challenge, owing to the short viraemia of infection, typically less than 10 days, and difficulties associated with repeated sampling of a blood-borne infection. Previous studies have attempted to sample people who are most at risk of being exposed such as in high-transmission areas during peak season and

close contacts of infected individuals[173,174]. This study has sampled household contacts of people presenting at enrolled medical facilities.

3.1.1 Chapter Aims

The principal aim of this chapter is to characterise the observable differences between infections of the circulating DENV viruses in the study population. I will therefore use virus serotyping data in combination with surveyed symptom data to investigate the following:

i. Describe the immune and virological responses following infection by different circulating DENV serotypes
ii. Investigate differences in both viral titres and disease progression based on evidence for previous exposure to DENV and symptomatic outcomes.

Contributions

The team based at the Institute Pasteur in Vietnam coordinated sample collection as well as viral RNA extraction and the serological diagnostic tests. The algorithm for determining primary or post-primary DENV infection using serological markers was developed by J. Biggs at LSHTM. All other analyses and experimentation unless stated were performed by me.

3.2 Methods

3.2.1 Study Population and Sample Collection

Individuals of all ages who resided in the communes of Vinh Hai, Vinh Phuoc, Vinh Tho, and Vinh Hoa in Nha Trang City, Vietnam, and who visited either a partner local Polyclinic or the city's Tropical Medicine Hospital were eligible for the study. In the polyclinic, those who presented with fever were eligible to be enrolled in the study and subsequently tested for dengue with an NS1 rapid test (DENGUE NS1 AG STRIP, #70700, BioRad). In the hospital, only those who were diagnosed with dengue fever on the basis of an NS1 rapid test as part of standard practice, were eligible to be enrolled in the study.

Enrolment was conducted between October 2016 and May 2019 at the polyclinic and between December 2016 and April 2019 at the hospital. Study teams then visited their given home addresses, where additional consenting household members were administered questionnaires and had blood samples taken at recruitment and again at an interval of 28 days.

3.2.2 Serological analysis

To characterise the presence and any changes to levels of IgM and IgG in the study population, samples were assayed using the Panbio® capture IgM and IgG ELISA kits (Cat. No.: 01PE10/01PE20, Alere, Brisbane, Australia). Briefly, kits encompass antigen capable of capturing host antibody specific to all four dengue serotypes and include plate-specific calibrators that normalise output optical density (OD) readings to generate standardised antibody panbio units. Pre-determined panbio unit serological thresholds categorised individuals as negative (IgM ≤ 9, IgG ≤ 18), equivocal (IgM 9–11, IgG 18–22) and positive (IgM ≥ 11, IgG ≥ 22) for dengue infections.

We used these thresholds to define participant seroconversions between visits. Primary infections were defined as negative to all immune markers at the first visit and then meeting the positive threshold to NS1, IgM or IgG at the second. We characterised those who were IgG positive at the first visit as historical and then those who seroconverted to post-primary

as an individual who was subsequently positive to IgM, NS1 or showed an increase of 2.2-fold in their IgG titre. Individuals could additionally remain either negative or historical without undergoing further seroconversion.

3.3 Results

3.3.1 Prevalence and distribution of DENV serotypes during the study

Patients presenting at an enrolling heath centre who tested positive for DENV NS1 through serological testing, were subsequently tested using two qPCR assays. The first was a non-serotype specific anti-DENV assay, chosen for sensitivity and comparability between serotypes. The second assay was used to serotype positive samples.

In total 701 of 838 NS1 +ve patients were also positive for DENV nucleic acid; and of these 551 were successfully serotyped by molecular sequencing, with positive results for all 4 serotypes (Figure 3.1A). The breakdown of DENV serotypes found showed similar numbers of D1 and D2 with 37.2% and 44.8% respectively, there were less D4 present, representing 17.2% of all positive tests; and only 0.7% of tests were positive for D3.

All 4 of these serotypes were present at some point in the study area during its duration from October 2016 to April 2019, with D2 replacing D1 as the dominant serotype by the end of the study (Figure 3.1B). There were also periods when all 4 were present at the same time. The serotype data alone indicates that the DENV season may peak around October-February every year. We can see a significant increase in all serotypes during October 2018-February 2019 which appears to be the result of a local outbreak event.

The study was performed in 4 communes within Nha Trang city itself. Using the serotype data generated, I subsequently investigated whether distinct clusters of cases and serotypes could be observed (Figure 3.2A). This resulting distribution shows cases were found throughout the study area with only limited evidence for close clustering of cases with all sequences and not by serotype (Figure 3.2B).

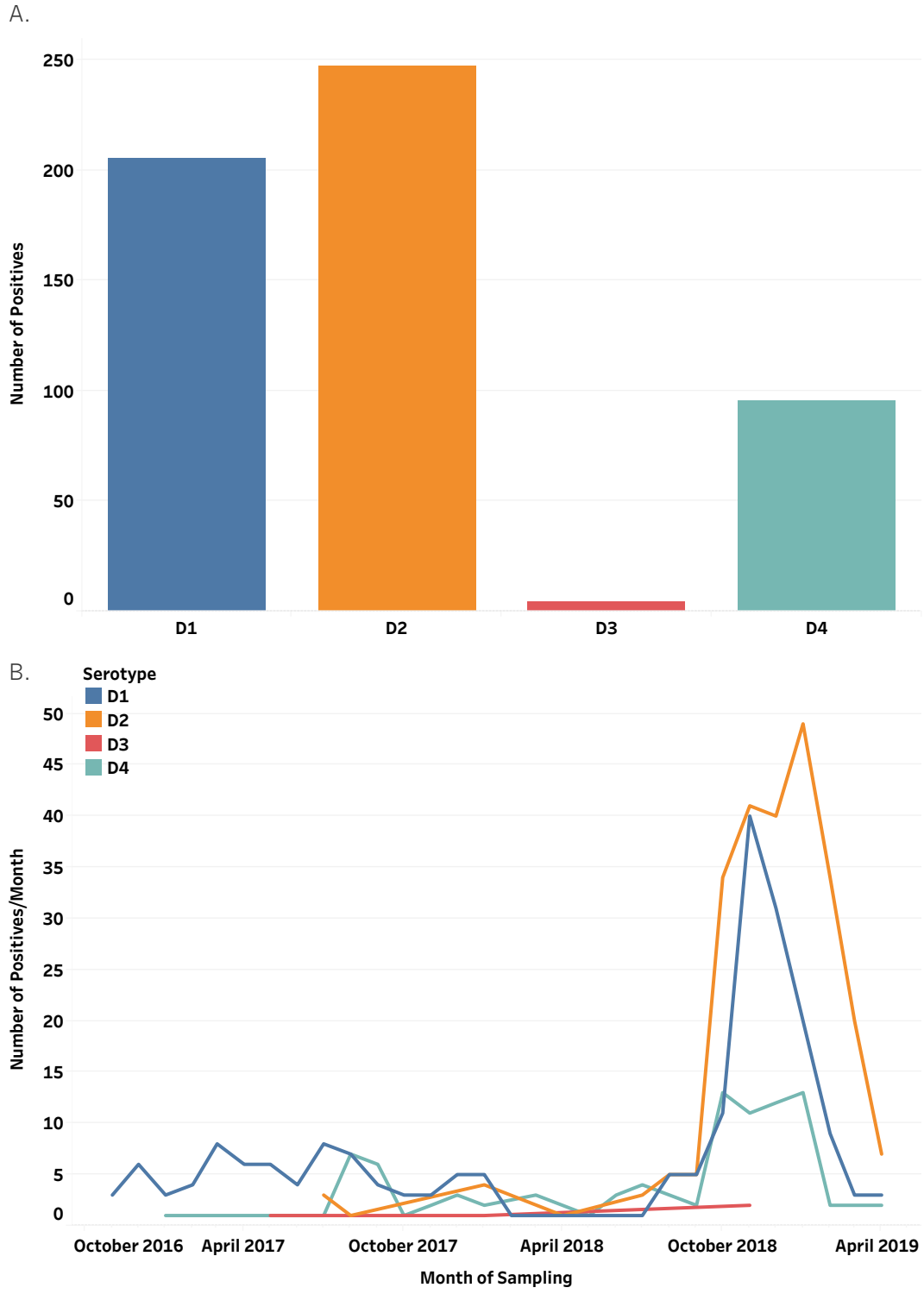


Figure 3.1 Distribution of dengue serotypes in the study area. A. Breakdown of positive serotyping qPCR results. B. Distribution of serotype prevalence by month over the study period.

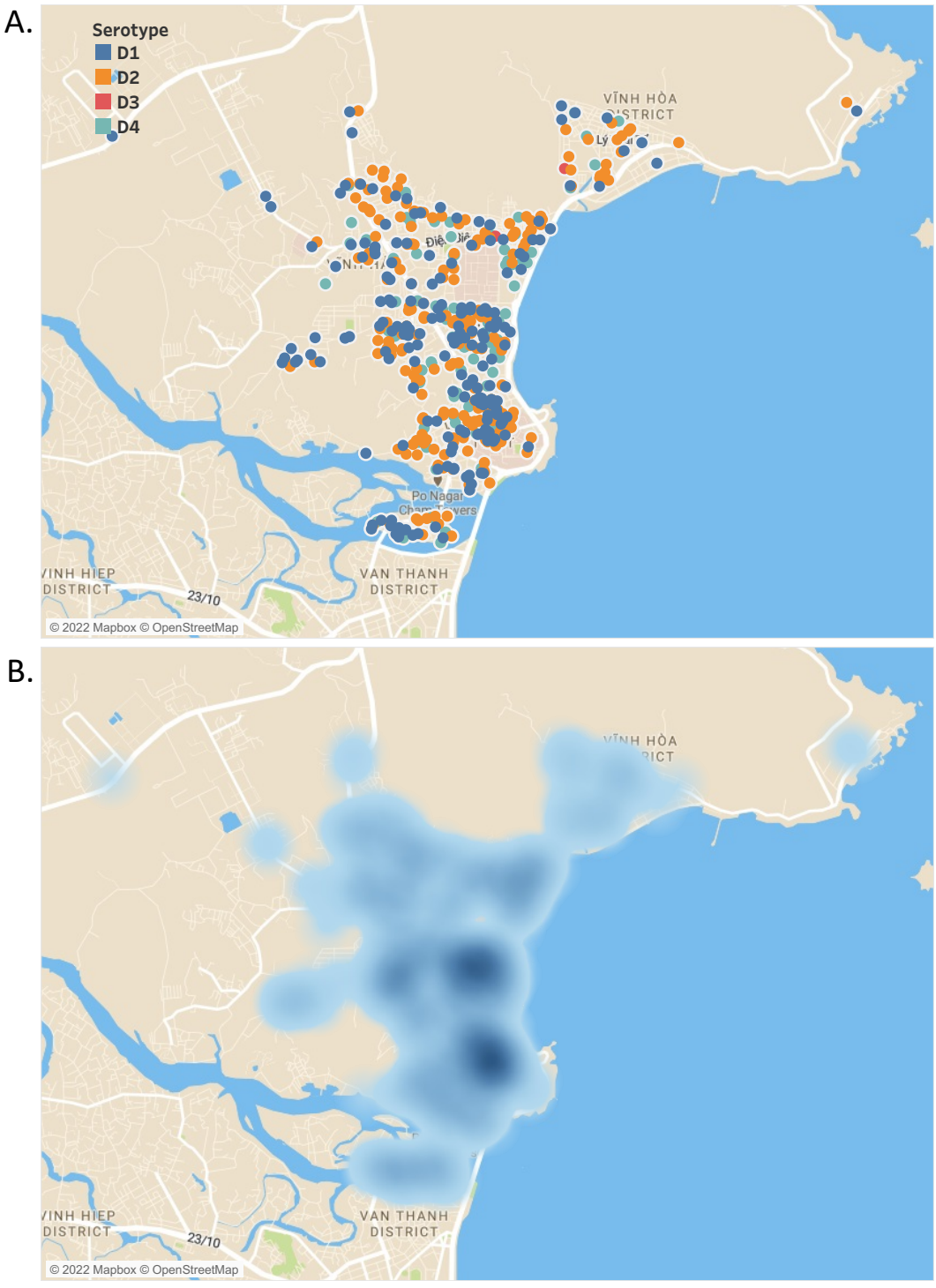


Figure 3.2 Map of study location within Vietnam, with location of serotyped subjects home address location plotted. A. Distribution of all samples which were successfully

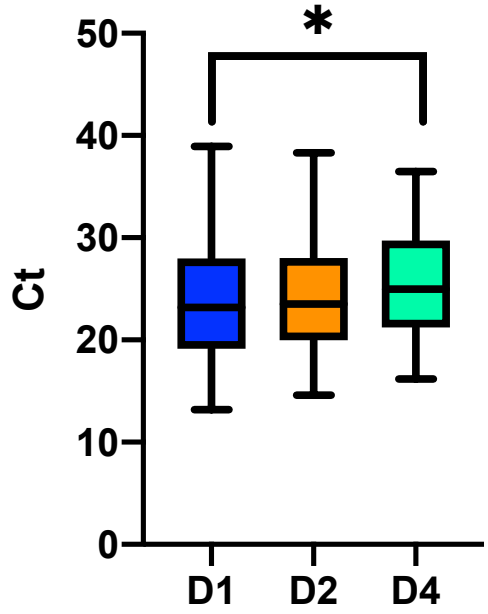
serotyped (n =516). B. Density plot of home addresses for all virus serotyped subjects, with darker colour indicating more samples within a 500m radius.

3.3.2 Observed viraemic differences between serotypes

In this study I was able to compare viraemia between subjects infected with DENV of three different serotypes, D1, D2 and D4. I initially compared average cycle threshold (ct) values between these serotypes using the qPCR results from the pan-DENV serotyping assay. The resulting average viral loads between serotypes showed limited differences with only D4 being significantly lower than D1 ($p=0.047$) (Figure 3.3A). With mean ct values for D1, D2 and D4 being 23.8, 24.4 and 25.5 respectively.

I therefore investigated how viraemia dropped according to self-reported disease day (Figure 3.3B). Overall viraemia changes over time appear to have a sigmoidal curve distribution, with individuals remaining at high viral load levels before quickly dropping and then a plateau was observed with significant variations in viral load by disease day. The viral serotypes appear to have a similar viraemia progression and the early stages of D4 infection are similar to the earlier observation of significantly lower titres.

A.



B.

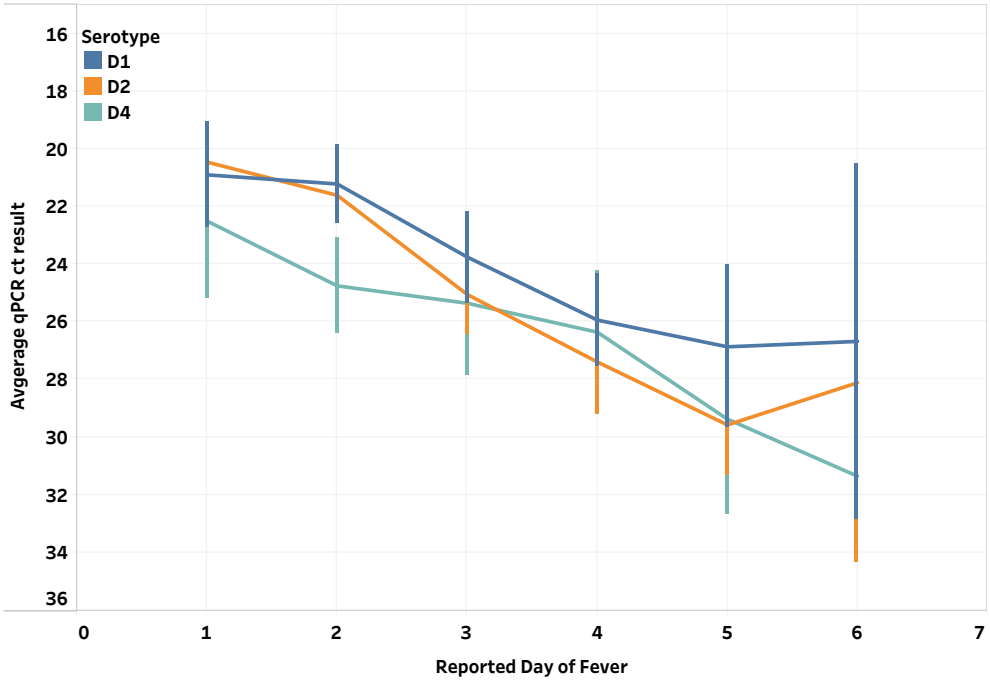


Figure 3.3 Viraemia differences between serotypes A. Box-and-Whisker Distribution of Ct levels between serotypes with range, median and upper and lower quartiles shown. B. Viraemia differences over time by self-reported day of infection with standard deviation at each day.

3.3.3 Antibody profiles of reporting population

To investigate the serotype specific variation of antibody responses to DENV, IgG and IgM were compared. In the 2935 individuals who were tested for the presence of IgG and IgM, seropositivity was 72.6% and 15.3% respectively. Overall seropositivity, where individuals were positive to either test, was 73.4% indicating that there was a high degree of prior exposure to DENV among the study population. All antibody response measures were highly variable across individuals as shown by the wide kit unit range and both displayed a skew to the left (Figure 3.4). Kit units for anti-DENV IgG ranged from 0.4 to 104.7 (mean $\pm$ S.D = 19.4 $\pm$ 25.9) (Figure 3.4A) all individuals with a value greater than 2.2 were considered positive for exposure to DENV. Kit units for anti-DENV IgM displayed a similar range from 0 to 107.3 (mean $\pm$ S.D = 8.9 $\pm$ 19.5) (Figure 3.4B), all individuals with a value greater than 9.9 were considered positive for recent DENV exposure with this assay (as previously described, ref)

I subsequently investigated the seroprevalence changes by age and used logistic regression to determine the likelihood of being exposed as age increases (Figure 3.4C). This showed that there is a high amount of seroconversion within the population at a young age with 10-year-olds being 54.5% likely to have already been exposed (95% CI 46.2-62.8) and by the age of 20 there is a 79% chance (95% CI 75.0-83.1).

Primary or post-primary status of individuals was assigned to patients who were presenting at an enrolled health facility based upon their reported day of infection and serological markers. In total 169 patients were identified as likely experiencing their first DENV infection so were primary infections whilst 640 were identified as post-primary infections.

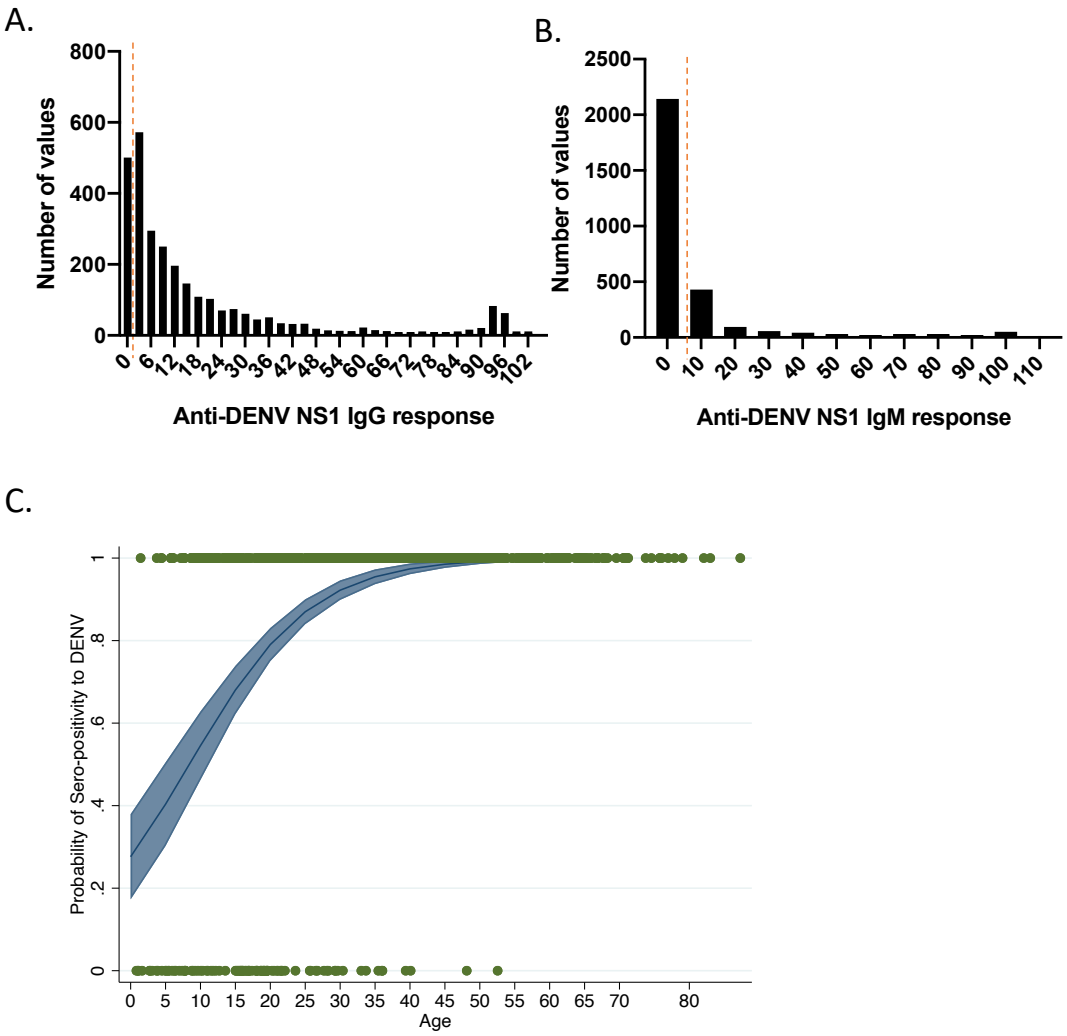


Figure 3.4 Inter-individual variability of serological responses to DENV of Patients and associated household members at their first visit and with complete data. A. Distribution of anti-DENV NS1 IgG kit units. Red dotted line represents calculated cut-off for a positive result (2.2). Seropositive=2135, Seronegative=800. B. Distribution of anti-DENV NS1 IgM kit units. Red dotted line represents calculated cut-off for a positive result (9.9). Seropositive=456, Seronegative=2479. C. Logistic regression of seroprevalence to DENV by age as generated by the study population. Blue shading is the 95% confidence intervals of predicted seroprevalence by age.

3.3.4 Effect of primary/post-primary status of viraemia

I next investigated the effects on viral load of serological status, with individuals compared based on whether they were experiencing their first or a subsequent infection.

Mean age of primary infection was used as a subsequent indicator for force of infection as previously described[175]. I therefore initially compared the distributions of ages among individuals classified as primary and post-primary (Figure 3.5A). This showed individuals experiencing a primary infection were significantly more likely to be younger than those with post-primary infection (mean 14.4 against 20.8, $p<0.0001$). Additionally, we see a difference in the distributions between the 2 groups with primary infections being significantly more skewed towards younger age groups ($p<0.0001$). There is a subset of people who are younger than 1 which have been called as post-primary infections, this is likely due to the prevalence of maternal IgG.

Symptomatic differences between primary and post-primary infections could contribute to a different trend of day of presentation, with a more pronounced symptoms causing care to be sought earlier than one where symptoms appear more slowly, thus indicating different biological factors involved in infection. Through comparing the self-reported day of infection between these two groups, I saw that primary infected individuals on average reported 1 day earlier than post-primary (Figure 3.5B), with primary individuals on average reporting on day 2 against day 3 for post-primary, with primary individuals being significantly more likely ($p<0.005$) to report earlier overall than post-primary individuals.

Due to the distribution of 1-6 days for the reported day of infection, I investigated whether a difference in viraemia could be observed both between which day of fever a person presented on, as previously, but results were divided between primary and post-primary cases (Figure 3.5C). As expected, viraemia reduced as day of reported infection increased for both cases presentation categories. There is a slight observable difference between the viraemia patterns however with individuals experiencing a primary infection showing a sigmoidal reduction of viral load, with a large drop in viremia between days 2 and 3, which was not observed in post-primary cases, however case numbers for most days were too small to show significant statistical differences.

Following this I compared overall viraemia between primary and post-primary individuals. Viraemia was observed in this study to be significantly lower ($p<0.005$) among people experiencing a post-primary infection than individuals who were a primary infection (Figure 3.5D). However, there is a 1-day delay between average reporting day which may give time for the virus to wane.

The distributions of primary and post-primary infections between the observed viral serotypes were compared (Figure 3.5E). Statistical tests comparing these distributions showed no significance between the 3 serotypes.

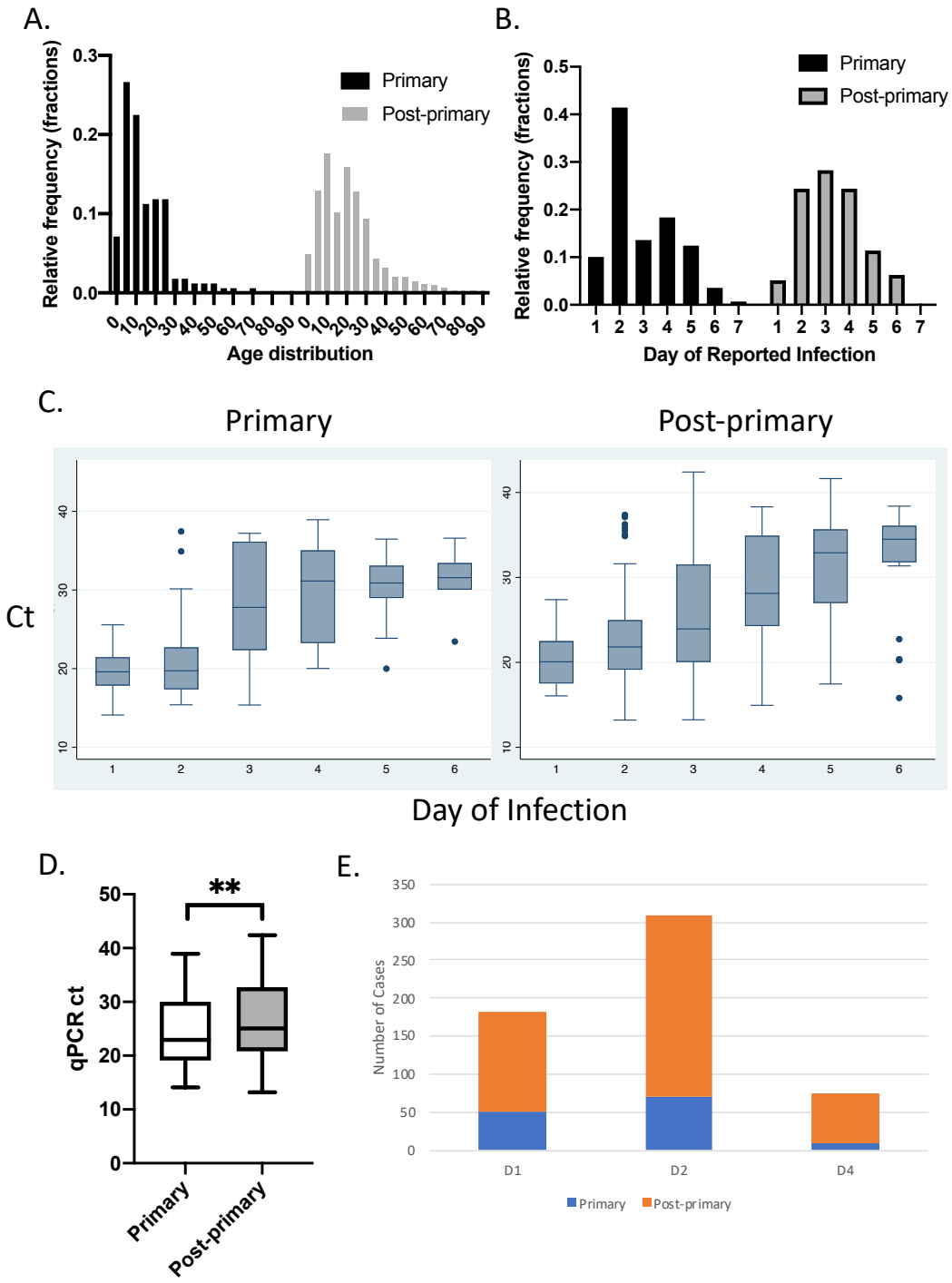


Figure 3.5 Effect of prior exposure to DENV on Viraemia. A. Distribution of Primary and Post-primary cases by viral serotype. B. Difference in Viraemia between all primary and post-primary cases. C. Effect of immune status (primary or post-primary) on Viral load throughout the course of infection. D. Difference in viraemia between primary and post-primary with a T-test of significance between the means shown. E. Break-down of distribution of reporting patients with primary and post-primary DENV infections.

Following the assignment of primary or post-primary status to individuals who were presenting at an enrolled clinic[175]. I used the serotype information and individuals age to calculate the respective force of infections of each serotype in the final year of the study, April 2018-April 2019 (Table 3.1). There were insufficient data points during this time period to determine this metric with D3 and D4. The resulting force of infections show that there is no significant difference between the serotypes, with similar values for both and equally wide and overlapping confidence intervals.

The overall force of infection calculation for this period, which determined by an NS1+ test result and not on a successful serotype result, was significantly lower as this included D3 and D4 which were less present during the sampling period, with a lower FOI, confidence intervals for this estimate were additionally much closer.

Table 3.1: Force of infection for serotypes D1, D2 and all patients across the final year of the study period (April 2018 – April 2019) with 95% confidence intervals.

	Force of infection (%)	95% confidence intervals
D1	8.9	[6.8-12.6]
D2	8.1	[6.0-12.3]
Overall	5.5	[5.2-5.8]

3.3.5 Symptomatic outcomes of reporting population

Severe DENV disease is a rare manifestation of DENV infection which if left undiagnosed and untreated can lead to death. The WHO has a list of warning signs which health professionals should be aware of among individuals reporting at health facilities with a suspected DENV infection. I used the presence of bleeding as the key indicator for warning signs among the list of symptoms given in the questionnaire of an enrolled patient.

I first investigated the distribution of reporting days for those with warning signs to determine whether it differed from the overall trend which we see among all other symptomatic infections. The resulting distributions were skewed to later days of disease presentation (mean $\pm$ S.D = 4.17 $\pm$ 1.42) (Figure 3.6A). This is significantly greater ($p<0.0001$) than the mean of 2.67 $\pm$ 1.20 which we see for cases presenting without warning signs.

I next compared the Ct distributions of viral load between patients with and without warning signs to investigate whether higher viral load was associated with a higher amount of warning signs. The resulting distributions were significantly different ($p<0.05$) with patients with warning signs generally having a higher Ct (Figure 3.6B). The difference is likely due to the significant difference in reported day however as I previously saw that viral load drops over time. It is therefore unlikely that the different manifestations of the disease have different viraemia timelines.

Warning signs have been reported to be more common in patients experiencing a post-primary infection instead of a primary one. I therefore investigated whether there was a difference in the proportions of enrolled patients presenting with warning signs between primary and post-primary dengue infections (Figure 3.6C). Interestingly there was no significant difference observed between these groups with both experiencing an almost equal proportion of warning signs ($p>0.05$). There were more post-primary cases enrolled within the study than primary (169:640) which could indicate that post-primary cases are more likely to be symptomatic.

I then investigated whether there was a difference in the risk of developing warning signs by serotype within our study population. The results showed that despite the large range in case

numbers between the serotypes there was no significant difference (Figure 3.6D) ($p>0.05$ between serotype).

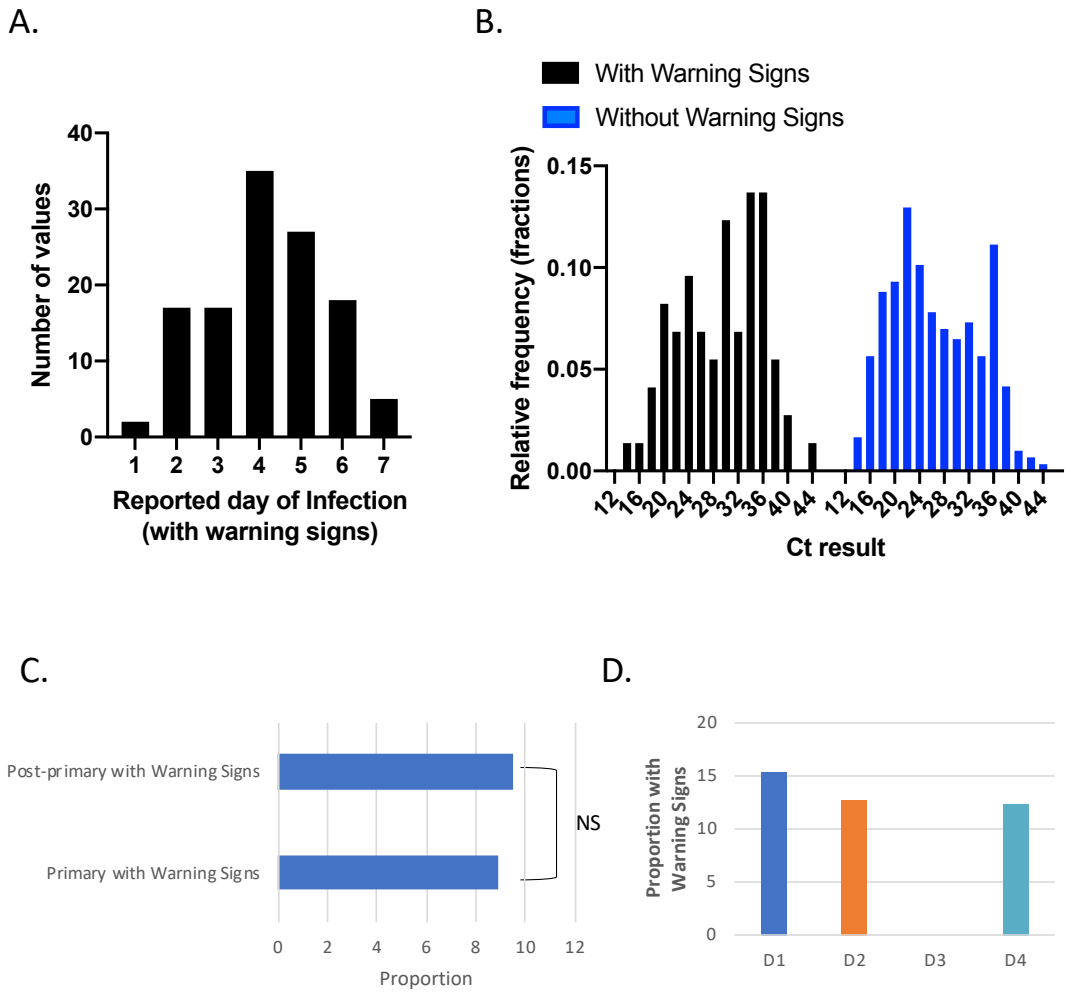


Figure 3.6 Investigation of predictors of severe disease. A. Distribution of reporting day of infection among those with warning signs. B. Distribution of Ct results of those reporting with and without warning signs. C. Proportion of those identified as primary or post-primary with warning signs. D. Proportion of successfully serotyped individuals identified as having warning signs.

3.3.6 Differences in viraemia between sub-clinical and reporting populations

I next investigated whether sub-clinical and asymptomatic infections would have comparable viraemia to individuals who presented with DENV at a medical facility enrolled in the study. Viraemia is often reported to be lower in people with mild or asymptomatic infections, suggesting that the ability to pass infection on to a mosquito is predicted to be less. Household contacts of index cases were tested at 2 visits at an average of 30 days apart for the presence of DENV and were asked if they had attended a clinic or were experiencing any DENV symptoms. Although none had attended a clinic, and few reported symptoms. Those who were positive for DENV by qPCR (26/1241), at either of the 2 timepoints, had DENV Ct results that were comparable to symptomatic index cases (who had all reported to a clinic), with no significant difference observed ($P>0.05$) (Figure 3.7A).

When sub-clinical individuals who reported having mild symptoms of DENV were compared to the reporting population, a similar pattern of viraemia change over time can be observed (Figure 3.7B). The individuals who were asymptomatic appeared to have a viraemia distribution which was comparable to Ct results found within symptomatic individuals, with the average result being around the equivalent of day 3 of the reporting cases (Figure 3.7B).

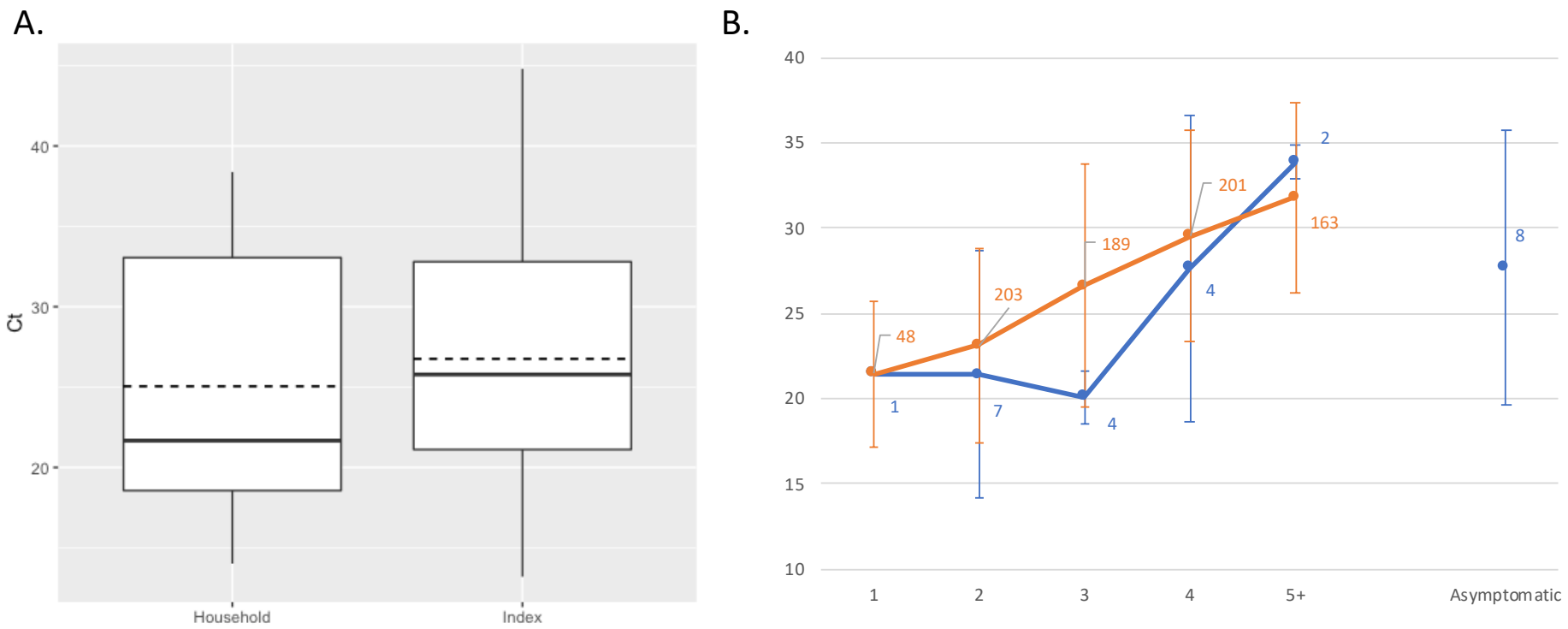


Figure 3.7: Comparison of viraemia between reporting Index cases and their associated sub-clinical Household contacts. A. Box-and-Whisker plot of distributions of viraemia between Index and Household cases, dotted line is the mean Ct result and solid line is the median. B. Viral profiles of viraemia over time according to self-reported day of fever if applicable with error bars showing standard deviation and numbers of individuals at each data point showed. Symptomatic individuals are shown by orange points whilst those who are sub-clinical but reported mild infection for a given number of days or were completely asymptomatic are shown in blue.

3.4 Discussion

Correctly identifying patients who are most at risk of developing a severe DENV infection is crucial in their successful management, which can significantly improve their prognosis[2]. During this study I assessed various virological and immunological markers which could be seen in the enrolled population presenting with DENV infections. As previous studies have shown that viral serotype, infection number and age could be indicators for disease severity[176], I initially investigated these metrics. I saw no statistically significant difference in disease outcome between either infecting viral serotype or whether they were primary or post-primary infected individuals. There was a small but significant difference between the viraemia of the infecting serotypes, with D4 being lower than D1 ($p<0.05$), interestingly another study has shown that D4 can present with lower viraemia than the other serotypes, with D1 and D2 generally having higher viraemia[177]. Differences in qPCR assay efficiency should not be ruled out however and future experiments could include viral plaque titration assays to compare differences in viraemia. These results suggest that among a reporting population there are other factors at play which may be causing severe infection. The majority of viraemic individuals enrolled in this study were people presenting at health facilities with disease, therefore asymptomatic and sub-symptomatic infections were likely missed, indeed the younger age of primary infections likely led to them being more likely to be taken to medical centres by a parent or guardian if they were suffering from an infection than an older independent person with a similar infection. The resulting viral profiles of disease day show, as expected, that the day of presentation following self-reported start of fever at a health facility, was the most significant predictor of viral load. This observation is unsurprising given the infection profile of DENV and is seen in other studies[161,178]. We observed that there were more post-primary infections than primary infections observed during the study period indicating that despite similar viraemia and observed risk of progression to severe disease, symptoms may be more severe in post-primary infections or indeed there more be more post-primary infections as most of the population will have already experienced a DENV infection by the age of 10.

The lack of significant difference between viral serotypes and their chances of developing severe disease suggests that there is little difference in the pathogenicity between viral

serotypes which were found in this population. Previous studies have both supported this finding and found differences in pathogenesis between serotypes [178,179]. Estimating the force of infection from the average age of the reporting population has been shown to be a good indicator for estimating the prevalence of DENV in a population[175]. I investigated whether any difference could be observed between the serotypes during a year of the study period with the greatest number of cases. Only D1 and D2 had sufficient case numbers to perform this metric and there was no significant difference observed between them. The force of infection calculated between serotyped and all samples with a positive DENV NS1 test result was significantly lower ($p>0.05$), this is likely due to the high viraemia required for a successful serotyping qPCR result, so force of infection measurements based on serotyping results are likely unreliable.

We have introduced a potential bias into our selection criteria against post-primary infection by using an NS1 test for selecting participants. This diagnostic test has been shown to be less sensitive for patients who are experiencing a post-primary infection, possibly due to competition between host antibodies and those used in the diagnostic test[180]. Future enrolment protocols could include qPCR diagnostics, especially for those who are experiencing a post-primary infection.

I used any report of bleeding in recorded questionnaire information as an indication of DENV disease with warning signs as it is listed as one of the warning signs for progression to severe disease from the WHO criteria and has additionally been strongly associated with severe disease[8,181]. Bleeding may be too vague of a descriptor here as an unexpectedly large number of individuals have been recorded as having severe disease, with over 10% of presenting positive individuals indicating that severe disease is not a rare outcome, although, as noted in other studies which investigate the prevalence of severe DENV, this is almost certainly a bias that has been introduced from looking at people presenting at clinics[182]. As not all so-called mucosal bleeding events have been associated with progression to severe infection, such as gum bleeding and epistaxis, future studies could instead use more specific notifiers of bleeding with hematemesis and gastrointestinal bleeding being the most highly significant warning signs[183].

This study highlights the difficulties of capturing sub-clinical infections, as despite two visits to index case households and over 2000 samples taken, only 26 contacts were positive by qPCR for DENV. Serological changes were more readily captured, with around 10% of household contacts showing seroconversion from naïve to IgG positive or serological evidence of recent exposure such as IgM positivity, more detailed results of which is covered in a later results chapter. This low sub-clinical capture success rate of virus may be a result of assuming that there is high transmission of the pathogen within a household as is the case with pathogens such as those spread through aerosols such as Influenza[184] and other vector-borne diseases like Malaria[185]. Indeed previous studies have shown slightly increased risk of DENV transmission within a household[173,174]. Despite this, the fact that it is spread by a day-biting vector makes it likely that other areas in endemic populations should be considered, the role of human movement during the day also likely plays a role in its transmission with one study noting that risk of DENV exposure is driven by individuals visiting places where there are infected mosquitos regardless of distance from the home[186]. A future study could therefore investigate the work, school or daytime contacts of index cases as well as their household members to identify more sub-clinical individuals.

Through comparing sub-clinical, including asymptomatics, and individuals reporting with more severe symptoms who sought medical help, I investigated whether there were any indications of differences in viraemia. As the resulting viraemia profiles were similar, despite the small sample size of sub-clinical and asymptomatic individuals, this suggests that asymptomatic individuals have sufficient virus in their blood for it to be passed on to a vector. Which has been shown in previous studies[92,166], although asymptomatic individuals in another similar study were found to have significantly lower average viraemia throughout the course of their infection[162]. Interestingly, the viral profile appeared to be more stable and at a lower Ct (indicating more virus) for the sub-clinical individuals before viraemia quickly dropped off, however as the sample size is small, significant conclusions are difficult to reach, further studies should therefore be carried out to add more data to this finding. The viral kinetics of these asymptomatic infections would be interesting to determine in future studies as a previous study has shown that despite lower average viraemia at each day there is a slower observed decay rate when the virus is sampled every day[162].

During this study I compared serological and genomic marker of infection to determine whether any were indicative of a severe DENV infection. Samples sent to me had already been assayed by NS1 antigen test which gave a simple positive or negative result. Future studies could compare whether there were any significant differences in titre of NS1 between these participants as this has been shown to be indicative of a more severe DENV infection[187,188].

Chapter 4 A Phylogenetic Study of Dengue Virus in Urban Vietnam shows Long-term Persistence of Endemic Strains

This chapter has been formatted as a journal article and has been submitted to the journal "Virus Evolution" awaiting review.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1704559	**Title**	Mr
First Name(s)	James		
Surname/Family Name	Ashall		
Thesis Title	A Genomic and Immunological Investigation of Dengue Virus Transmission		
Primary Supervisor	Martin Hibberd		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Virus Evolution
Please list the paper's authors in the intended authorship order:	James Ashall; Sonal Shah; Joseph R. Biggs; Renee Chang; Yalda Jafari; Oliver J. Brady; Huynh Kim Mai; Le Thuy Lien; Hung Do Thai; Hien Anh Thi Nguyen; Dang Duc Anh; Chihiro Iwasaki; Noriko Kitamura; Marnix Van Loock; Guillermo Herrera-Taracena; Freya Rasschaert; Liesbeth

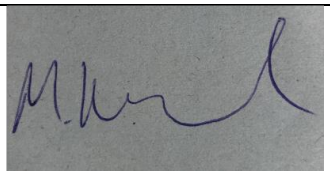
	Van Wesenbeeck; Lay-Myint Yoshida; Julius Clemence R. Hafalla; Stephane Hue; Martin L Hibberd.
Stage of publication	Accepted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Sequences were generated by myself and Renee Change under my supervision. Relevant samples for processing was done by SS and myself. Data analysis and writing of the paper was done by myself with advice from SH and MLH. All authors reviewed drafts.

SECTION E

Student Signature	
Date	22/06/22

Supervisor Signature	
Date	22/06/2022

A Phylogenetic Study of Dengue Virus in Urban Vietnam shows Long-term Persistence of Endemic Strains

James Ashall[1]; Sonal Shah[1]; Joseph R. Biggs[2]; Renee Chang[1]; Yalda Jafari[2]; Oliver J. Brady[2,8]; Huynh Kim Mai[3]; Le Thuy Lien[3]; Hung Do Thai[3]; Hien Anh Thi Nguyen[4]; Dang Duc Anh[4]; Chihiro Iwasaki[5]; Noriko Kitamura[5]; Marnix Van Loock[6]; Guillermo Herrera-Taracena[7]; Freya Rasschaert[6]; Liesbeth Van Wesenbeeck[6]; Lay-Myint Yoshida[5]; Julius Clemence R. Hafalla[1]; Stephane Hue[2,8]*; Martin L Hibberd[1]*.

1) Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK.
2) Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.
3) Pasteur Institute of Nha Trang, Nha Trang, Vietnam.
4) National Institute of Hygiene and Epidemiology, Hanoi, Vietnam.
5) Paediatric Infectious Diseases Department, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan.
6) Janssen R&D, Janssen Pharmaceutica NV, Beerse, Belgium
7) Janssen Global Public Health, Janssen Research & Development, LLC, Horsham, Pennsylvania, USA.
8) Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK.

*Equal contributions

4.1 Abstract

4.1.1 Background

DENV causes repeated outbreaks of disease in endemic areas, with patterns of local transmission strongly influenced by seasonality, importation via human movement, human immunity, and vector control efforts. Understanding how each of these interact to enable endemic transmission (continual circulation of local virus strains) is largely unknown. During periods of low transmission, no cases can be reported in an area for extended periods of time, perhaps wrongly implying eradication a local strain from that area.

4.1.2 Methods

Individuals who presented at a clinic or hospital in four communes in Nha Thrang, Vietnam were initially tested for DENV antigen presence. Enrolled positive individuals then invited their corresponding household members to participate, and those who enrolled were also tested for DENV. The presence of viral nucleic acid was confirmed using qPCR and subsequent viral samples were then whole genome sequenced using a target enrichment library preparation technique and Illumina Miseq sequencing technology. Generated consensus genome sequences were then analysed using phylogenetic tree reconstruction to categorise sequences into clades with a common ancestor, enabling investigations of both viral clade persistence and introductions. Hypothetical introduction dates were additionally assessed using a molecular clock model that calculated the time-to-the-most-recent-common-ancestor (TMRCA).

4.1.3 Results

We obtained 511 DENV whole genome sequences covering four serotypes and more than 10 distinct viral clades. For 5 of these clades, we had sufficient data to show that the same viral lineage persisted for at least several months and others much longer. By comparison with other published sequences from elsewhere in Vietnam and around the world, we saw at least two different viral lineages were introduced into the population during the study period (April 2017-2019). Next, by inferring the TMRCA from the construction of molecular-clock

phylogenies, we predicted that two of the viral lineages had been present in the study population for over a decade.

4.1.4 Conclusions

We observed 5 viral lineages co-circulating in Nha Trang from three DENV serotypes, with two likely to have remained as uninterrupted transmission chains for a decade, suggesting cryptic persistence in the area, even in during periods of low reported incidence.

4.2 Introduction

Dengue virus (DENV) is one of the most important arthropod-borne viral infections in the world today, with roughly 2.5 billion people at risk of infection[1], mainly in south-east Asia, South America and the Caribbean. Globally there are estimated to be 100-400 million new infections each year and around 40,000 deaths[7,71]. The number of confirmed infections reported to the World Health Organisation (WHO) has increased every year since 1990, primarily due to improvements in testing infrastructure and increased numbers of outbreaks, peaking in 2019 with 5.2 million cases and 4032 deaths[8].

DENV is an arthropod-borne virus of the *Flaviviridae* family that includes four genetically related yet antigenically distinct serotypes (D1, D2, D3, and D4). The virus is spread by the bite of *Aedes* mosquitoes, primarily *A. aegypti*[189], which are distributed in most tropical and subtropical areas of the world. Households have been proposed as the primary location for the insect vector[186], particularly those within areas of critical population density and without access to piped drinking water[190], with lower insect abundances associated with non-residential districts[191]. Vector control through prevention of biting, breeding site removal and insecticide are currently the principal methods of controlling DENV in endemic populations[192]. There are also new vector control measures, which are attempting to reduce or remove the vector population through the release of sterilised male mosquitoes[193], as well as strategies that reduce the ability of mosquitoes to transmit the virus by introducing those infected with the *Wolbachia* bacteria[91] that have had promising success in reducing reported DENV disease [194]. A vaccine for all for DENV serotypes was developed in 2015[70], although a requirement for pre-testing before use has limited its widespread adoption[195]. There is also the possibility for antiviral treatment in the near future[42] that might have the potential for prophylactic intervention in outbreaks[196].

The persistence of DENV in a population is dependent upon several factors that involve the interplay between the human host and insect vector[94]. The primary human factors are

therefore population density and how many susceptible individuals remain or are added to a population as the virus spreads through it. This is particularly challenging to measure with DENV due to the complex cross-reactive role of antibodies between the four serotypes, associated with both immune protection and enhancement of infection[56,153,197]. Mosquito vectors and the impact of the environment on their lifecycle are the primary reason why DENV displays a seasonal pattern that is often closely related to rainfall and temperature, conditions vital for the development of the larval stages of their lifecycle[198]. Epidemiological data can indicate breaks in transmission during periods in which environmental conditions are unfavourable to mosquito breeding[14], yet because a high percentage of cases are asymptomatic and/or unreported, the virus could persist at levels below the threshold for detection[92,175].

Previous studies have shown that locally circulating genotypes and strains from previous seasons are capable of contributing to seasonal spikes in case numbers and even outbreaks[199–201]. Despite this, the amount of time DENV serotypes and strains remain in an endemic population despite increasing immunity, control measures and periods of very low circulation remains poorly studied. Additionally, the role of introduced strains in seasonal outbreaks against a background of highly endemic strains has not been extensively investigated.

The phylogenetic analysis of viral genomes enables the identification of closely related viruses in an infected population, with similar viral genetic identity being indicative of a recent transmission event between sampled cases[131]. This approach allows the determination of epidemiological linkage between observed infections without the reporting biases inherent to traditional contact tracing methods[127]. DENV is a good candidate for this type of analysis, as it typically only infects a host for a very short amount of time (between 4 and 7 days in humans and up to the 1 month lifespan of the mosquito[176]) before being transmitted to a new host, ensuing the evolution dynamics, that is the evolutionary rate of the virus, are on a scale

that closely correlates with transmission[202–204]. By using these analyses, the transmission characteristics of DENV from an international scale can be inferred, thereby determining likely origins of viral import that can have implications for control measures[205].

Classifying viruses based on their genetic similarity using phylogenetic information gives us the most reliable metric of viral persistence in a population. Phylogenetic information coupled to molecular clock inference can additionally be used to determine the timing of epidemiological events along a phylogeny, such as the time of the most recent common ancestor (TMRCA) of a viral clade, which can be utilised to estimate a time of introduction of a particular strain or clade to a country or how long it has been in circulation in the area.

4.3 Methods

4.3.1 Study Population

The study was conducted at Nha Trang population-based cohort study site in central Vietnam. Enrolment of the study cohort was conducted in the communes of Vinh Hai, Vinh Phuoc, Vinh Tho, and Vinh Hoa in Nha Trang City, Vietnam. Individuals of all ages, who resided in the selected communes and who visited either a local polyclinic (between October 2016 and May 2019) or the city's Tropical Medicine Hospital (between December 2016 and April 2019) with suspected dengue were deemed eligible. In the polyclinic, those who presented with suspected dengue fever were approached, and those who consented to be enrolled in the study were subsequently tested for dengue with an NS1 rapid test (DENGUE NS1 AG STRIP, #70700, BioRad). In the hospital, only those who were diagnosed with dengue fever on the basis of an NS1 rapid test, were enrolled in the study.

Study teams then visited the index case's home addresses, where additional consenting members of the household were also enrolled and had blood samples taken. A questionnaire was also completed by all enrolled individuals in order to gather information on clinical and demographic information relevant to this study, such as age, time of fever onset, home GPS coordinates and typical daytime location.

Historical DENV case numbers were acquired from the Nha Trang Preventive Medicine Center, Khanh Hoa CDC and Tropical Diseases Hospital and were based on clinical signs and symptoms.

4.3.2 DENV RNA detection

NS1-positive serum samples were processed using Qiagen viral RNA extraction kit and tested by real-time reverse transcription-quantitative polymerase chain reaction (RT-qPCR). Viral RNA detection and viral load quantification were determined using LightMix® Dengue Virus EC kit (Cat. -No. 58-0700-96, TIB MolBiol, Berlin, Germany), which can identify all four DENV

serotypes, with Luna® Universal Probe One-Step RT-qPCR Kit (New England Biolabs, USA). RT-qPCR assay was then performed in Applied Biosystems® 7500 Fast Dx Real-Time PCR instrument (Thermo Fisher Scientific, USA) following cycling conditions in the manufacturer's protocol.

The CDC real-time RT-PCR assay was used to determine the serotype of DENV in the infected individuals. The protocol used was based on Santiago *et al.,* 2013[83] with minor adjustments. The Ultraplex 1-Step Tough Mix qPCR kit (Quantabio) was substituted for the recommended kit with the manufacturer's instructions followed.

4.3.3 Sequence Generation

RNA extracted from DENV qPCR positive patient sera were processed for whole-genome sequencing on the Illumina Miseq platform. Two library preparation techniques were used; for samples which were positive by serotyping assay (to serotypes 1 or 2), an amplicon sequencing approach was used based on the protocol developed by Josh Quick for Zika virus[139]. Briefly, multiple 400bp amplicons were designed using the primal scheme software (available at ***http://primal.zibraproject.org/***) which is based on primal3. 10 whole genome sequences from temporally and spatially close outbreaks were used to design the sequencing primers.

For sequences with an unknown serotype but were positive for DENV RNA by the qPCR singleplex assay, a target enrichment approach was followed. 120nt length biotinylated probes were designed by running the CATCH programme[141] on available DENV 1,2,3,4 genomes from NCBI GenBank. The Agilent SureSelect$^{XT\ HS}$ Target Enrichment kit and protocol was followed with the revisions recommended for pathogen sequencing[206].

4.3.4 Phylogenetic analysis

Consensus DENV full genomes of the target viruses were generated from the FASTQ files produced by Illumina MiSeq sequencing. Raw files were screened for quality and trimmed to remove primer sequence bias using Fastp[146]. Cleaned FASTQ files were then aligned to a serotype-specific reference genome using the Burrow-Wheeler Alignment for short-reads (BWA-MEM)[147] along with the SAMtools view and sort packages[148]. Reference DENV genomes were extracted from GenBank: DENV1: JQ045626, DENV2: GU131898 and DENV4:NC_002640. Aligned reads were called using SAMtools mpileup and iVar consensus[149] programs with a resulting consensus genome sequence in FASTA format. The genotype and serotype of the sampled viral populations were confirmed from the generated consensus genomes, using the DENV typing tool available from the Genome Detective resource[17] which uses phylogenetics and pairwise distance within and between groups to known references, to assign a genotype to a sequence.

Processed consensus genomes were aligned against all full-length DENV genomes available in the NCBI Virus Variation Resource Database[207] (as of 01/03/2022), per serotype, using MAFFT[143]. The resulting alignments were visually checked and manually adjusted in AliView[208] to remove the UTR sequencing regions, as these greatly differed in length between publicly available sequences, and sequences which contained gaps where sequencing read depth (<200) was insufficient. Maximum likelihood (ML) phylogenies were generated using IQTREE [209], under the best fitting model of nucleotide substitution (as determined by the model finding procedure implemented in IQTREE) and 1,000 ultrafast bootstrap replicates for branch support assessment. The resulting trees were edited, and mid-point rooted using FigTree.

Time-scaled phylogenies and root-to-tip linear regressions were inferred for each serotype with TreeTime[210], using the ML trees generated with IQTREE. Time trees were estimated using the treetime function, with up to 500 interactions, taking into account covariation when estimating rates of nucleotide substitutions along the phylogeny and with automatic rerooting of the tree to maximize the clock-like signal. The “mugration” model for ancestral state reconstruction implemented in TreeTime was used to infer the most likely migration

patterns of the sampled viruses along the phylogeny, using the sampling location of the genomes as discrete traits (See Supplementary Table 1).

4.3.5 Selection analysis

We investigated positive evolutionary pressure across the genomes of the 2 clades with the largest case numbers and which also spanned over the greatest time period (D1.I and D4.I) using the MEME[211] and FUBAR[211] tools from datamonkey.org[212] to investigate the potential presence of episodic and pervasive selection pressures respectively. The dengue polyprotein amino acid sites which the models predicted were under significant positive pressure were investigated for frequency, both before and during the period of elevated case numbers.

4.4 Results

4.4.1 Genetic diversity of DENV1-4 genomes isolated in Nha Trang, Vietnam, between 2016-2019

A total of 511 DENV full-length genomes were generated from the individuals sampled over the study period (2016-2019). The breakdown of the serotype and genotype distribution of the sequences, as determined by the Genome Detective, is given in **Figure 4.1**. Viruses from all 4 serotypes were observed, with a majority of D1 (37.8%) and D2 (41.9%) and fewer D4 (18.8%) detected. D3 was represented by a single genome (0.2%) and was excluded from subsequent analyses. Only one genotype within serotypes D1, D3 and D4 was identified (Genotype I, III and I, respectively), whilst with serotype D2 sequences were of two distinct genotypes, namely II and V.

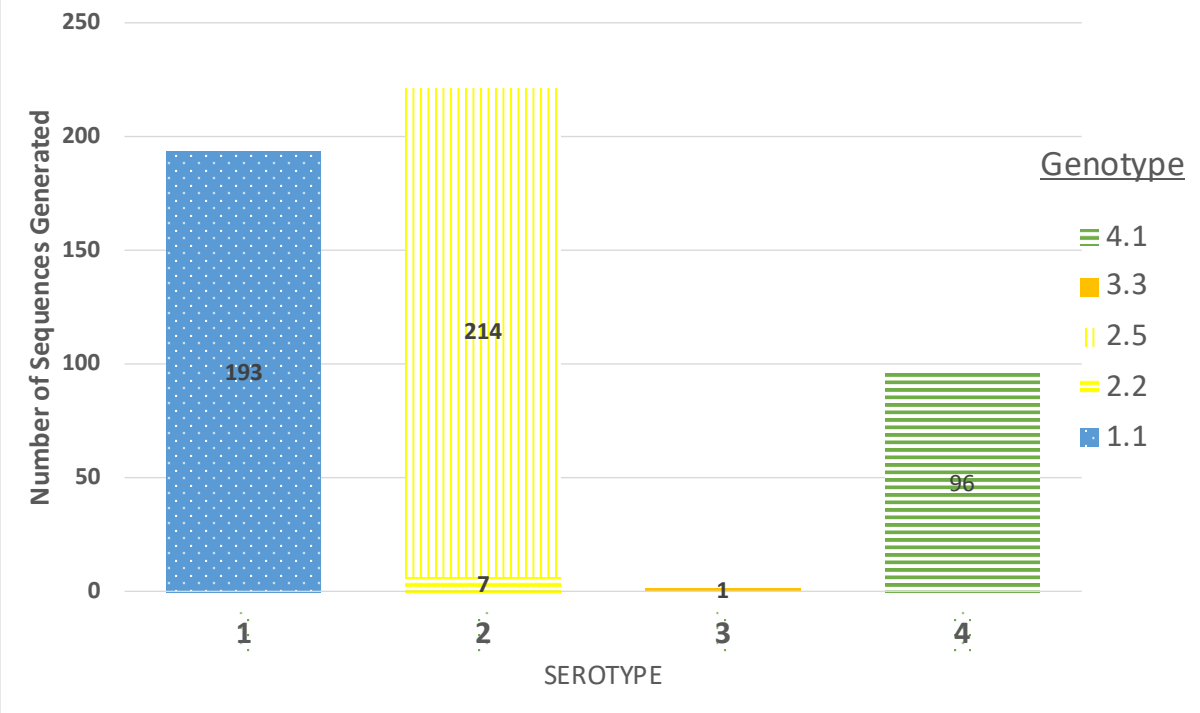


Figure 4.1: Stacked bar chart of number of full genome sequences generated during the study and their breakdown by serotype and genotype.

Maximum likelihood phylogenies of the three studied serotypes combined with international full-length sequences available from GenBank confirmed the presence of multiple DENV lineages amongst the sampled infections (**Figures 4.2, 4.3 and 4.4**). Clades were identified as clusters of sequences with the same geographical origin according to the mugration modelling in TreeTime and here are labelled according to their serotype and then given a clade number (for example "Clade 1.I" is the largest clade identified for D1). The majority of the Nha Trang D1 genomes formed three distinct clades: (I) a large main cluster containing 173 local sequences (Clade 1.I in **Figure 4.2**; 90% of all Nha Trang D1 genomes) and 35 sequences from neighbouring countries and other parts of Vietnam (branch support 98.6%), (ii) a smaller cluster of 13 sequences containing 12 local sequences and 1 from elsewhere in Vietnam (Clade 1.II in **Figure 4.2**; support 100%), and (iii) a cluster mainly made of sequences from southern Vietnam and neighbouring countries (n=17), but also including 7 sequences from Nha Trang (Clade 1.III in **Figure 4.2**; branch support 100%). These findings suggested multiple introductions of DENV genotype D1 into Nha Trang prior to the commencement of the study and over the course of the 3 years of the sampling in our study.

The D2 phylogeny also indicated at least two introductions of this serotype in the area (**Figure 4.3**), with one large clade of 218 genotype V sequences, including 214 from this study in Nha Trang and four from southern Vietnam, China and Cambodia (Clade 2. I in **Figure 4.3**; 100% branch support) and a second smaller clade (Clade 2.II; branch support 100%) of 7 genotype II genomes from Nha Trang.

All but one D4 genomes formed a large clade of 101 sequences (Clade 4.I in **Figure 4.4**; branch support 96%) and included sequences from China, Cambodia and elsewhere in Vietnam. The remaining D4 genome clustered distinctly with sequences isolated from Thailand (branch support 99%).

At least one large clade of Nha Trang sequences (i.e., Clades 1.I, 2.I and 4.I) was observed in each of the phylogenies (**Figures 4.2-4**), suggesting that, for each serotype, a vast majority of the studied cases were sampled within local transmission chains. The average number of

nucleotide substitution per sites within each clade are listed in **Supplementary Table 1 and depicted in Supplementary Figure 1**.

Isolated sequences (sequences with no close relative from the study also in the phylogeny) were observed for serotypes D1 and D4, indicating independent introductions that either did not spread in the population or, more likely, were not sampled. Conversely, we observed that D1 and D2 sequences clustered very closely with other sequences isolated from Nha Trang and Vietnam. There were far fewer D4 sequences available, consequently clustering was limited to a group of sequences imported into China.

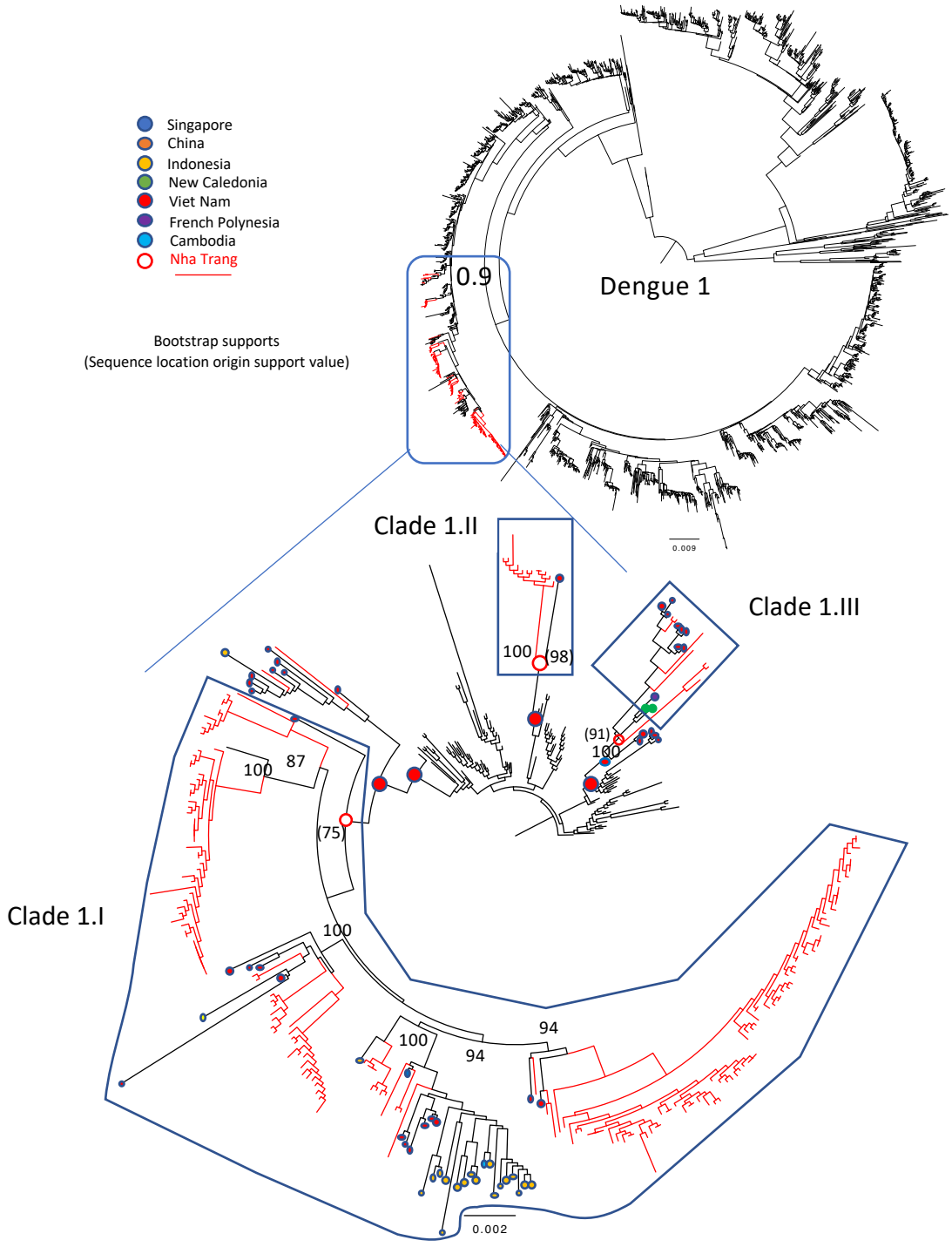


Figure 4.2: Midpoint rooted maximum-likelihood trees of 2446 publicly available full genome length DENV1 sequences (black terminal branches) with 193 DENV1 genomes generated in this study (red branches). The relevant region of the tree has been isolated as a separate tree. Branch supports, calculated as bootstrap scores from 1,000 replicates, are indicated on relevant branches. The sampling location of the publicly available sequences in a Nha Trang cluster is indicated by a coloured dot (see legend). The most likely country

of origin of relevant internal nodes is indicated by a coloured node, with probability values given in brackets.

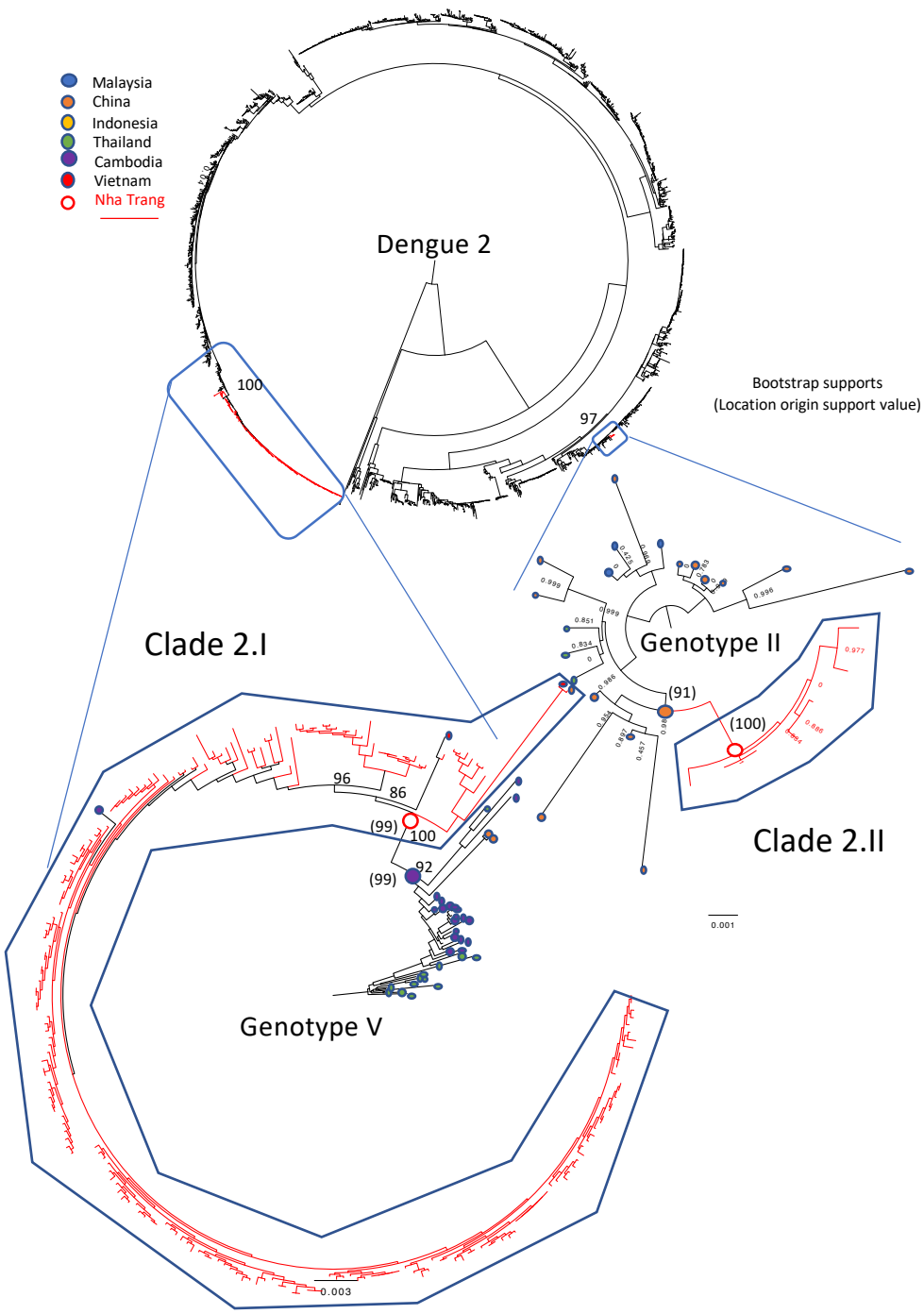


Figure 4.3: Midpoint rooted maximum-likelihood trees of 1891 publicly available full genome length sequences with the 221 full-length D2 genomes generated in this study shown as red branches. The relevant regions of the tree have been isolated as separate trees with the difference in genotypes of these sequences shown. The given country of isolation for the most genetically similar corresponding sequences is marked as a coloured dot at the end of the branch. Bootstrap support values over 70 and on relevant supporting

branches have been shown except for genotype II which is small enough to show all bootstrap supports

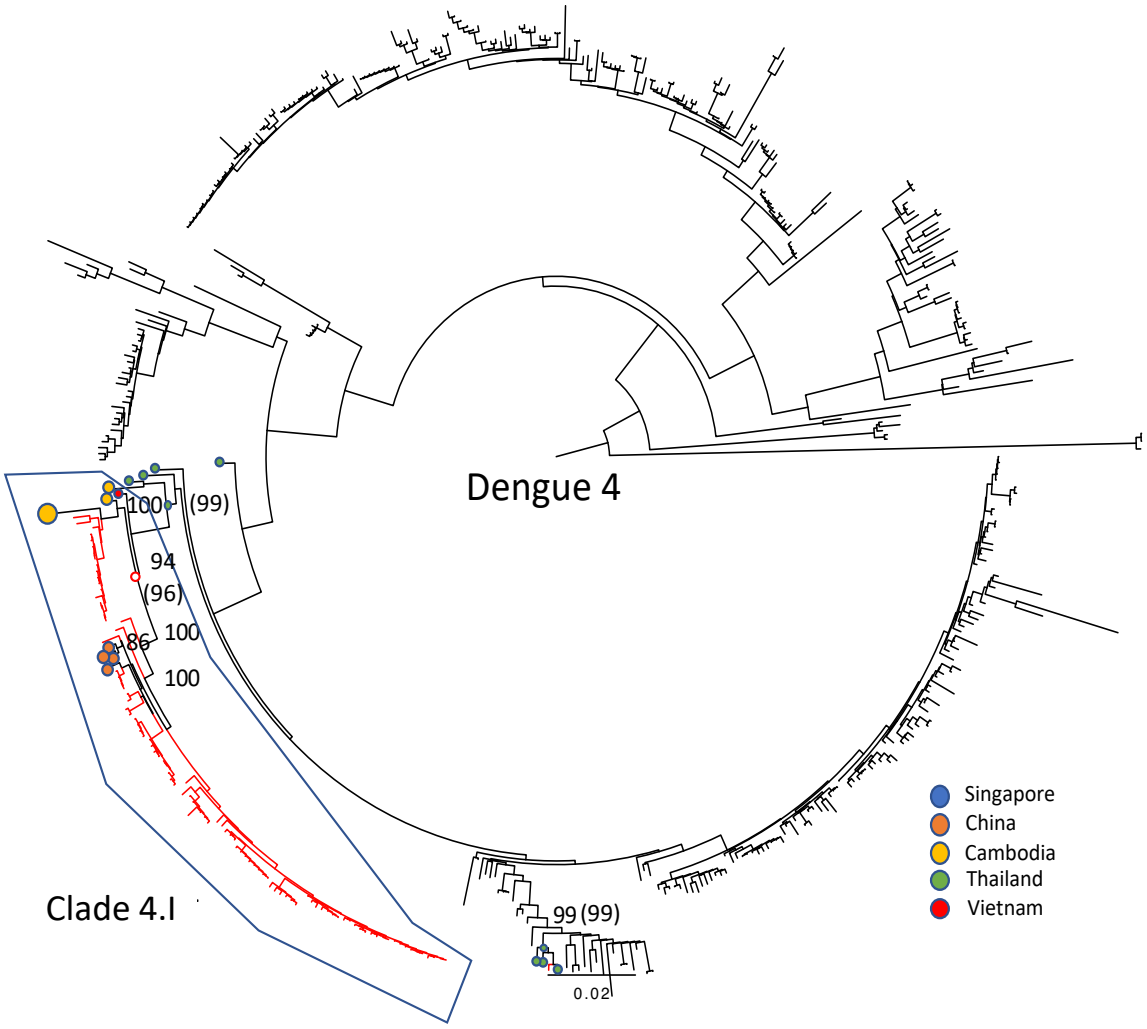


Figure 4.4: Midpoint rooted maximum-likelihood trees of 407 publicly available full genome length sequences with the 96 D4 genomes generated in this study shown as red branches. The given country of isolation for the most genetically similar corresponding sequences is marked as a coloured dot at the end of the branch. Bootstrap support values over 70 and on relevant supporting branches have been shown.

4.4.2 Evidence for Dengue virus persistence in the study area

The sampling time interval of the viral genomes within each local clade indicated that these chains of transmission persisted for between 121 days (Clade 2.II) and 878 days (Clade 1.I) during the sampling period in the area (**Figure 4.5**). This provided a complex picture of serotype prevalence shifting over time, but with at least two distinct serotypes present every month of the sampling interval from April 2017 onwards. Serotypes D1 and D4 were ubiquitous and remained in the population for at least 2 years, despite periods of low incidence and even after the introduction of D2 (first sample: August 2018) and the resulting outbreak of 2018-19. Our sampling strategy did not change during the sampling period, suggesting that it was the increase in reporting dengue cases during the outbreak that allowed more samples to be collected, that led to the consequent increased sensitivity to detect more clades during this period. This absence of strain replacement and long-term persistence of multiple serotypes in the population in the local area suggest sufficient levels of susceptible hosts and on-going transmission to maintain the lineage all year round, even between DENV outbreak periods, when incidence is low, and despite vector control in the affected areas.

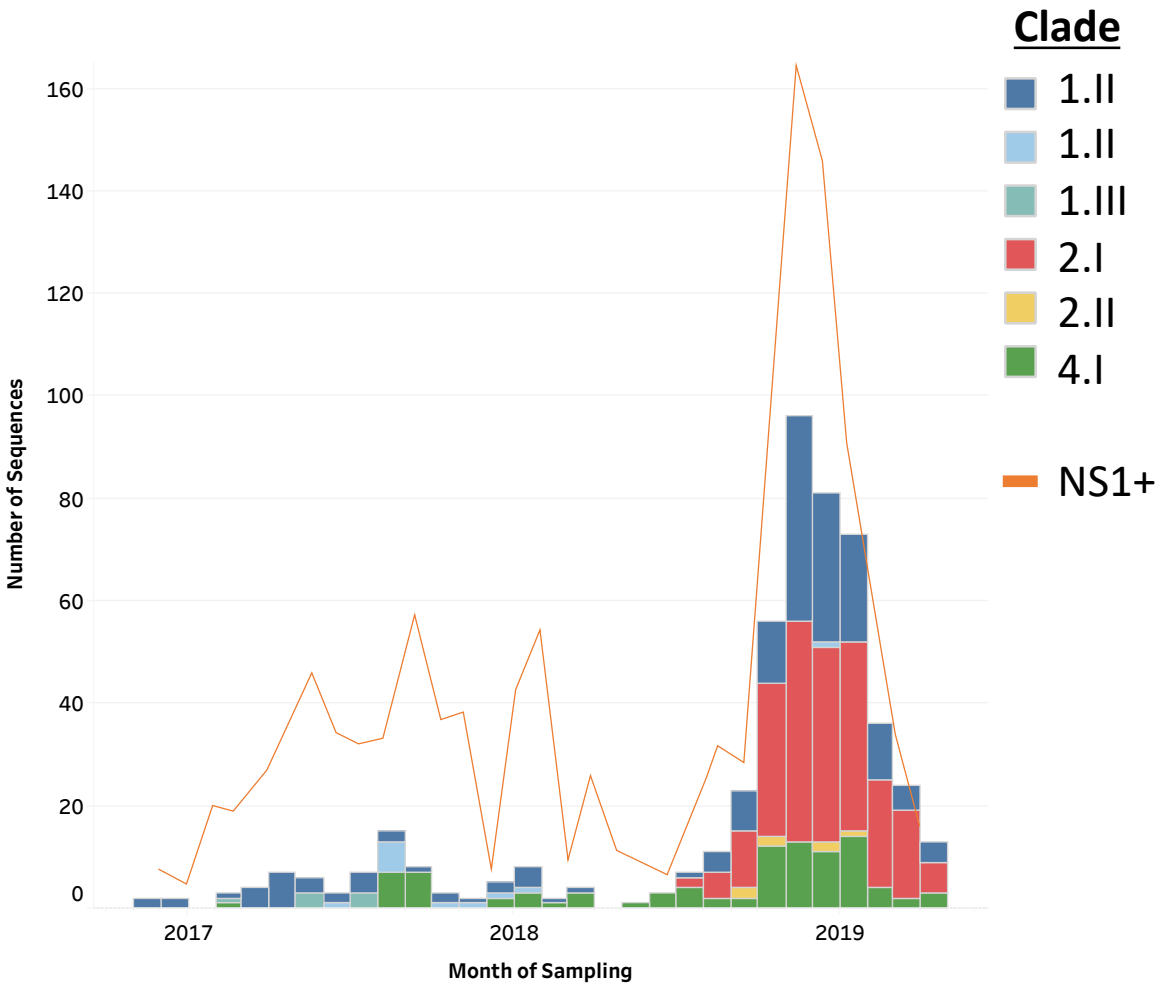


Figure 4.5: Number of sequences in each clade sampled each month from October 2016 to April 2019 with total number of NS1 positive individuals recorded during the same time period.

Ancestral state reconstruction revealed that, of the 6 clades identified in Nha Trang across the three serotypes, 4 had an ancestral node inferred to be located in Nha Trang with a probability greater than 75%: Clade 1.I, Clade 1.II, Clade 2.I, Clade 4.I. These clades were considered to represent local transmission chains and were further analysed. Inversely, the most likely place of origin of Clade 2.II was estimated to be China (**Figure 4.3**; probability 91%) with and was excluded from the persistence time quantification.

The complete time of persistence of these local transmission clusters was estimated from dated phylogenies as the time span between the estimated time of introduction of a clade's founder virus in the area and the time of sampling of the most recent isolate in that cluster (**Figure 4.6**), using the sampling dates of the Nha Trang genomes to calibrate the molecular clock.

Calculating the time to the most common recent ancestor (TMCRA) of the identified clades suggested that the first of the clades to be imported into Nha Trang was clade 4.I, with a date of introduction as early as September 2008 (90% CI: May 2007 – Jan 2010), followed by clade 1.I. (Mar 2009; 90% CI: Jul 2006 – Nov 2011), clade 2.I (Jun 2014; 90% CI: Jul 2012 – May 2015), clade 1.II (Jun 2017; 90% CI: Feb 2017 – Aug 2017) clade 2.II (Jun 2018; 90% CI: Mar 2018 – Sep 2018) (**Figure 4.7 and Supplementary Table 2**). Two of the clades showed evidence of at least one decade of local persistence, with intervals of 3,887 days (clade 4.I) and 3,707 days (clade 1.I) between the estimated time of introduction and most recent sampling. We see introduction events of novel clades (Clades 1.II and 2.II) to the population during a low case-number season, these remained and were observed during the peak outbreak season in 2019. No introduction events were noted during or within the previous year of 2019.

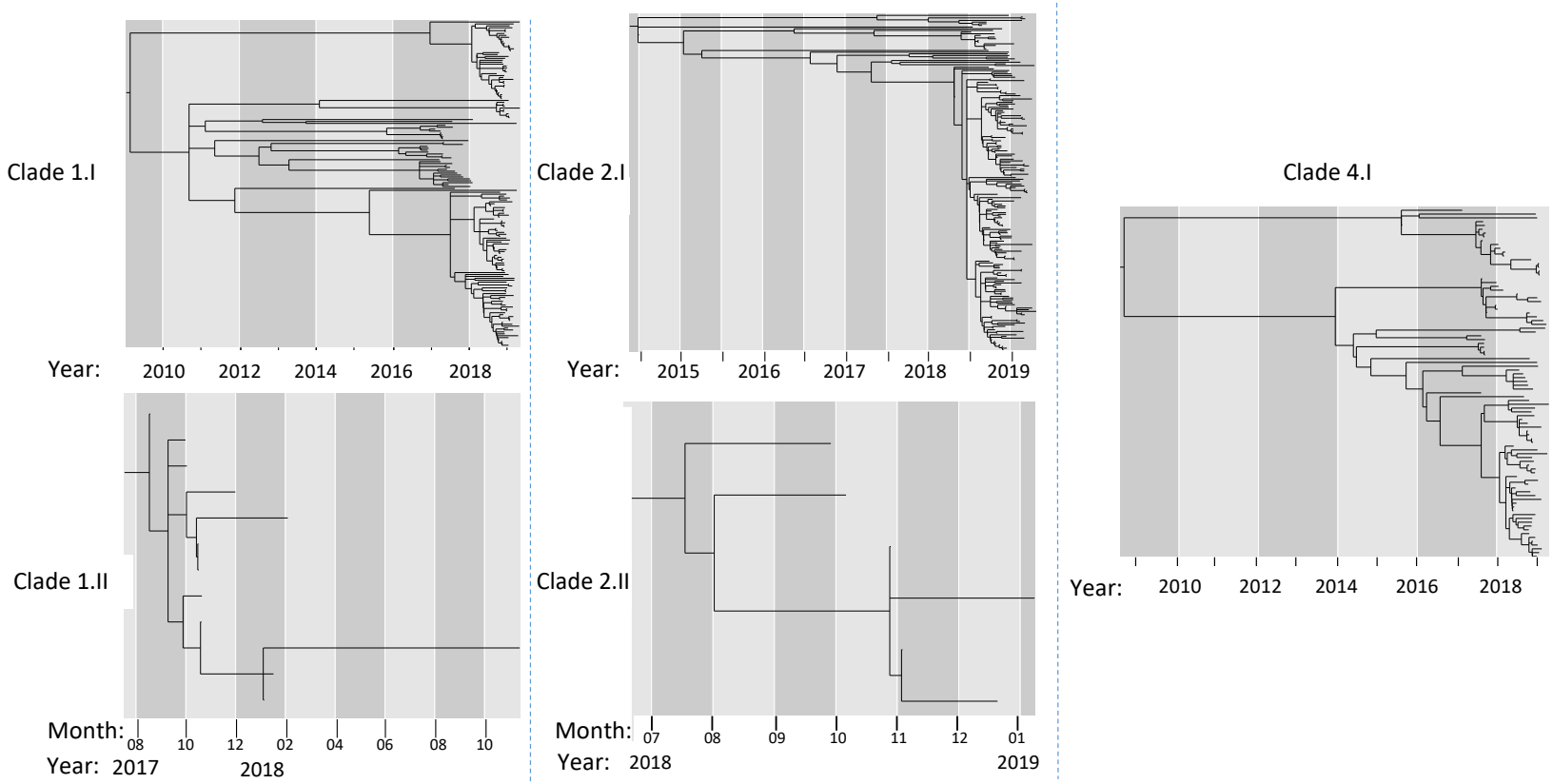


Figure 4.6: Dated phylogenies of 5 DENV clades sampled in Nha Trang between 2016 and 2019. Branch lengths correspond to time, with the year and month of the node shown in x axis.

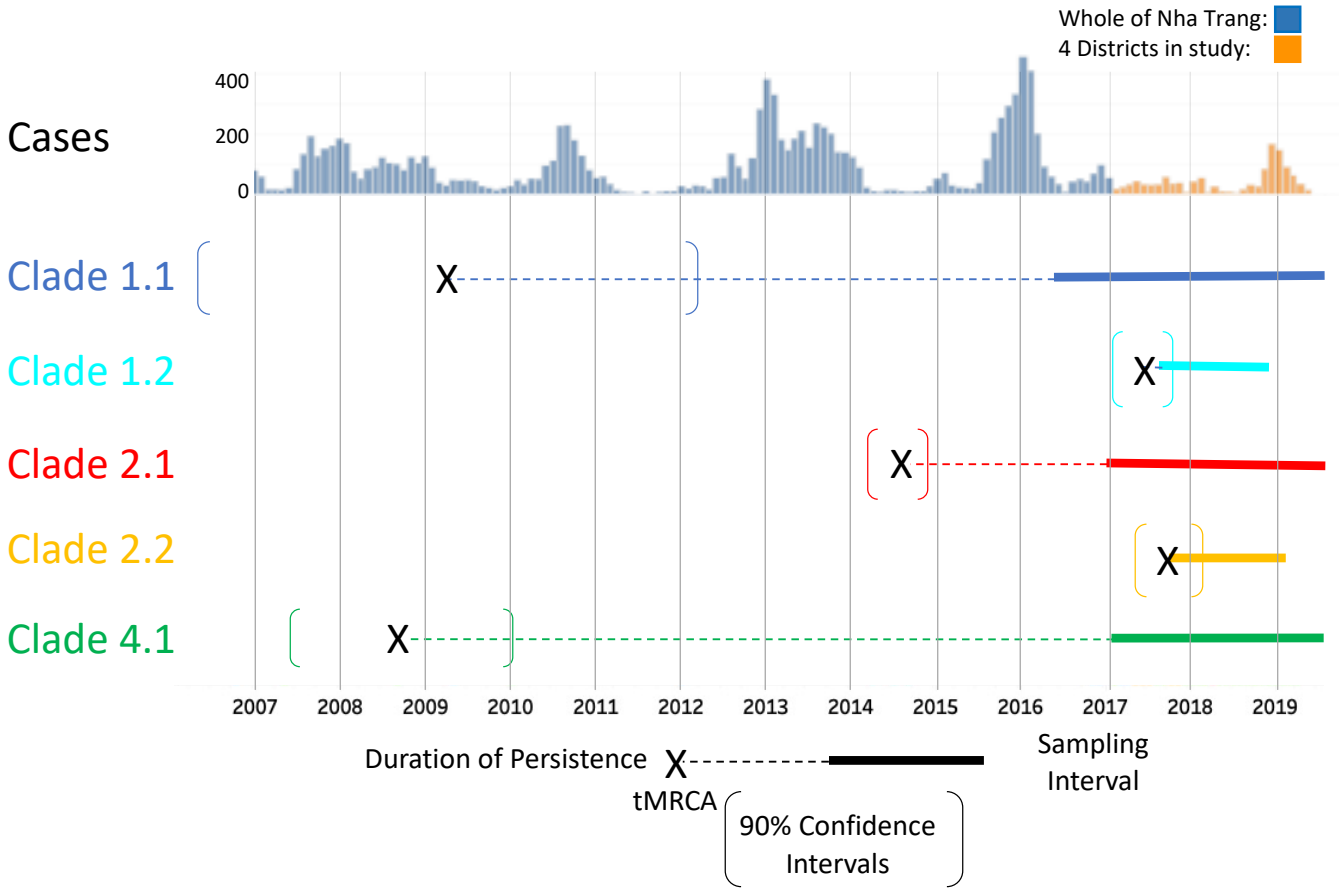


Figure 4.7: Timescale of sampled and predicted historical samples based on linear regression analysis of phylogenetic diversity overlayed on historical and recently reported DENV cases in Nha Trang. Sequences have been divided into serotype then coloured by clade according to genetic similarity and putative node origin. Fixed lines denote that a sequence has been sampled from at least a 2-month time period, whilst dotted lines are the hypothesised TMRCA denoted as an X with 90% confidence intervals denoted by square brackets.

4.5 Discussion

This study provides strong evidence of the persistence of locally circulating DENV strains throughout its almost three-year duration, determined using phylogenetic techniques. This is despite significant fluctuations in case numbers and control measures implemented by the Vietnam National Dengue control program[213]. In addition to the continued prevalence of the clades that were directly sampled, we also estimated that two of these circulating lineages had likely been present in the area for a decade. Introduction events of new viral lineages that were distinct from those in persistent circulation were also observed during the study period.

Categorising clades according to the phylogenetic similarity of the sampled viral sequences, showed that clades 1.I and 4.I persisted for the longest time period, whilst the more recently introduced clade 2.I clade became the dominant clade sampled during the final year of our study. In this final year, an increase in case numbers compared to other years occurred that coincided with the rainy season in this area of Viet Nam[214] although interestingly all recent putative introduction events appeared to occur outside of these peak years. This was a pattern that was also reported across the whole of Viet Nam, and conforms to the cyclic nature of DENV infections that has been reported elsewhere, with cases peaking roughly every 3 years[11,215]. Our results suggest that some of these recent increases in cases may have been related to the recent introduction of two new D2 clades into the population (Clades 2.I and 2.II), although further serology work would be required to confirm this effect and the effects of increased and improved sampling and diagnosis cannot be excluded.

The successful establishment of a new viral lineage has been suggested to be dependent on the historical exposure to different DENV serotypes of a population, potentially generating an immunological niche of more susceptible individuals to the new serotype, with previous studies showing that novel serotype introductions into a naïve population are more likely to become established, and cause greater case numbers, than those for which there is already a serotype in long-term

circulation[105,216]. For establishment of viral lineages of a similar serotype to historical exposures, previous investigations have suggested that positive selection for epidemic potential may have occurred, which could potentially explain the increase in case numbers. In our analysis of both pervasive and episodic changes on the genome, we detected a mixed picture, (Supplementary Table 3), with Clade 1.1 showing some evidence of selection associated with the later increase in numbers, while clade 4.1 did not. Given that the numbers of all three main clades were observed to increase simultaneously during this later period, it seems likely that environmental factors, such as the number of mosquitos, were more important than virological adaptation.

The low case numbers sampled at certain times of the year, with prolonged periods of weeks and even months where no clade members were sampled, could reflect successful control strategies. However, phylogenetic inference has now enabled us to show the subsequent re-emergence of the same clade at later dates, with the expected low number of observed mutations accumulated during the missing time period. The virus therefore continues to be cryptically transmitted in the population and is likely not detected due to a lack of sampling during this period. Sampling is generally lower during periods which also coincide with unfavourable conditions for the vector[28]. Such a period of low transmission, aggravated by a high prevalence of underlying asymptomatic and mildly symptomatic transmission (estimated to compromise up to 84% of total DENV infections in a population[92]), is suggestive of unsampled sub-clinical infection as the most likely explanation for viral persistence of these clades.

Our data also provides evidence for viral persistence despite attempts of local control, although continued re-importation of the same viral clade, probably from nearby areas, is also possible. We suspect this may have happened with clades 1.I and 4.I, where we see distant ancestral branches between subclades (Figure 6). This implies missed sampling in the local area over the many years since the predicted tMRCA of these clades (Figure 4.5). This highlights one of the limitations with this study where incomplete sampling of the total infected population will likely cause us

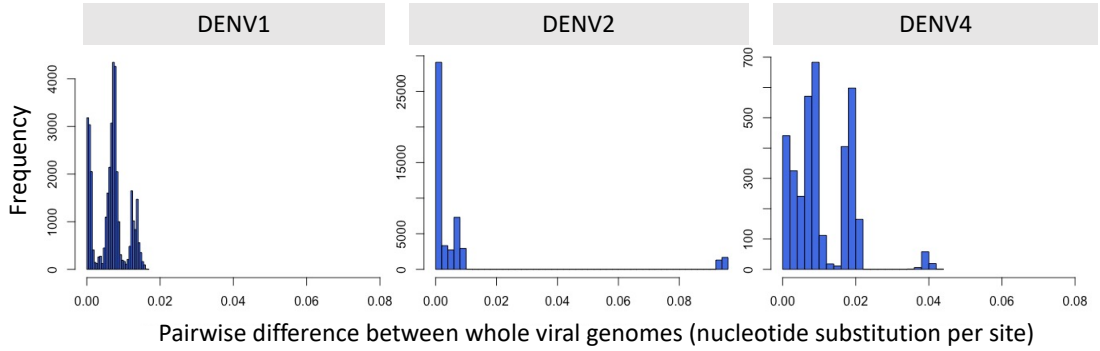
to not fully capture the total viral diversity of each clade, which can cause an underestimation of the calculated tMRCAs, and in turn of the time of introduction of the observed lineages. However, despite this we still observe prolonged periods of persistence in the population.

These findings have implications for future control strategies for DENV, as we have shown that there is continued, often unreported DENV transmission year-round. Vector control could therefore be implemented all year, as well as during peak periods as is currently favoured[217]. The cyclic nature of DENV case numbers, with peak outbreak periods occurring approximately every three years in this population[31] indicates a role for increased surveillance and control measures during outbreak years and potentially the use of vaccines or prophylaxis if available[42]. The discovery of imported cases into the area from other regions of Viet Nam as well as from abroad, one of which rapidly became the dominant circulating lineage, highlights a potential role of border sampling of international visitors particularly from DENV endemic regions with a high prevalence of unseen DENV serotypes (such as D3 of which very few cases were observed in this study period).

4.5.1 Conclusions

We have identified epidemiological and phylogenetic evidence of uninterrupted transmission chains that have lasted for the 3 years of our sampling period, and through the use of TreeTime we have inferred that some of these transmission chains may have persisted for more than 10 years. During this persistence, infected cases per year have increased and decreased, implying that environmental changes (which highly influence mosquito numbers) are more likely to be the cause of observed changes in human infections, rather than the introduction of clades with specific viral adaptations. In addition, our data also imply that anti-dengue strategies had only limited success in preventing transmissions, even in the low seasons or non-outbreak years, possibly due to the importance of unreported (potentially sub-symptomatic) cases.

Supplementary figures



Supplementary Figure 1: Distribution of raw pairwise differences between sequenced genomes across DENV1,2 and 4

Supplementary Table 1: Distribution of average p-distances among sampled clades

Clade	Size	Nucleotide substitutions/site ($x10^{-3}$)	Standard Error ($x10^{-3}$)	Genome Length (base pairs)	Average mutation sites/genome
1.1	173	9.68	0.52	10415	100.85
1.2	12	0.44	0.098	10415	4.57
2.1	214	2.46	0.18	10411	25.61
2.2	7	9.68	0.52	10411	100.81
4.1	95	9.80	0.49	10594	103.83

Supplementary Table 2: Summary of clade persistence findings from both sampling data and estimated tMRCA inference

Serotype	Clade	Sampling dates			tMRCA (90% CI)			
		First	Last	Duration (days)	Estimated root date	Min	Max	Duration (days)
1	I	28/11/2016	25/04/2019	878	03/2009	07/2006	11/2011	3707
	II	15/06/2017	10/12/2018	543	06/2017	02/2017	08/2017	557
2	I	19/07/2018	26/04/2019	281	06/2014	07/2013	05/2015	1790
	II	10/09/2018	09/01/2019	121	06/2018	03/2018	09/2018	222
4	I	17/02/2017	24/04/2019	796	09/2008	05/2007	01/2010	3887

Supplementary Table 3: Positive selection analysis of clades which spanned the full study period. MEME and FUBAR analysis have been performed with sites with significance values less than 0.05 and greater than 0.9 shown. The distribution of amino acids before and after the elevated case number period are shown.

Clade	*AA location*	*Viral protein*	*MEME (p-value)*	*FUBAR (probability)*	*Percentage before elevated cases*	*Percentage during elevated cases*
1.I	183	Membrane Glycoprotein Precursor	NO	YES (0.978)	100% Alanine	37.3% Alanine : 62.7% Threonine
1.I	878	NS1	NO	YES (0.961)	100% Isoleucine	63.2% Isoleucine : 36.8% Valine
1.I	375	Envelope	YES (0.01)	YES (0.931)	61.1% Phenylalanine : 38.9% Leucine	100% Phenylalanine
4.I	3164	NS5	YES (0.01)	YES (0.96)	83.3% Cysteine : 16.7% Histidine	100% Cysteine
4.I	174	prM	YES (0.03)	NO	100% Threonine	97.2% Threonine : 2.8% Isoleucine
4.I	670	Envelope	YES (0.04)	NO	95.8% Threonine : 4.2% Serine	98.6% Threonine : 1.4% Isoleucine

Acknowledgements

We would like to thank those who participated in the study and the health care personnel who collected data from patients at the Pasteur Institute of Nha Trang and the Hanoi National Institute of Hygiene and Epidemiology for their work in collecting and managing data.

Funding

Financial support for this study were provided by the United Kingdom Medical Research Council (www.mrc.ac.uk) as well as Janssen (Johnson & Johnson, Project number: ITPMZG4810), support for the population study site was provided by the Japan Program for Infectious Diseases Research and Infrastructure, Japan Agency for Medical Research and Development (AMED) under Grant Number: JP21wn0125006. O.J.B. was funded by a Sir Henry Wellcome Fellowship funded by the Wellcome Trust (206471/Z/17/Z).

The funders had no role in study design, data collection and analysis; decision to publish; or preparation of the manuscript.

Competing interests

I have read the journal's policy and the authors of this manuscript have the following competing interests: MVL, GHT, FR and LVM are employees of Johnson & Johnson.

Data Availability

Processed FASTA files have been deposited at the National Center for Biotechniology Information (NCBI).

Chapter 5 A mixed methods study of dengue virus transmission in urban Vietnam

5.1 Abstract

Background

Human infection by the dengue virus is transmitted by the mosquito vector Aedes aegypti, which is thought to reside in close proximity to affected cases. However, identifying viral transmission chains, which are critical to preventing transmission, is complicated by the difficulties in tracking infected mosquitos and the large number of asymptomatic dengue cases that do not report for diagnosis. In the study presented here, we have used a community-based household survey to identify symptomatic and asymptomatic dengue cases; and used viral whole genome sequencing to identify transmission chain pairs, to determine likely sites of vector-born infection.

Methods

We identified 563 households with individuals having symptomatic infection (index cases) within 4 communes of Nha Trang Vietnam and 35 of their dengue positive but pre-symptomatic household contacts (HHCs), using NS1 rapid tests. From these participants, 551 qPCR positive dengue infections were identified, and from these we obtained 511 DENV whole genome sequences. Clusters of related viruses were first generated based on their phylogenetic similarity, using a threshold of a maximum of 1 nucleotide difference between genomes. These clusters were then further refined into transmission pairs according to the symptom's onset date, using the program 'outbreaker'. We identified pairs as being infected in parallel (presumed bitten by the same already infected mosquito), sequential (with a mosquito infection cycle occurring between cases) or ambiguous (where we could not distinguish between parallel or sequential), according to the predicted dengue infection periods within the mosquito. We then investigated the transmission pair's most likely exposure

location, using home and daytime addresses. From 365 of the Index households, 805 HHCs were also studied by paired serology to investigate whether serological markers could be used to capture transmissions missed by our viral sequencing strategy

Principal Findings

Viral genome transmission analysis identified 89 transmission pairs from 141 individuals (as both parallel and sequential infections can occur from the same individual). The majority of the pairs were parallel infections that occurred outside of the index case household. Distance analysis suggested that most transmission pairs were infected within the 500m threshold that the mosquito vector is estimated to fly in its lifetime, although most pairs were not within a household (suggesting neighbourhood transmission). We also identified 19 cases where transmission was likely to have occurred at the given daytime address, rather than a home address, implying human movement as important mediators of the viral dispersal seen in the study location. Our study did not detect sequential transmissions within the household, probably because of infrequent sampling of HHCs. However, serological analysis of HHCs suggested that this may have occurred, although at least half of HHCs had different infecting serotypes from their Index cases, supporting our virological findings that neighbourhood and daytime locations were driving most of the infections.

5.2 Introduction

Dengue virus (DENV) is associated with one of the largest burden of arthropod-borne viral infections, with roughly 2.5 billion people at risk of infection[1] mainly in south-east Asia, South America and the Caribbean. An estimated 400 million infections occur each year, with a mortality rate ranging from 5-20%[3] for the most severe forms (dengue haemorrhagic shock).

Despite being considered a day-biting mosquito *Aedes aegypti,* is most active during the hours of dawn and dusk[218]. It is also considered as having a short flight range, especially in urban environments, with the blood-feeding females flying not more than 100-500m from their emergence sites[219–221]. Thus, DENV transmission is believed to mainly occur in a small radius, with within-household infection being the primary route of transmission[189–191] , although direct evidence of this is rare. Indeed, investigating the transmission dynamics of vector-borne diseases is complicated by the addition of the vector phase of the disease, which necessity in-depth knowledge of the behaviour and characteristics of three different organisms: the virus, the mosquito and the human host . This makes traditional techniques such as contact tracing difficult to conduct, a difficulty further enhanced by the observation bias resulting from asymptomatic DENV infections.

Figure 5.1: Comparison of contact tracing maps between a respiratory and vector-borne disease when an infected individual walks from their household to school. With a respiratory infection range of 2m, it is straightforward to distinguish between the school and the home as infection site, but if the two sites are within 500m, an infected mosquito may have flown between the two sites

Whole genome sequencing of pathogens from clinical samples has emerged as an effective tool for describing transmission clusters and can also provide supplementary evidence of transmission chains to inform researchers on how a virus spreads[106,107]. DENV has been classified into four serotypes (DENV-1, DENV-2, DENV-3, and DENV-4), based on antigenic properties and genetic differences, which can reach up to 40% across the full genome[222 223]. This level of genetic diversity, both with and between serotypes, makes DENV an attractive candidate for molecular epidemiology studies. Yet this has not been applied comprehensively to dengue virus, perhaps partially because of the difficulties of including the mosquito vector and the difficulties of sampling dengue infections sufficiently well when many are unreported or asymptomatic cases.

As DENV is principally spread by A. *aegypti*, a vector that can take a single blood meal from multiple people, there can be a complex transmission cycle which can range from, multiple people being exposed at the same time by the same mosquito (parallel exposure), to around 38 days between cases if the maximum incubation and lifecycle times of the virus and mosquito are taken into account (sequential transmission)[224–227]. This is illustrated in Figure 5.2.

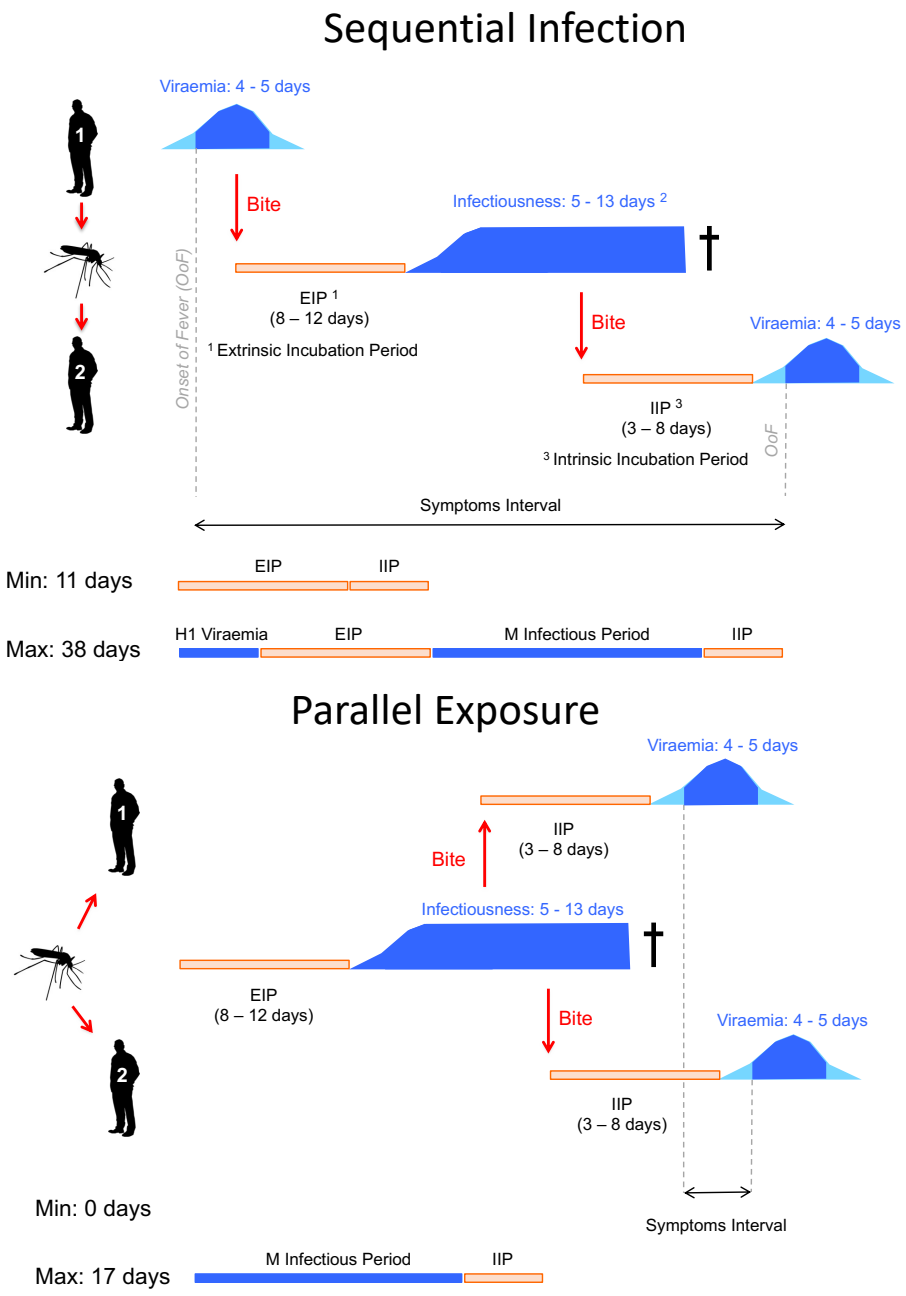


Figure 5.2: Estimates of human and mosquito infection periods are used to distinguish different infection profiles of sequential and parallel infections involving human hosts and the mosquito vector. When a suspected transmission occurs in humans with a symptom interval between 11 and 17 days, it is ambiguous whether it can be called a parallel of sequential infection.

Dengue virus is commonly diagnosed using serological tests for antibody status as well as viral detection assays for antigen or viral genomic nucleic acid presence. As the 4 serotypes of DENV are highly cross-reactive to antibody responses nucleic acid detection is the most reliable method for serotyping of active DENV infection, whilst convalescent patients must be serotyped using plaque reduction neutralisation assays (PRNTs) of their sera.

Following infection with DENV, serological responses differ depending on prior history with the disease. Primary exposures illicit a typical serological response with an initial IgM response which wanes over the next 6 months, followed by a typically long-lasting IgG response which increases in titre following clearance of the infection. Subsequent infections with a different serotype are associated with a reduced IgM response and an increase in the titre of previous IgG responses which are thought to be non-neutralising and can lead to an enhanced infection through antibody-dependent enhancement (ADE).

Combinations of various serological assays can therefore be used to identify DENV naïve, active primary infections and active post-primary infections as well as various changes in serostatus. A previous study has developed an algorithm to assist in characterising the serostatus of individuals using IgG, IgM and PCR or antigen ELISAs/Rapid Diagnostic tests in combination with self-reported disease day[228].

PRNTs are a valuable tool for serotyping DENV infections. There are several factors which are known to limit their reliability, the presence of IgM is known to induce a significant cross-reactive neutralising response, the assay is also known to be more unreliable with more DENV infections a person has recovered from, with significant cross-reactivity between all 4 serotypes with each successive exposure. They are therefore most useful in individuals who are experiencing their second dengue infection, while having had their first DENV exposure more than 6 months previously and so no longer have anti-DENV IgM present but have sufficient viremia to enable molecular typing of their second infection. qPCR during acute viraemia is the most

accurate method for serotyping DENV infection but this is generally only possible when cases report in the first few days of fever.

Here we present data from a large-scale study into household DENV transmission in Nha Trang city, Vietnam with the transmission dynamics of the virus in the population assessed from the immunology, viral sequencing, and epidemiological data.

5.2.1 Chapter Aims

With many viral and bacterial infections, being in close contact, such as living in the same accommodation as an infected individual, increases the risk of exposure to the pathogen. However, for vector-borne diseases, especially those transmitted by a day-biting-flying vector, the exposure risk within a household is unclear. One of the original aims of this study was therefore to characterise the amount of intra-household DENV transmission that occurred in this endemic urban study area. I then sought to determine whether other routes of transmission in the population as a whole could be determined. In order to investigate these, I sought to combine genomic qPCR and sequencing data with serological and epidemiological data of presenting individuals and household contacts to understand how the virus transmits between the population. Therefore, the aims of this chapter are to:

i) Use collected serological and epidemiological data to explore identified household exposures
ii) Identify putative transmission pairs and exposure clusters within households and the general population using generated viral genomic sequences combined with epidemiological data and investigate their likely exposure locations

Contributions

Sample collection, storage and shipment were conducted as previously stated by the team at Pasteur Institute, Nha Trang, Vietnam. Plaque-reduction neutralisation assays were performed by the National Institute of Hygiene and Epidemiology in Hanoi, Vietnam. All other analysis unless otherwise stated were performed by me.

5.4 Methods

5.4.1 Study cohort

Enrolment of the study cohort was conducted in the communes of Vinh Hai, Vinh Phuoc, Vinh Tho, and Vinh Hoa in Nha Trang City, Vietnam, after local ethical review and approval from the Vietnamese National Institute of Hygiene and Epidemiology (study approval number IRB-VN01057). Individuals of all ages, who resided in the selected communes and who visited either a local polyclinic (between October 2016 and May 2019) or the city's Tropical Medicine Hospital (between December 2016 and April 2019) with suspected dengue were deemed eligible. In the polyclinic, those who presented with suspected dengue fever were approached, and those who consented to be enrolled in the study were subsequently tested for dengue with an NS1 rapid test (DENGUE NS1 AG STRIP, #70700, BioRad). Individuals who tested positive were enrolled as index cases. Of those who tested negative for NS1, every tenth subject was asked to return for a second test 30 days later to act as controls. In the hospital, only those who were diagnosed with dengue fever on the basis of an NS1 rapid test were enrolled in the study.

Following enrolment, study teams visited the index case's home addresses, where consenting members of the household were also enrolled and had blood samples taken for dengue testing. After consent was obtained, a questionnaire was completed by all enrolled individuals in order to gather information on clinical, demographic and behavioural information relevant to this study, such as age, time of fever onset, home GPS coordinates, recent travel history and typical daytime location. All information was anonymised prior to the analyses.

5.4.2 Phylogenetic analysis

FASTQ files produced from Illumina Miseq runs as described in chapter 4, were processed using the following pipeline in order to generate consensus sequence of the target virus. Raw files were screened for quality and trimmed to remove primer sequence bias using the program fastp[146]. Cleaned FASTQ files were then aligned to

a reference using the Burrow-Wheeler Alignment for short-reads (BWA-MEM)[147] along with the SAMtools view and sort packages[148]. Reference DENV genomes were from GenBank: DENV1: JQ045626, DENV2: GU131898 and DENV4:NC_002640. Aligned reads were then called using the SAMtools mpileup and iVar consensus[149] programs with a resulting consensus sequence given in a FASTA format.

Processed sequences were aligned using MAFFT[143] against the available full-length DENV genomes downloaded from the Virus Variation Resource database provided by NCBI. The resulting alignment was then checked and manually adjusted in AliView[208] to remove sequences had more than 10% of their genomes missing. Maximum likelihood phylogenetic trees were generated using IQTREE[209] with the model finder tool to select the best fitting model of nucleotide substitution and 1000 bootstrap resamples to estimate branch supports. Resulting trees were visualised and edited using FigTree with rooting at the mid-point and tips arranged in ascending order.

5.4.3 Identifying transmission pairs

In order to identify likely transmission pairs from the phylogenies, the following workflow was applied to the generated sequences (summarised in Figure 5.3): The program TreeCluster[229] was used to identify clusters of genomes that shared a common ancestor and differed by a maximum of 1 nucleotide across the whole sequence. These clusters were considered to represent recent dengue transmission events amongst the participants represented by the viral genomes. Clustered Individuals whose transmission times were outside the expected time for a sequential DENV transmission (greater than 38 days between symptom onset) were excluded from the putative transmission chain. For clusters containing more than 2 individuals, the program outbreaker2[230] was used to reconstruct transmission chains within a cluster, determine the most likely infector/infected relationships between the members of the cluster and exclude from the analysis those members with serial intervals exceeding the estimated maximum duration of 38 days. Gamma distributions of shape 5 and scale 0.4 for parallel infections, and of shape 3 and scale 4 for sequential were used to model dengue virus incubation period generation time, respectively, in the transmission chain inference (these values were acquired based

on the calculated transmission times between the two infection categories). Only infector/infectee pairs inferred with a confidence >90% in a cluster were selected for further analysis.

The putative transmission pairs thus identified were characterised as either parallel, ambiguous or sequential based on the serial interval of the linked cases. Pair members with symptoms onsets intervals of less than 11 days were likely exposed at the same time by the same mosquito and were characterised as being exposed in parallel (see Figure 5.3). Pair members with symptoms onsets intervals of more than 17 days (but less than 38 days, as described above) were characterised as being infected through sequential transmission. Pair members with symptoms onsets intervals between 11 and 17 days were classified as ambiguous. It is possible for an individual to be part of more than one transmission pair. This led to transmission pair members being counted in several chain of infections, reflecting the possibility that a human can be bitten by multiple mosquitos.

The most likely location of transmission was determined by taking the shortest given distance between the pairs according to their household and daytime locations. For instance, if two individuals shared the same home address but had different daytime address, the household was considered the most likely location of infection.

5.4.4 Identifying exposure to DENV among household contacts

Household contacts of index cases who had tested positive, were sampled twice within 7 days of their index case enrolling and again 30 days later.

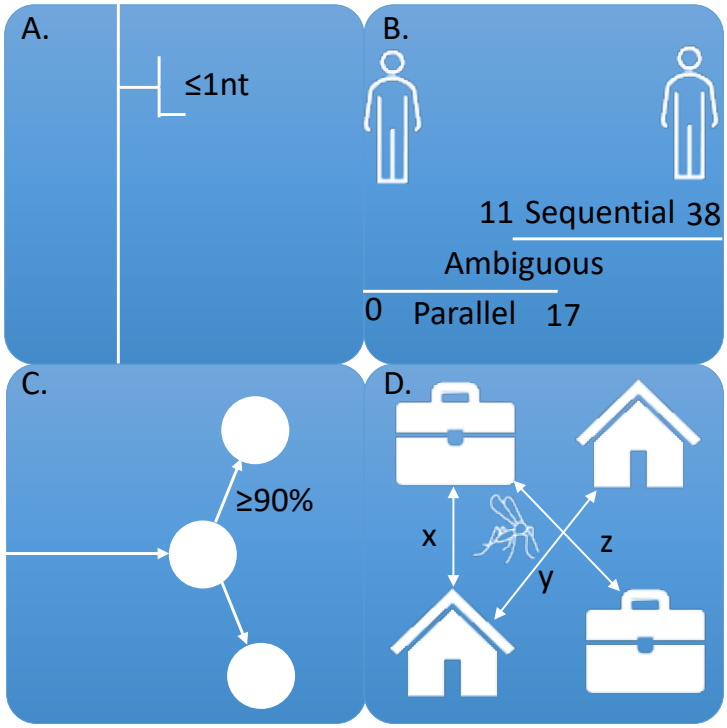


Figure 5.3: Workflow for identifying putative transmission pairs from whole genome viral sequences. A. Treecluster was used to identify clusters of individuals whose virus was within 1nt between aligned sequences. B. Identified clusters were sorted according to the biological probability based on the combined lifecycle of the mosquito and viraemia profiles into sequential transmission, parallel exposure and ambiguous between the two. C. Identified clusters which contained biologically possible tranmissions were sorted to infer who infected whom using outbreaker2 with pairs containing a support of at least 90% chosen. D. In order to establish the most likely exposure/transmission location, the closest distance between home and given most frequent daytime location was measured.

5.5 Results

5.5.1 Recruitment

During the study period 2,287 patients presented and were enrolled at either a participating hospital or clinic. From these enrolled subjects, a total of 838 index cases tested positive for DENV by NS1 rapid antigen test and were invited to enrol their household contacts, with subjects from 563 index case households consenting. We obtained samples from 1241 household contacts at first visit and 861 at their second follow-up visit, at an average of 30 days later. None of the household contacts reported to a clinic or hospital with dengue like symptoms during their enrolment period, although 84 reported experiencing fever during their study period on their questionnaire and 38 tested positive to DENV by NS1 test.

I tested all NS1+ index and household contacts for the presence of dengue virus by qPCR and identified 894 positive subjects. From these DENV positive subjects, 701 samples were identified as having sufficient viral RNA to attempt whole genome sequencing (PCR Ct value < 32) with 511 of them successfully whole genome sequenced. A summary of the demographic and serological data collected from index cases and their associated household contacts is shown in Table 5.1.

Table 5.1: Characteristics of the study participants. Case numbers and resulting serological assays of index cases and associated HHCs with total numbers of households surveyed given (HH). HHCs were visited twice with an average of a 30-day interval

	Enrolled Index cases (n = 838)	Household contacts Visit 1 (n =1241)	Household contacts Visit 2 (n= 861)
Average Age	**18.7 (0-85)**	**33.2 (0-90)**	**34.1 (0-88)**
Gender:			
Male	**483 (57.6%)**	**563 (45.4%)**	**382 (44.4%)**
Female	**355 (42.4%)**	**678 (54.6%)**	**479 (55.6%)**
Average disease day	**3.4 (1-7)**		
Diagnostic data:			
NS1+	**838 (100%)**	**35 (2.8%)**	**3 (0.3%)**
NS1-		**1206 (97.2%)**	**858 (99.7%)**
IgG+	**486 (58%)**	**1068 (86.1%)**	**765 (88.9%)**
IgG-	**197 (23.5%)**	**169 (13.6%)**	**92 (10.7%)**
Untested IgG	**155 (18.5%)**	**4 (0.3%)**	**4 (0.3%)**
IgM+	**284 (33.9%)**	**142 (11.%)**	**105 (12.2%)**
IgM-	**398 (47.5%)**	**1033 (83.2%)**	**722 (83.9%)**
Untested IgM	**155 (18.5%)**	**66 (5.3%)**	**34 (3.9%)**
Disease presentation:			
Dengue Fever	**716 (86.4%)**	**68 (5.5%)**	**5 (0.6%)**
With warning signs	**121 (14.4%)**	**16 (1.3%)**	**2 (0.2%)**
No symptoms	**1 (0.1%)**	**1157 (93.2%)**	**85 (99.2%)**

During the study period and among the 805 HHCs with serology sampled at both visits, 101 showed evidence of active disease at the first visit (12.5%). Further serological analysis identified that 18.7% of historical cases and 31.3% of negative household members experienced an infection between visits 1 and 2 (Table 5.2). Therefore, household members who were negative at V1 were more at risk of DENV infection compared to historically infected members (Risk ratio 1.7 [1.2 – 2.3], as previously shown in [4]). Overall, 20.5% of all household members of NS1+ve index cases capable of experiencing a subsequent infection (negative and historical at V1), did seroconvert, compared to 9.9% of returning NS1- Index case controls, (Risk ratio: 2.07 [1.58-2.56]) (Table 5.2).

Table 5.2: The number of household members that experienced dengue infections in the households of reported NS1+ index cases, compared with NS1 -ve index cases returning after 1 month (acting as controls).

Index Cases	DENV Immune status of household member		Proportion of seroconversion	
		N	%	n
NS1+	Historical	605	18.7	113
	Negative	99	31.3	31
	Combined	**704**	**20.5**	**144**
NS1-	Historical	111	8.1	9
	Negative	31	16.1	5
	Combined	**142**	**9.9**	**14**

5.5.2 Transmission chain analysis

Using full genome sequence data generated as previously, I then sought to identify transmission pairs who were likely infected by the same mosquito. The threshold of 1 SNPS between whole DENV genomes yielded 72 clusters of 201 individuals across the 3 serotypes analysed, with the distribution of individuals within each cluster across the 3 serotypes analysed shown in Figure 5.4, where we can see the majority of clusters contain a pair of individuals (48/72 (67%)). After removing individuals who declared symptoms more than 38 days apart and running outbreaker2 on clusters with multiple individuals, we had 90 putative transmission clusters of 143 individuals remaining. This increase in number of clusters is because an individual can be involved in 3 different transmission events – parallel, ambiguous and sequential, so can be counted multiple times in different clusters.

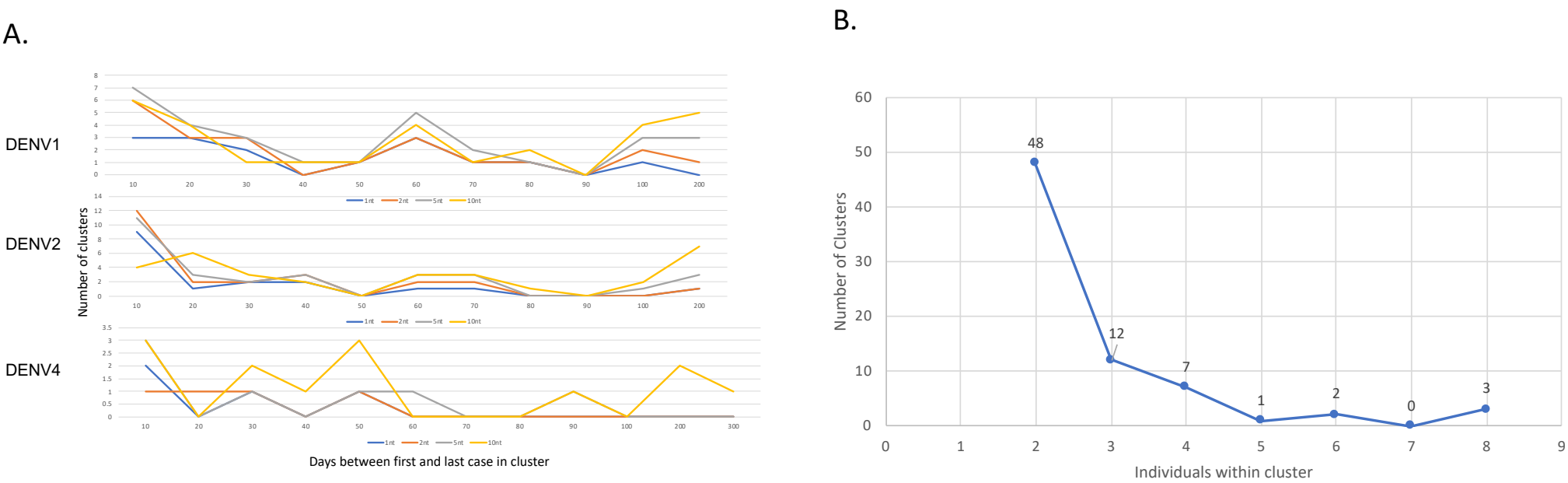


Figure 5.4: Optimisation of SNP threshold for determining clusters. A. Comparison of SNP thresholds with time distributions between individuals within identified clusters. B. Distribution of cluster sizes identified using TreeCluster with an SNP difference threshold of 1nt.

These were then characterised as parallel, sequential or ambiguous based upon the time differences between symptom onsets. I identified 49 parallel, 22 sequential and 18 ambiguous transmissions (Table 5.3). 8 of these transmissions occurred between subjects of the same household, that were all parallel transmissions involving their index case (Table 5.3). The lack of ambiguous or sequential transmissions reported within the same household may be a consequence of the lack of sampling over the 1-month enrolment period, as additional sampling between standard timepoints required dengue symptoms to be reported to initiate them (and no HHC reported dengue symptoms to a clinic). However, to investigate what happened between sample points further, we explored these households using serological analysis (further below).

I next investigated the likely exposure locations of the identified clusters using the data given for daytime and home address locations. This showed that 42 transmissions likely occurred between households, given that the household addresses of the transmission pairs were closer than their given daytime addresses (Table 5.3). Indeed, the median distance between households forming transmission pairs was 182.3m, which is similar to the commonly reported 150m distance used to identify clusters in Singapore[231]. However, parallel exposures had a significantly shorter distance (median 153.6m) than sequential transmissions (median 470.8m) ($p<0.05$), suggesting that there may be mosquito migration over time. While all pairs identified were parallel infections involving index cases (n = 49), many transmission pairs showed that members of the same household were infected with different viruses from different transmission pairs, implying that multiple infection sites were responsible.

I also identified that 4 pre-/asymptomatic household members formed pairs with someone outside of their household and the remaining 7 potential within-household transmission pairs, for which we had sequence data for both individuals, did not form clusters together and did not meet the requirements for forming a transmission pair with any of the other sequences generated.

Analysis of the 22 pre-/asymptomatic household members who formed transmission pairs, also showed that one of them (who was asymptomatic) formed a sequential transmission pair with a symptomatic case who enrolled 21 days after the pre-/asymptomatic household member gave their sample. The most likely infection site for this transmission was near the household location of the asymptomatic case (which was also given as their most frequent daytime location) and the daytime location of the symptomatic case, with the distance between these being 204.6m (in contrast to the distance between their households, which was 991.1m apart).

We found that daytime locations were involved in 39/89 (43.8%) transmissions as compared to 70/89 (78.7%) that involved household locations with 39 pairs involving a combination of daytime locations or both daytime and household against 50 involving just household transmissions. Suggesting that both daytime locations (household or other) can be a risk factor for acquiring the disease and that there is a risk of transmission to residential areas that are close to locations which people may migrate to during the day, such as schools or commercial districts.

Table 5.3: Distribution of transmission pairs identified characterised by time interval and optimal distance between cases. Average distances are given as the median.

Closest location between pair	Parallel	Sequential	Ambiguous	Total	Median distance (m)
Between Household	22	7	12	42	182.3
Within household	8			8	0
Between Daytime	5	9	5	19	607.6
Between Household and Daytime	13	6	1	20	494.6
Total	49	22	18	89	
Median distance (m)	153.6	470.8	519.7	347.6	

I next investigated the actual distances between identified transmission pairs to determine how likely observed transmission events were, given the estimated 500m maximum flight range of the principal mosquito vector[220], using the shortest distance between either the household or most frequent daytime addresses of the pair members, the distance between their home addresses or the distance between their daytime address (Table 5.3). I found that transmission between households were the most common route of infection sampled (51 out of 89 pairs), with primarily parallel infections between multiple people often being exposed at a similar time in households which are within the flight range of a mosquito. Sequential infections were most likely to occur between daytime addresses (n=9/22). This indicates a role for daytime transmission in transmitting the virus across an urban population, whilst households most likely provide large clusters of localised infections from a small number of hosts (Figure 5.5).

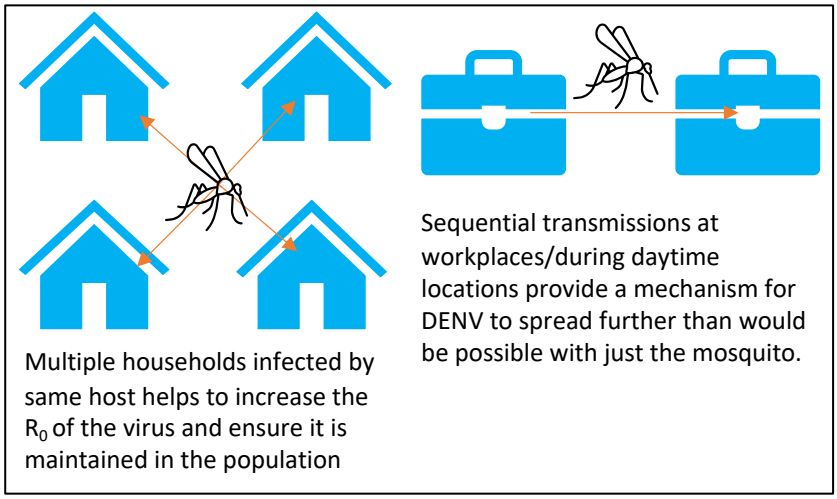


Figure 5.5: Key findings from analysis of identified transmission pairs and distances between them

From examining the range of distances found (Figure 5.6B) we can see a distinct clustering of distances within the estimated flight range of a mosquito in its lifetime (<500m). The average distance between parallel exposures is significantly lower than was found with sequential infections (P=0.0062 using Mann Whitney test) (Figure 5.6A). The median distances found for most categories analysed (Table 5.3) were additionally within this distance threshold with the exception of transmissions occurring between daytime locations (607.6m).

Despite the overall median transmission distance being 347.6m, there are significant outliers for all transmission categories which cannot be explained by mosquito transmission between the hypothesised transmission sites as the distance is far beyond the vector flight range. Indeed, many putative parallel infections are over a very large distance when using the 3 potential exposure locations this could imply a role for human movement in spreading the disease further than the mosquito is able to and that more daytime locations of participants are required to be sampled.

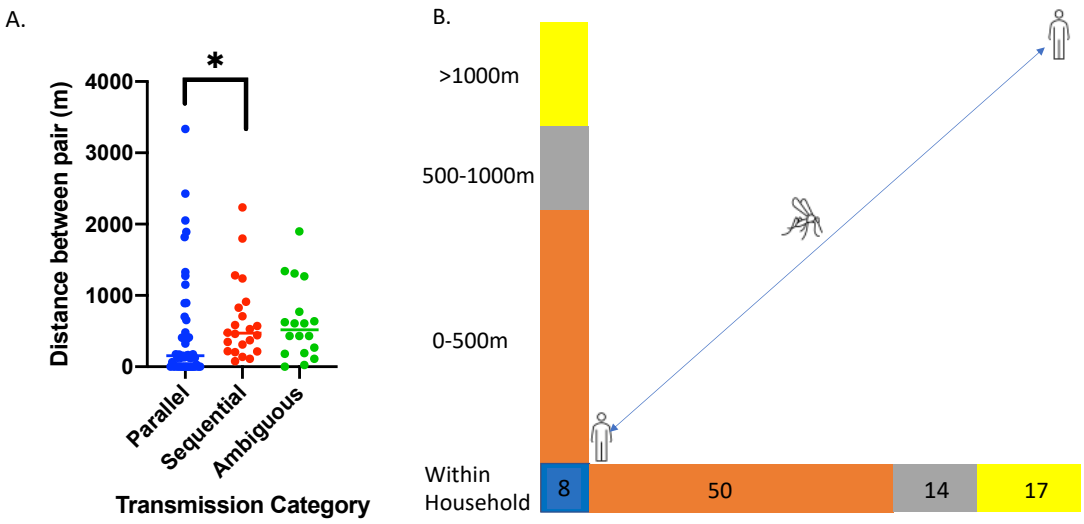


Figure 5.6: Distance analysis of transmission pairs identified. A. Plot showing the breakdown of distances between identified transmission pairs according to their transmission category, with the median value highlighted. B. Distances between transmission pairs stratified by distance.

Mapping and comparing the household and daytime locations of identified transmission pairs shows that while the sampled population live over a wide area of the study site and the given daytime locations are over an even wider range, with some being outside of the city limits (Figure 5.7). There are distinct clusters where higher densities of the study population reside and others where they spend the daytime hours, these are generally separate, however there are several examples where these locations overlap, which could have provided an opportunity for viral migration. We found one example of an asymptomatic individual who was implicated in sequential transmission to a patient who subsequently presented at a hospital or clinic. Interestingly, the most likely transmission route (based on shortest possible given distance between the two) for this patient was between the household of the asymptomatic patient (which they have also given as their most common daytime location) and a daytime address for the symptomatic patient (Figure 5.7). This suggests that pre- or asymptomatic DENV infected individuals are involved in viral transmissions which I have shown in Chapter 3 that they can have similar viraemia levels, so it is possible virologically.

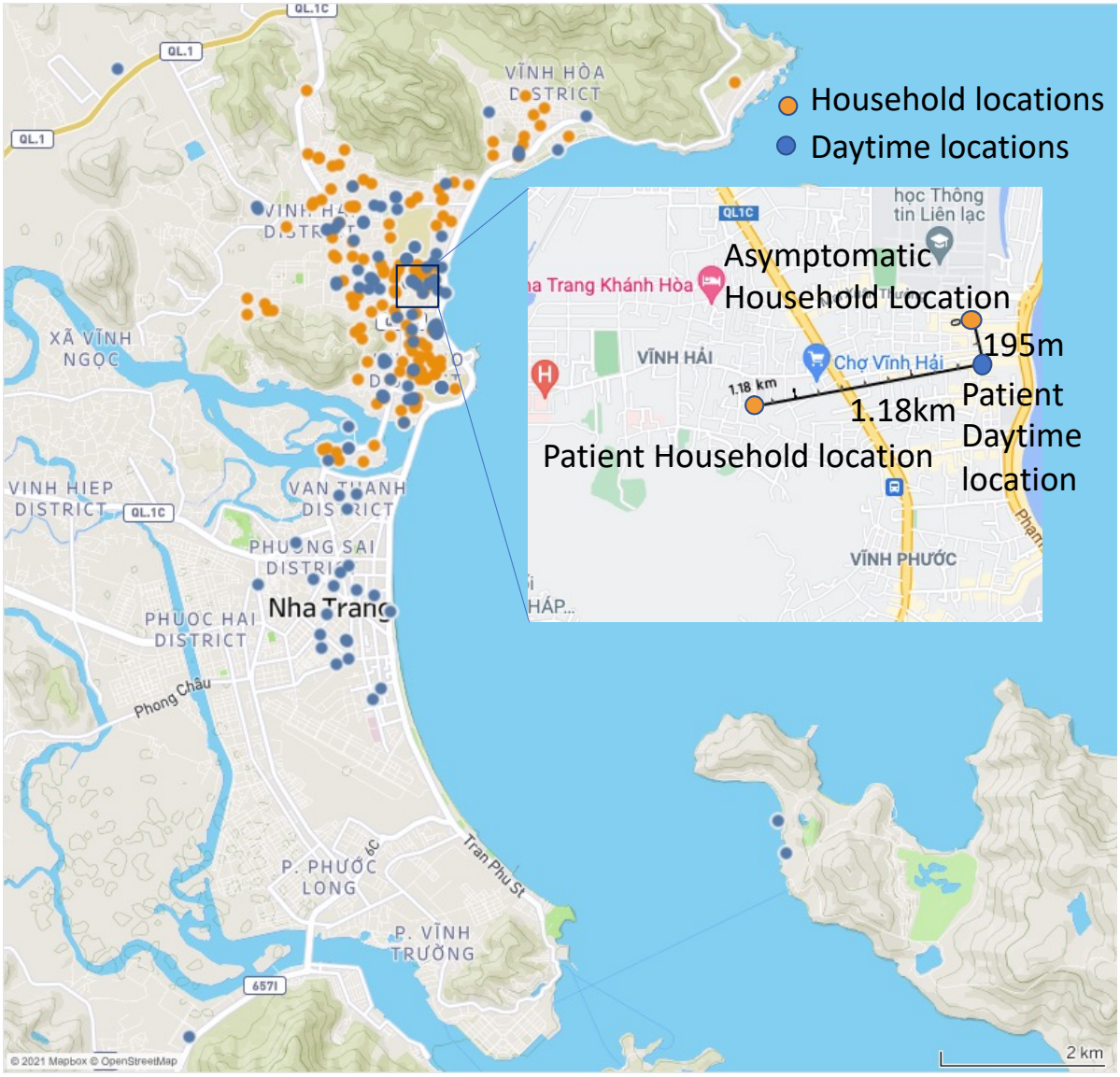


Figure 5.7: Map showing sampled pairs household and given daytime locations. Here we show a sequential infection between an asymptomatic individual and a reporting index case. One daytime location in the city of Cam Nghĩa, 27km south is not shown.

5.5.3 Serological evidence for transmission

As we were not able to observe sequential transmission pairs within households using sequenced genomes, we sought to determine if unobserved potential sequential transmissions had occurred between samplings, using serology. To do this, we first investigated if the Index and their HHCs shared the same serotype response, in a subset of 23 households with their index cases and 81 of their HHCs, using PRNT (Table 5.4). We showed that of the 14 HHCs (from 81) that seroconverted, the antibody response at the second visit from 7 matched their index case serotype (while 7 did not). Of 40 HHCs that did not seroconvert, 31 already had neutralising antibodies to the serotype of the Index case, while 7 remained naive. In 16 cases, the PRNT result was unclear due to cross-reactivity between serotype results, a commonly reported problem with PRNT assays[232], especially when IgM is present (as was observed in 6 cases). However, 5 of these individuals who did not seroconvert to the serotype showed a significant increase in IgG titre between visits and 6 were positive for IgM indicating recent exposure.

5.5.4 Combining serological and genomic data reduces the number of transmissions consistent with intra-household events.

This study initially used a concurrent NS1+ result between an index case and their associated HHC tested within 7 days as evidence for parallel exposure, implying that both were exposed at the same time and location to the same mosquito. However, by applying genomic and phylogenetic data from isolated viruses from these HHC-index pairs I sought to determine whether these represented true parallel exposures (Table 5.5). As DENV has 4 serotypes, a mismatch in serotype between the pairs would indicate that this was not a parallel exposure, indeed when qPCR serotyping was done on the pairs, 6 out of 21 pairs with a successful qPCR serotyping result had such a mismatch, which makes it impossible that they were exposed to the same virus.

Our analysis showed that 6 out of the 23 HHCs with a successful qPCR serotyping result, were infected with a virus of serotype different to that of their index case or

other HH members. This suggests that these infections were acquired outside of the HH rather than within. We also explored to what extend two individuals from the same household infected with the same DENV serotype were truly linked, as opposed to individually infected outside of the HH by different strains of the same serotype. We restricted this investigation to parallel transmissions, because the number within household sequential transmissions we identified was too small. In our diagnostic analysis, 35 HHCs (from 32 households) were positive by NS1 test at the first visit (Table 5.1), suggesting parallel infections with their index case. While this suggested that 17 of the pairs could be within household transmissions, phylogenetic reconstruction analysis showed that only 9 of the 17 pairs which could be whole genome sequenced (52.9%), clustered with each other (Figure 5.8). The remaining 8 did not form monophyletic groups with their index case of other HH members, suggesting infections from independent sources. Altogether, 39.1% of Household members with full molecular characterization were compatible with transmission pairs, compared to the 74% suggested by serology typing. Thus, if sequential transmissions are similar to parallel transmissions, then the PRNT analysis suggesting 50% compatibility for within household sequential transmission (Table 5.4) might be an overestimate of the true number of within household sequential transmissions.

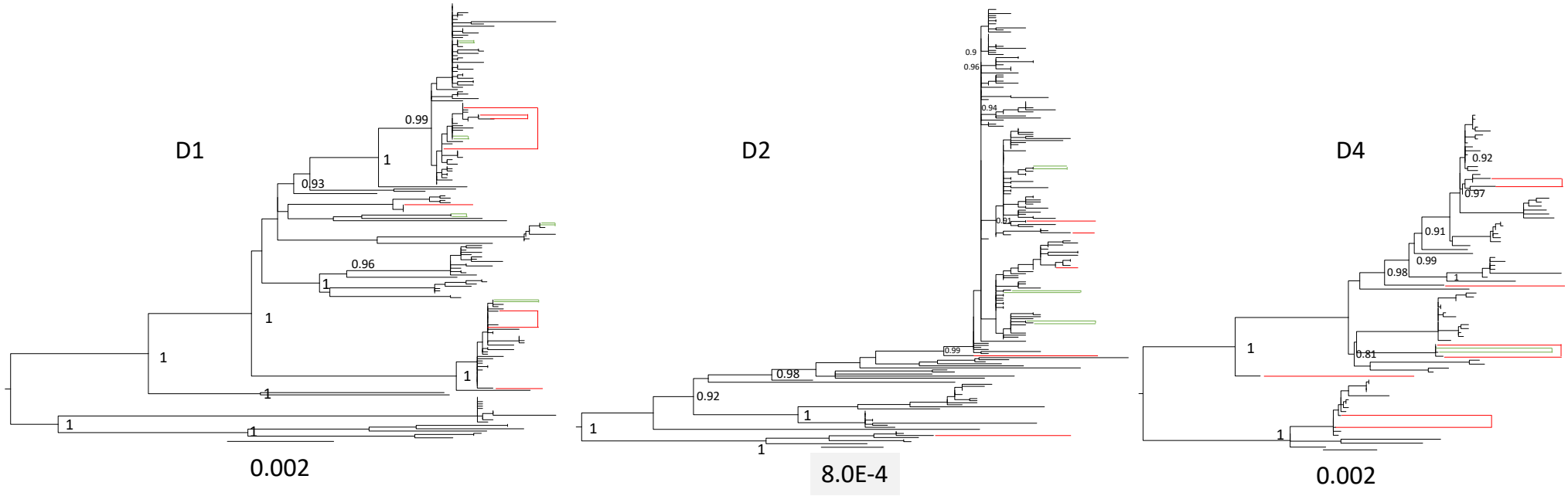


Figure 5.8: Midpoint rooted maximum-likelihood phylogenetic trees for full genome length sequences generated during study for serotype D1, D2 and D4. Branch supports, calculated as bootstrap scores from 1000 replicates, are indicated on relevant branches. Predicted household parallel exposure pairs are show with green connecting lines (9 pairs) whilst those which are not are shown with red lines (14 HHCs).

Of the NS1+ pairs where there was a sequence available for both, and who were the same serotype, only 9/17 clustered with each other (Table 5.5), the remaining pairs (8/17) clustered closer to another individual or were separated by several branches implying their common ancestral sequence was several generations ago. These additional tests significantly dropped the overall number of identified parallel exposure events from the starting 35 observations (Figure 5.9).

Table 5.4: Breakdown of serotype seroconversions observed among HHCs of index cases with a positive qPCR result.

Sequential Seroconversion		**Count**
	Same serotype	7
	Different serotype	7
No Seroconversion		
	Prior immunity	31
	Remained susceptible	9
Unclear Seroconversion		27

Table 5.5: Parallel exposures between participants inferred by various detection techniques.

Inferred parallel exposure	Serology (index & HHC NS1+ visit 1)	Molecular serotype (same as index)	Transmission pair by sequencing (genetic similarity)
Yes	35	17	7
No	0	6	13
Unknown	0	12	15

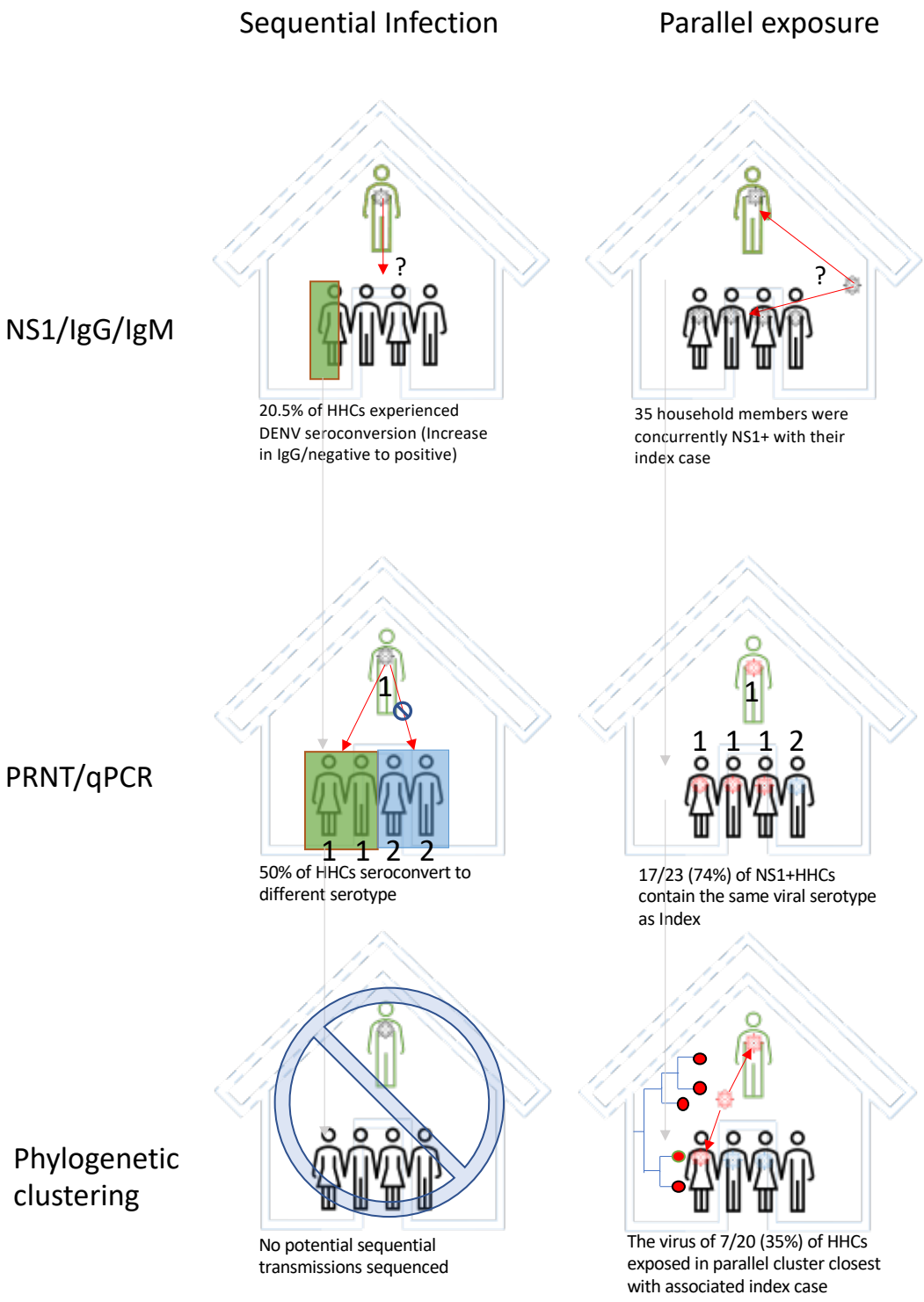


Figure 5.9: Representation of transmission pairs by the different experimental techniques employed. Transmission pairs identified have been divided into parallel and sequential exposure routes to show how different testing techniques affect the total number of observed cases.

5.6 Discussion

This work has used a household survey and DENV whole genome sequencing to identify 90 high confidence transmission pairs that enable a data driven analysis of the complex patten of infection that occurs during dengue outbreaks. Our closest proximity distance analysis suggested that the majority of transmissions occurred between households (47.8%), with transmissions involving out of household, daytime addresses also common (43.3%), whereas transmissions pairs within the same household were relatively rare (8.9%), despite these generally being thought of as the primary location of disease exposure[186]. This has important implications for preventing DENV transmission and suggests that targeting control measures in areas other than just the index identified households, would be more effective.

Previous studies have found it difficult to identify intra-household transmission using viral detection assays and have instead relied on changes to serological status of household members[173,174]. One of the aims of our study design was to look for transmission from index cases to contacts within the same household, as we enrolled household contacts of the index case within three days of the index case reporting and monitored them twice during one month of follow up. However, none of the contacts reported sufficient dengue symptoms to prompt them to report to a clinic, despite 296/840 showing evidence of recent infection or seroconverting during the follow up, 6 reporting fever and 35 being dengue NS1 positive at the first visit and then 3 at the second. This reflects the high proportion of dengue infections that are known not to report to a clinic[92,166] (often classed as asymptomatic, subclinical or non-reporting subjects) and reduced our opportunities to obtain samples from enrolled HHCs outside the serologically determined protocol sampling windows, when we may have obtained more dengue PCR positive serum samples.

At first sampling (<3 days after the index case), 7 of the HHCs were classed as parallel transmissions using the viral sequencing data in combination with our transmission identification method, but as none of the second visit samples were PCR positive, we were not able to detect any sequential or ambiguous transmission pairs within the

same household. We therefore used our serological neutralization assay to investigate a sub-set of those HHCs who sequentially seroconverted between our household visits and found that only half of them suggested sequential infections were to the same serotype as the index case. This suggests that, while sequential infections may have occurred, many may not have been between the index case and their HHC. This is similar to our findings with the parallel infections, where most were confirmed to be between households, rather than within them. Due to the high cross-reactivity associated with the neutralisation used, as well as the relatively high expense, future studies could instead test associated household members more frequently, in order to attempt to capture the viraemic phase, therefore enabling the use of genomic techniques to investigate transmission.

By combining serological, epidemiological and genomic analysis tools, we have increased the confidence of identifying the transmission route of the virus and have further characterised these transmissions by including the life cycle of the mosquito (creating parallel and sequential infections) to gain additional information that could be used by intervention strategies. The thresholds for distinguishing parallel and sequential infections were based on extensive literature that reflects the biting habits of the mosquito, which are known to bite multiple people during their lifetime, and even during a single egg laying blood meal[233], and the well-established infection cycle of DENV in mosquitos and humans. However it is important to highlight that these thresholds are likely to vary, particularly with temperature[226], and have not been studied specifically in our study site, which may mean that more transmissions were ambiguous than we anticipated, although this does not affect our estimates of transmission infection sites.

We also identified that in parallel infected pairs, despite the short times and distances between them, many were not infected in their own household along with the reporting index case, but rather they were more likely infected in neighbouring households or daytime addresses. For sequential infections, where the transmission chain between pairs takes longer to complete, we additionally observed greater distances between addresses, which is consistent with the mosquito moving over

time. However, some infection pairs are clearly outside of the mosquito lifetime flight range and suggest that human movement may account for as much as half of all transmission pairs.

Confirmation of this would require further study, that was able to specifically characterise the now recognisably important subjects' daytime movements. We identified transmission locations using the shortest distances between the possible address combinations reported for the transmission pairs, but as only two addresses were recorded, it is possible that the transmission occurred at another location. More thorough recording of participant daytime activities, GPS or contact tracing using mobile phone data such as has been done with diseases such as SARS-COV2[234] could help identify this potential location.

This study recruited asymptomatic subjects (those dengue cases with mild or no symptoms, that did not report to a clinic), with 144/704 (20.5%) identified by serology as showing evidence for DENV exposure and 26/1241 confirmed by PCR, from which 24/26 whole genome sequences were obtained. While asymptomatic cases are known to be important in the dengue infection burden, their role in transmission chains is less clear. Amongst our subclinical subjects who were PCR positive, the questionnaire identified 18 individuals who reported a fever, which allowed us to compare the quantify the dengue viremia in these subjects' using qPCR. We showed that the viremia of subclinical subjects with and without fever was comparable, and that both were similar to symptomatic cases. This suggests that asymptomatic and subclinical infections are equally likely to transmit as more severe symptomatic cases. Indeed, we identified an asymptomatic case that formed a sequential pair with a subsequent symptomatic case (interestingly, most likely infected at their work daytime address which was near the asymptomatic subject's home address).

A potential explanation for our inability to detect sequential transmission within a household appears to be suggested by the PRNT results. It would appear that a significant amount of HHCs have already been exposed to one or more DENV serotypes, so would be protected against future infections, especially those of the

same serotype as the index case. Previous studies have indicated that exposure to DENV serotypes can produce protective antibodies against other serotypes for many months and lifelong infection to the same serotype[165,168,235]. Limiting the potential spread of the virus especially in highly endemic regions.

When conducting studies to determine exposure routes and locations it is important to use the correct diagnostic test. The results generated in this study showed that different diagnostic tests can overestimate inferred exposure clusters among household members who are positive concurrently. Contact tracing of positive individuals has traditionally been used to infer transmission clusters, however, as DENV has a mosquito vector, this is difficult to achieve via this method. Despite 35 pairs being inferred as exposed to the same virus using a DENV antigen test, this dropped to 7 when further analysis took place investigating serotype and genome similarity. Although this method also had a higher failure rate, requiring a full-length genome for both pair members and a serotype results, it shows how phylogenetic information can radically change the conclusions from more simple diagnostic approaches inferring transmission.

Chapter 6 Conclusions and Future Work

In this thesis I have described 3 distinct results chapters with the aim of exploring the immunological risk factors for severe disease and using generated viral genomic information to understand how DENV is able to persist in a population and infer potential transmission routes despite the intermediate insect vector step.

Throughout the genomics work I have tried to explore alternative sequencing approaches to better generate whole viral DENV genomes from clinical samples. I covered the detailed methodology behind each approach followed as well as a preliminary comparison between them in chapter 2, I also explored a novel viral sequencing approach which could be implemented for future work. While the high costs associated with sequencing meant that comparison experiments were impossible, sequences ran on all three systems tried showed that the TE approach more consistently yielded full-length genome coverage with more even coverage.

In chapter 3 I focussed on characterising the immune responses of the study population to DENV and whether any virological or immunological risk factors could be identified which could provide information on the complicated role of previous DENV exposure and the subsequent risk of more severe disease in a later infection. The results from these analyses yielded little significant findings with both primary and post-primary cases being equally as likely to present with disease. Viraemia profiles appeared to be different between the infection numbers with primary infections remaining viraemic for longer however the results were not significant and were most likely an artifact of primary infections presenting earlier, possibly due to their age. Interestingly, although there was a small but significant difference in average viraemia between D1 and D4, this had no implication on disease manifestation, with all 3 commonly found serotypes in this study having similar prevalence's of disease. I identified 26 household contacts who were positive for DENV, and 8 of these were completely asymptomatic with the remainder having sub-clinical symptoms. Analysis of their viraemia showed no significant difference

between these sub-clinical individuals and the reporting population, indicating that they can reach the required viraemia to transmit the disease onto a mosquito vector.

For chapters 4 and 5, NGS was performed on viral RNA isolated from DENV positive individuals who were enrolled in the study with full-length genomic data generated. I applied a variety of phylogenetic and epidemiological techniques to this data to investigate how long various sub-clades of the virus are able to persist in a population and to reconstruct putative transmission pairs. Chapter 4 focussed on applying these phylogenetic techniques to calculate estimated importation dates from the amount of viral diversity found in the identified clades, even though the vast majority of samples used in these metrics were from presenting cases, therefore excluding the large number of sub-symptomatic individuals who would have further contributed to the diversity, I still estimated prevalence times of over 10 years for two of the circulating clades. This indicates both a failure of control measures to eradicate them over the 10-year timespan, and the difficulty of a newly introduced clade to replace those which are already established.

In chapter 5, generated genomic data was used to attempt to infer transmission between enrolled participants despite the ability of the mosquito vector to transmit the virus to untraceable individuals. I developed a workflow which combined viral genomic similarity with biological information from the lifecycle of DENV within humans and mosquito vectors. Using this workflow, I investigated whether individuals were likely being exposed at their reported most frequent daytime location or at their households. Using the shortest possible distance between putative transmission pairs, I identified both inter-household and daytime address transmissions and exposures, as well a combination of the two. I identified a sequential transmission from a sub-clinical individual to a reporting index case supplying evidence that individuals who have none to mild symptoms are involved in the disease transmission in endemic areas.

6.1 Immunological and viral risk factors for DENV disease outcome

Identifying immunological risk factors for severe DENV infection can help prioritise palliative treatment, or targeted treatments should they become available, towards those most at need. Many endemic populations can experience very high case numbers during peak outbreak years which can stretch medical facilities despite the low mortality and hospitalisation rate of the virus.

DENV has the potential to infect a wide age range of individuals and the 4 serotypes mean that people can be infected multiple times, showing different symptoms with each subsequent infection. Post-primary infections have been regarded as having the potential to cause more severe DENV infection[183], however the mechanisms and causes behind this observation are unclear. The phenomenon ADE is implicated with causing increased viral infection of immune cells which would normally clear the virus, there are conflicting studies as to whether this additionally leads to increased viraemia or a change in the infection time course when compared to a primary infection[20,66,236]. The results generated in this study show no significance in previous infection status on disease outcome, with equal proportions of primary and post-primary infections within the reporting population presenting with disease. However, there were very few cases of severe disease in this study, making estimates of this outcome very difficult. As discussed previously this does not take into account whether either primary or post-primary infections contained more asymptomatic individuals who did not report or whether the younger age category of primary infections meant that they were more likely to report despite having milder symptoms.

Different DENV serotypes and within these, sub-genotypes, have been associated with differences in disease outcomes[93,167,216]. Identifying viruses associated with more severe disease could be an opportunity for a screening process to identify those who are most at risk of severe disease. Despite these previous studies suggesting some differences in disease outcome between different viruses, the only significant difference I observed was that people with D1 infections were slightly more likely to

report with symptoms than those with D4 infections, which does support some other studies which have suggested that D4 infections can be milder. Evidence discussed in Chapter 4 implies that the viral population within the study area was stable, with few observed introductions and that some of the strains in circulation had been there for over a decade. The only introduced strain which was identified to have spread significantly in the population was the Cosmopolitan genotype of serotype D2. This strain has not been associated with significantly more disease[237] but has been found to be able to establish itself in countries such as Taiwan and China were the virus has only ever been shown to be seasonally introduced previously[238]. Interestingly, I only detected 2 D3 serotypes, this could be a result of it being less pathogenic so fewer individuals reported with infection, or a lack of the virus being present in the population. Serotyping evidence from the PRNT assay undertaken on a subset of the samples found no individuals with only D3 reactive antibodies and among individuals with difficult to call serotype immunity due high cross-reactivity and repeat exposure, titres to D3 were very low, further suggesting limited exposure in the general population. Indeed, another study in a similar geographical area and time have shown a low prevalence of D3 so it is more likely the latter[239].

Asymptomatic and sub-clinical infections are modelled to be the most common categories of infection[92] but are the least studied owing to the difficulty in capturing asymptomatic but viraemic individuals. In this study I managed to identify 26 sub-clinical household contacts of presenting index cases. Of these, 18 had mild symptoms and 8 were completely asymptomatic. That these sub-clinical individuals were sufficiently viraemic to permit onward transmission of DENV is contraversial[95,162,166,172,174]. However, our data showed similar viremia levels between both groups, and despite the low number of cases in the sub-clinical group, the viral kinetics was very similar.

6.2 Identifying pathogen persistence in populations using phylogenetic techniques

Determining how long a virus is able to remain in a population, as opposed to being regularly reintroduced, is essential for adapting control strategies evaluating their success. In chapter 4 genomic data of DENV samples in an endemic urban population estimated that most of the viral strains had been in circulation for years before the study was started and some may have been in the population for up to 10 years. Using genomic data in this way to estimate persistence time in a population can be applied to other pathogens which can cause seasonal outbreaks of disease and whose continued maintenance throughout the year is unknown. This method is beneficial over others such as genotyping with qPCR or serotyping as it takes into account the multiple lineages that can be present in a population[240].

Although phylogenetic inference of viral diversity is useful for identifying the likely persistence of endemic strains[14,200,241], there are limitations to this inference which are primarily impacted by the sampling of the virus. For example, the mosquito vector is a key source of viral diversity which has not been sampled during this study. Additionally the mosquito provides a mechanism for the virus to persist in a population for longer than would be possible with only humans as hosts owing to the lifespan of the mosquito being longer than the time humans are viraemic (up to 30 days against 10 days[226,242]). This could impact the calculated diversity which has been inferred using only human isolated viruses. Future studies could aim to sample the mosquito virus to determine how much it contributed to the overall diversity.

An additional sampling limitation associated with this method was a result of the majority of DENV cases do not present at a health facility, so the full pathogen population diversity was not captured. However, through sampling sub-clinical contacts of index cases, we attempted to mitigate this, although only 26/511 sequences were from such individuals. As discussed previously, this lack of fully captured diversity will likely only cause an underestimation of the likely time of introduction so for some of the clades identified in the population, their prevalence

could be over 10 years. Future work could attempt to sample more of these asymptomatic individuals to assess the impact of this sub-sampling on the overall diversity.

The phylogenetic techniques used in this study can be readily applied to other pathogens for which there are better control measures, for example malaria and SARS-COVID-19, where active control strategies through vector control, treatment or vaccination can be assessed. Any changes in the sampled diversity and any evidence of new clade importations from generated global phylogenetic trees would be more useful metrics than case numbers alone.

6.3 Using NGS to investigate pathogen dynamics during an outbreak

Determining exposure routes and locations is important for many infectious agents for understanding risk factors for acquiring the disease and how it transmits through a population, which can have implications for control. During this study I have investigated viral transmission using a variety of means including NGS of patient viruses. I have shown that NGS has the potential to allow us to more accurately determine both exposure locations and potentially identify transmission routes of a vector-borne pathogen around a population.

Serological and non-specific testing for DENV were shown to be more unreliable than genomic techniques for inferring transmission. This is likely due to the impact of the vector spreading the virus further than with other infections, such as those transmitted by respiratory route. These traditional tests however have an important role in quickly and easily diagnosing DENV infections which is vital in endemic areas where case numbers can be extremely high.

Mapping the cases over time based upon household locations showed that the pathogen was rapidly disseminated around the study area which would limit the effectiveness of control strategies which are focussed on household locations. I

showed that in order to investigate the dispersal patterns of DENV, daytime locations must be included. Factoring in human movement explains how DENV is able to spread so quickly around the population, it is able to travel far further than would be possible with just the mosquito vector alone. This is likely a result of the day-biting habits of the vector, it would be interesting therefore to compare the dispersal patterns between a disease spread by a night-biting mosquito, where daytime human movement would in theory have reduced impact, and DENV where we see rapid spread around the population during peak season.

NGS enabled me to reconstruct putative transmission pathways of the virus around the population, as part of this I was able to identify likely exposure location. This was only using two given locations of individuals sampled, ideally future studies could use tracking data or more detailed responses to determine whether putative transmission pairs ever came into flight range of a mosquito between their disease periods. This would help to validate the identified pairs and any potential hotspots which could be targeted for vector control. The NGS data was critical in determining true exposures among household members who were positive at the same time and without this data household exposure would be overestimated. This shows how constructing phylogenetic trees using NGS data has real world implications for identifying exposure locations. Indeed from this data I was able to identify daytime locations as being more significantly involved in transmission than was anticipated, other studies have postulated that household locations are likely the most significant exposure route for DENV in an endemic urban environment[162,173,243,244], therefore have suggested vector control measures should be prioritised into heavily residential areas. However, the NGS data from this preliminary study showed that household members who were positive at the same time were more likely to have been exposed to different viruses than from the same source.

Although sub-clinical infections are hypothesised to make up the majority of the infections[3,92], this study highlights the difficulty in capturing those infections. In order to reconstruct sequential transmissions using NGS data, virus was required to be isolated from both the index case and their household contact at the second visit 30

days later. Unfortunately, none of these second visits yielded virus which could be sequenced. Future studies could look at either encouraging household members to visit a health facility if they encountered any symptoms of DENV at all, even very mild, or sample them more frequently within a 30-45 day period. This highlights that at the moment capturing sequential DENV infections, especially those involving sub-clinical individuals, with NGS techniques is challenging. Despite this, I was successful in identifying one putative sequential transmission from a sub-clinical individual (who had no symptoms at all) to an individual who presented at a clinic within the time window for a sequential transmission. Inferring transmission from NGS data in combination with case information remains difficult for many pathogen causing clinical infections[131], however the relatively high mutation rate of DENV and the high time between cases gives time for mutations between cases to develop[205] which is a major limitation with some diseases. This meant that the software that I used to call likely transmission events could reconstruct transmission pairs with biologically plausible timeframes.

Chapter 7 References

1. Guzman, M. G. *et al.* Dengue: a continuing global threat. *Nat. Rev. Microbiol.* **8**, S7–S16 (2010).

2. World Health Organization. Dengue: guidelines for diagnosis, treatment, prevention, and control. *Spec. Program. Res. Train. Trop. Dis.* x, 147 (2009). doi:WHO/HTM/NTD/DEN/2009.1

3. Hasan, S., Jamdar, S. F., Alalowi, M. & Al Ageel Al Beaiji, S. M. Dengue virus: A global human threat: Review of literature. *J. Int. Soc. Prev. Community Dent.* **6**, 1–6 (2016).

4. Ang, L. W. *et al.* A 15-year review of dengue hospitalizations in Singapore: Reducing admissions without adverse consequences, 2003 to 2017. *PLoS Negl. Trop. Dis.* **13**, e0007389 (2019).

5. Messina, J. P. *et al.* The current and future global distribution and population at risk of dengue. *Nat. Microbiol. 2019 49* **4**, 1508–1515 (2019).

6. Solórzano, J. O. *et al.* The Epidemiology of Dengue in the Americas Over the Last Three Decades: A Worrisome Reality. *Am. J. Trop. Med. Hyg.* **82**, 128–135 (2010).

7. Zeng, Z., Zhan, J., Chen, L., Chen, H. & Cheng, S. Global, regional, and national dengue burden from 1990 to 2017: A systematic analysis based on the global burden of disease study 2017. *EClinicalMedicine* **32**, (2021).

8. Dengue and severe dengue. Available at: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. (Accessed: 11th January 2022)

9. Geographical distribution of dengue cases reported worldwide, 2021. Available at: https://www.ecdc.europa.eu/en/publications-data/geographical-distribution-dengue-cases-reported-worldwide-2021. (Accessed: 23rd April 2022)

10. SN, B. *et al.* Epidemic dynamics revealed in dengue evolution. *Mol. Biol. Evol.* **27**, 811–818 (2010).

11. Cummings, D. A. T. *et al.* Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nat. 2004 4276972* **427**, 344–347 (2004).

12. Phanitchat, T. *et al.* Spatial and temporal patterns of dengue incidence in northeastern Thailand 2006-2016. *BMC Infect. Dis.* **19**, 1–12 (2019).

13. Lai, S. *et al.* Seasonal and interannual risks of dengue introduction from South-East Asia into China, 2005-2015. *PLoS Negl. Trop. Dis.* **12**, e0006743 (2018).

14. Kyle, J. L. & Harris, E. Global Spread and Persistence of Dengue. *http://dx.doi.org/10.1146/annurev.micro.62.081307.163005* **62**, 71–92 (2008).

15. Slon Campos, J. L., Mongkolsapaya, J. & Screaton, G. R. The immune response against flaviviruses. *Nature Immunology* **19**, 1189–1198 (2018).

16. Fonseca, V. *et al.* A computational method for the identification of dengue, zika and chikungunya virus species and genotypes. *PLoS Negl. Trop. Dis.* **13**, (2019).

17. Vilsker, M. *et al.* Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **35**, 871–873 (2019).

18. Mendoza, M. Y., Salas-Benito, J. S., Lanz-Mendoza, H., Hernández-Martinez, S.

& Del Angel, R. M. A putative receptor for dengue virus in mosquito tissues: Localization of a 45-KDA glycoprotein. *Am. J. Trop. Med. Hyg.* **67**, (2002).

19. Rodenhuis-Zybert, I. A. *et al.* Immature dengue virus: A veiled pathogen? *PLoS Pathog.* **6**, (2010).

20. Flipse, J., Wilschut, J. & Smit, J. M. Molecular Mechanisms Involved in Antibody-Dependent Enhancement of Dengue Virus Infection in Humans. *Traffic* **14**, 25–35 (2013).

21. Clyde, K., Kyle, J. L. & Harris, E. Recent Advances in Deciphering Viral and Host Determinants of Dengue Virus Replication and Pathogenesis. *J. Virol.* **80**, (2006).

22. Friebe, P., Peña, J., Pohl, M. O. F. & Harris, E. Composition of the sequence downstream of the dengue virus 5' cyclization sequence (dCS) affects viral RNA replication. *Virology* **422**, 346–356 (2012).

23. Nasar, S., Rashid, N. & Iftikhar, S. Dengue proteins with their role in pathogenesis, and strategies for developing an effective anti-dengue treatment: A review. *J. Med. Virol.* **92**, 941–955 (2020).

24. Song, H., Qi, J., Haywood, J., Shi, Y. & Gao, G. F. Zika virus NS1 structure reveals diversity of electrostatic surfaces among flaviviruses. *Nat. Struct. Mol. Biol.* **23**, 456–458 (2016).

25. Flamand, M. *et al.* Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion. *J. Virol.* **73**, 6104–6110 (1999).

26. Duyen, H. T. L. *et al.* Kinetics of plasma viremia and soluble nonstructural protein 1 concentrations in dengue: differential effects according to serotype and immune status. *J. Infect. Dis.* **203**, 1292–1300 (2011).

27. Colón-González, F. J. *et al.* Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Med.* **18**, e1003542 (2021).

28. Lee, H. S. *et al.* Seasonal patterns of dengue fever and associated climate factors in 4 provinces in Vietnam from 1994 to 2013. *BMC Infect. Dis.* **17**, 218 (2017).

29. Rabaa, M. A. *et al.* Dengue Virus in Sub-tropical Northern and Central Viet Nam: Population Immunity and Climate Shape Patterns of Viral Invasion and Maintenance. *PLoS Negl. Trop. Dis.* **7**, e2581 (2013).

30. Nha Trang Population 2022 (Demographics, Maps, Graphs). Available at: https://worldpopulationreview.com/world-cities/nha-trang-population. (Accessed: 24th March 2022)

31. Quyen, D. Le *et al.* Epidemiological, Serological, and Virological Features of Dengue in Nha Trang City, Vietnam. *Am. J. Trop. Med. Hyg.* **98**, 402 (2018).

32. Culshaw, A., Mongkolsapaya, J. & Screaton, G. R. The immunopathology of dengue and Zika virus infections. *Curr. Opin. Immunol.* **48**, 1–6 (2017).

33. Bhatt, P., Sabeena, S. P., Varma, M. & Arunkumar, G. Current Understanding of the Pathogenesis of Dengue Virus Infection. *Curr. Microbiol.* **78**, 17 (2021).

34. Simmons, C. P., Farrar, J. J., van Vinh Chau, N. & Wills, B. Dengue. *N. Engl. J. Med.* **366**, 1423–1432 (2012).

35. Fernando, S. *et al.* Patterns and causes of liver involvement in acute dengue infection. *BMC Infect. Dis.* **16**, 1–9 (2016).

36. Deen, J. & von Seidlein, L. Paracetamol for dengue fever: no benefit and potential harm? *Lancet Glob. Heal.* **7**, e552–e553 (2019).

37. Gubler, D. J., Suharyono, W., Tan, R., Abidin, M. & Sie, A. Viraemia in patients with naturally acquired dengue infection. *Bull. World Health Organ.* **59**, 623–630 (1981).

38. Lehnert, R., Pletz, M., Reuss, A. & Schaberg, T. Antiviral Medications in Seasonal and Pandemic Influenza: A Systematic Review. *Dtsch. Arztebl. Int.* **113**, 799 (2016).

39. Lee, V. J. *et al.* Oseltamivir ring prophylaxis for containment of 2009 H1N1 influenza outbreaks. *N. Engl. J. Med.* **362**, 2166–2174 (2010).

40. Pang, J. *et al.* Discovery and Validation of Prognostic Biomarker Models to Guide Triage among Adult Dengue Patients at Early Infection. *PLoS One* **11**, (2016).

41. Lim, S. P. Dengue drug discovery: Progress, challenges and outlook. *Antiviral Res.* **163**, 156–178 (2019).

42. Kaptein, S. J. F. *et al.* A pan-serotype dengue virus inhibitor targeting the NS3–NS4B interaction. *Nat. 2021 5987881* **598**, 504–509 (2021).

43. Agudelo Higuita, N. I., White, B. P., Franco-Paredes, C. & McGhee, M. A. An update on prevention of malaria in travelers. *Ther. Adv. Infect. Dis.* **8**, (2021).

44. Prestwood, T. R., Prigozhin, D. M., Sharar, K. L., Zellweger, R. M. & Shresta, S. A Mouse-Passaged Dengue Virus Strain with Reduced Affinity for Heparan Sulfate Causes Severe Disease in Mice by Establishing Increased Systemic Viral Loads. *J. Virol.* **82**, 8411–8421 (2008).

45. Zompi, S. & Harris, E. Animal models of dengue virus infection. *Viruses* **4**, 62–82 (2012).

46. Yauch, L. E. & Shresta, S. Dengue virus vaccine development. *Adv. Virus Res.*

88, 315–372 (2014).

47. Guirakhoo, F. *et al.* Construction, safety, and immunogenicity in nonhuman primates of a chimeric yellow fever-dengue virus tetravalent vaccine. *J. Virol.* **75**, 7290–7304 (2001).

48. Sridhar, S. *et al.* Effect of Dengue Serostatus on Dengue Vaccine Safety and Efficacy. *N. Engl. J. Med.* **379**, (2018).

49. Dengue vaccine: WHO position paper, September 2018 - Recommendations. *Vaccine* **37**, 4848–4849 (2019).

50. Chau, T. N. B. *et al.* Dengue in Vietnamese Infants—Results of Infection-Enhancement Assays Correlate with Age-Related Disease Epidemiology, and Cellular Immune Responses Correlate with Disease Severity. *J. Infect. Dis.* **198**, 516–524 (2008).

51. Halstead, S. B. Pathogenesis of Dengue: Dawn of a New Era. *F1000Research* **4**, (2015).

52. Montoya, M. *et al.* Symptomatic Versus Inapparent Outcome in Repeat Dengue Virus Infections Is Influenced by the Time Interval between Infections and Study Year. *PLoS Negl. Trop. Dis.* **7**, (2013).

53. Anderson, K. B. *et al.* A shorter time interval between first and second dengue infections is associated with protection from clinical Illness in a school-based cohort in Thailand. *J. Infect. Dis.* **209**, 360–368 (2014).

54. Halstead, S. B., Mahalingam, S., Marovich, M. A., Ubol, S. & Mosser, D. M. Intrinsic antibody-dependent enhancement of microbial infection in macrophages: disease regulation by immune complexes. *Lancet. Infect. Dis.* **10**, 712–22 (2010).

55. Lin, C.-F. *et al.* Endothelial Cell Apoptosis Induced by Antibodies Against Dengue Virus Nonstructural Protein 1 Via Production of Nitric Oxide. *J. Immunol.* **169**, 657–664 (2002).

56. Dejnirattisai, W. *et al.* Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. *Nat. Publ. Gr.* **17**, 1102–1108 (2016).

57. Moi, M. L., Lim, C.-K., Takasaki, T. & Kurane, I. Involvement of the Fc receptor IIA cytoplasmic domain in antibody-dependent enhancement of dengue virus infection. *J. Gen. Virol.* **91**, 103–111 (2010).

58. Chan, K. R. *et al.* Leukocyte immunoglobulin-like receptor B1 is critical for antibody-dependent dengue. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2722–2727 (2014).

59. da Silva Voorham, J. M. *et al.* Antibodies against the envelope glycoprotein promote infectivity of immature dengue virus serotype 2. *PLoS One* **7**, e29957 (2012).

60. Tian, Y., Grifoni, A., Sette, A. & Weiskopf, D. Human T Cell Response to Dengue Virus Infection. *Frontiers in Immunology* **10**, (2019).

61. Mathew, A. *et al.* Dominant recognition by human CD8+ cytotoxic T lymphocytes of dengue virus nonstructural proteins NS3 and NS1.2a. *J. Clin. Invest.* **98**, 1684–1691 (1996).

62. Gagnon, S. J., Ennis, F. A. & Rothman, A. L. Bystander Target Cell Lysis and Cytokine Production by Dengue Virus-Specific Human CD4 + Cytotoxic T-Lymphocyte Clones . *J. Virol.* **73**, 3623–3629 (1999).

63. Kurane, I. *et al.* Human T cell responses to dengue virus antigens. Proliferative responses and interferon gamma production. *J. Clin. Invest.* **83**, 506–513

(1989).

64. Chaturvedi, U. C. *et al.* Sequential Production of Cytokines by Dengue Virus-Infected Human Peripheral Blood Leukocyte Cultures. *J. Med. Virol* **59**, 335–340 (1999).

65. Mustafa, A. S., Elbishbishi, E. A., Agarwal, R. & Chaturvedi, U. C. Elevated levels of interleukin-13 and IL-18 in patients with dengue hemorrhagic fever. *FEMS Immunol. Med. Microbiol.* **30**, 229–233 (2001).

66. Guzman, M. G., Alvarez, M. & Halstead, S. B. Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: An historical perspective and role of antibody-dependent enhancement of infection. *Arch. Virol.* **158**, 1445–1459 (2013).

67. Wang, T. T. *et al.* IgG antibodies to dengue enhanced for FcγRIIIA binding determine disease severity. *Science* **355**, 395 (2017).

68. Azami, N. A. M., Takasaki, T., Kurane, I. & Moi, M. L. Non-Human Primate Models of Dengue Virus Infection: A Comparison of Viremia Levels and Antibody Responses during Primary and Secondary Infection among Old World and New World Monkeys. *Pathogens* **9**, 1–16 (2020).

69. Durbin, A. P. & Whitehead, S. S. The dengue human challenge model: has the time come to accept this challenge? *J. Infect. Dis.* **207**, 697–699 (2013).

70. Flasche, S. *et al.* The Long-Term Safety, Public Health Impact, and Cost-Effectiveness of Routine Vaccination with a Recombinant, Live-Attenuated Dengue Vaccine (Dengvaxia): A Model Comparison Study. *PLoS Med.* **13**, (2016).

71. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).

72. King, C. C. *et al.* Comparative analysis of full genomic sequences among different genotypes of dengue virus type 3. *Virol. J.* **5**, 63 (2008).

73. Nguyen, T. H. T. *et al.* Methods to discriminate primary from secondary dengue during acute symptomatic infection. *BMC Infect. Dis.* **18**, (2018).

74. Cucunawangsih, Lugito, N. P. H. & Kurniawan, A. Immunoglobulin G (IgG) to IgM ratio in secondary adult dengue infection using samples from early days of symptoms onset. *BMC Infect. Dis.* **15**, (2015).

75. Cassedy, A., Parle-McDermott, A. & O'Kennedy, R. Virus Detection: A Review of the Current and Emerging Molecular and Immunological Methods. *Front. Mol. Biosci.* **8**, 76 (2021).

76. Raafat, N., Blacksell, S. D. & Maude, R. J. A review of dengue diagnostics and implications for surveillance and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **113**, (2019).

77. Fritzell, C. *et al.* Current challenges and implications for dengue, chikungunya and Zika seroprevalence studies worldwide: A scoping review. *PLoS Negl. Trop. Dis.* **12**, e0006533 (2018).

78. L'Huillier, A. G. *et al.* Evaluation of euroimmun anti-zika virus igm and igg enzyme-linked immunosorbent assays for zika virus serologic testing. *J. Clin. Microbiol.* **55**, 2462–2471 (2017).

79. Hunsperger, E. A. *et al.* Performance of Dengue Diagnostic Tests in a Single-Specimen Diagnostic Algorithm. *J. Infect. Dis.* **214**, 836–844 (2016).

80. Chanama, S. *et al.* Analysis of specific IgM responses in secondary dengue virus infections: levels and positive rates in comparison with primary infections. *J. Clin. Virol.* **31**, 185–189 (2004).

81. Sirivichayakul, C., Sabchareon, A., Limkittikul, K. & Yoksan, S. Plaque reduction neutralization antibody test does not accurately predict protection against dengue infection in Ratchaburi cohort, Thailand. *Virol. J.* **11**, 1–5 (2014).

82. Santiago, G. A. *et al.* Performance of the Trioplex real-time RT-PCR assay for detection of Zika, dengue, and chikungunya viruses. *Nat. Commun. 2018 91* **9**, 1–10 (2018).

83. Santiago, G. A. *et al.* Analytical and clinical performance of the CDC real time RT-PCR assay for detection and typing of dengue virus. *PLoS Negl. Trop. Dis.* **7**, e2311 (2013).

84. Domingo, C. *et al.* 2nd International external quality control assessment for the molecular diagnosis of dengue infections. *PLoS Negl. Trop. Dis.* **4**, (2010).

85. Tsuzuki, A. *et al.* Can Daytime Use of Bed Nets Not Treated with Insecticide Reduce the Risk of Dengue Hemorrhagic Fever Among Children in Vietnam? *Am. J. Trop. Med. Hyg.* **82**, 1157 (2010).

86. Ferguson, N. M. Challenges and opportunities in controlling mosquito-borne infections. *Nat. 2018 5597715* **559**, 490–497 (2018).

87. Bowman, L. R., Donegan, S. & McCall, P. J. Is Dengue Vector Control Deficient in Effectiveness or Evidence?: Systematic Review and Meta-analysis. *PLoS Negl. Trop. Dis.* **10**, (2016).

88. Ross, P., Callahan, A., Yang, Q., … M. J.-E. and & 2020, undefined. An elusive endosymbiont: Does Wolbachia occur naturally in Aedes aegypti? *Wiley Online Libr.* **10**, 1581–1591 (2020).

89. Walker, T., Johnson, P., Moreira, L., Nature, I. I.-O.- & 2011, undefined. The w Mel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. *nature.com*

90. Moreira, L., Iturbe-Ormaetxe, I., Jeffery, J., Cell, G. L.- & 2009, undefined. A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium. *Elsevier*

91. Utarini, A. *et al.* Efficacy of Wolbachia-Infected Mosquito Deployments for the Control of Dengue. **384**, 2177–2186 (2021).

92. ten Bosch, Q. A. *et al.* Contributions from the silent majority dominate dengue virus transmission. *PLoS Pathog.* **14**, e1006965 (2018).

93. Zou, C. *et al.* Virulence difference of five type I dengue viruses and the intrinsic molecular mechanism. *PLoS Negl. Trop. Dis.* **13**, e0007202 (2019).

94. Rico-Hesse, R. Dengue Virus Virulence and Transmission Determinants. in *Dengue Virus* **338**, 45–55 (Nature Publishing Group, 2010).

95. Tsai, J. J., Lin, P. C., Tsai, C. Y., Wang, Y. H. & Liu, L. T. Low frequency of asymptomatic dengue virus-infected donors in blood donor centers during the largest dengue outbreak in Taiwan. *PLoS One* **13**, (2018).

96. Kucharski, A. J. *et al.* Using paired serology and surveillance data to quantify dengue transmission and control during a large outbreak in Fiji. *Elife* **7**, (2018).

97. Baillie, G. J. *et al.* Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *J. Virol.* **86**, 11–18 (2011).

98. Christenbury, J. G. *et al.* A method for full genome sequencing of all four serotypes of the dengue virus. *J. Virol. Methods* **169**, 202–206 (2010).

99. Aw, P. P. K. *et al.* Next-Generation Whole Genome Sequencing of Dengue Virus. *Methods Mol. Biol.* **1138**, 175–195 (2014).

100. Watson, S. J. *et al.* Viral population analysis and minority-variant detection

using short read next-generation sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, (2013).

101. Wilm, A. *et al.* LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).

102. Houldcroft, C. J., Beale, M. A. & Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **15**, 183 (2017).

103. Thomson, E. *et al.* Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. *J. Clin. Microbiol.* **54**, 2470–2484 (2016).

104. Wohl, S., Schaffner, S. F. & Sabeti, P. C. Genomic Analysis of Viral Outbreaks. *Annu. Rev. Virol.* **3**, 173–195 (2016).

105. Sim, S. & Hibberd, M. L. Genomic approaches for understanding dengue: Insights from the virus, vector, and host. *Genome Biology* **17**, (2016).

106. Lipsitch, M. *et al.* Transmission dynamics and control of severe acute respiratory syndrome. *Science (80-.).* **300**, 1966–1970 (2003).

107. Kiss, I. Z., Green, D. M. & Kao, R. R. Disease contact tracing in random and clustered networks. *Proc. R. Soc. B Biol. Sci.* **272**, 1407–1414 (2005).

108. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet. 2012 135* **13**, 303–314 (2012).

109. Carroll, M. W. *et al.* Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* **524**, 97 (2015).

110. Ho, Z. J. M. *et al.* Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. *Lancet Infect.*

Dis. **17**, 813–821 (2017).

111. Lemey, P. *et al.* Molecular Footprint of Drug-Selective Pressure in a Human Immunodeficiency Virus Transmission Chain. *J. Virol.* **79**, 11981–11989 (2005).

112. Guan, Y. *et al.* Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8950–8955 (2002).

113. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

114. Guindon, S. & Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **52**, 696–704 (2003).

115. Kumar, S. Molecular clocks: Four decades of evolution. *Nature Reviews Genetics* **6**, 654–662 (2005).

116. Vazquez-Prokopec, G. M., Montgomery, B. L., Horne, P., Clennon, J. A. & Ritchie, S. A. Combining contact tracing with targeted indoor residual spraying significantly reduces dengue transmission. *Sci. Adv.* **3**, (2017).

117. Dennis, A. M. *et al.* Integration of Contact Tracing and Phylogenetics in an Investigation of Acute HIV Infection. *Sex. Transm. Dis.* **45**, 222–228 (2018).

118. GISAID - Initiative. Available at: https://www.gisaid.org/. (Accessed: 27th April 2022)

119. Aggarwal, D. *et al.* The role of viral genomics in understanding COVID-19 outbreaks in long-term care facilities. *The Lancet Microbe* **3**, e151–e158 (2022).

120. Geoghegan, J. L. *et al.* Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. *Nat. Commun. 2020 111* **11**, 1–7 (2020).

121. Putri, N. D. *et al.* Whole-Genome Sequencing of SARS-CoV-2 Infection in a Cluster of Immunocompromised Children in Indonesia. *Front. Med.* **9**, 320 (2022).

122. Kousathanas, A. *et al.* Whole genome sequencing reveals host factors underlying critical Covid-19. *Nat. 2022* 1–10 (2022). doi:10.1038/s41586-022-04576-6

123. Banerjee, A. *et al.* RNA-Seq analysis of peripheral blood mononuclear cells reveals unique transcriptional signatures associated with disease progression in dengue patients. *Transl. Res.* **186**, 62-78.e9 (2017).

124. Zanini, F. *et al.* Virus-inclusive single-cell RNA sequencing reveals the molecular signature of progression to severe dengue. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E12363–E12369 (2018).

125. Peck, K. M. & Lauring, A. S. Complexities of Viral Mutation Rates. *J. Virol.* **92**, (2018).

126. Hossain, A. D. *et al.* Effectiveness of contact tracing in the control of infectious diseases: a systematic review. *Lancet Public Heal.* **7**, e259–e273 (2022).

127. Grubaugh, N. D. *et al.* Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol. 2018 41* **4**, 10–19 (2018).

128. Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics* **10**, 540–550 (2009).

129. Eshleman, S. H. *et al.* Analysis of genetic linkage of HIV from couples enrolled in the HIV prevention trials network 052 trial. *J. Infect. Dis.* **204**, 1918–1926 (2011).

130. Cottam, E. M. *et al.* Integrating genetic and epidemiological data to determine

transmission pathways of foot-and-mouth disease virus. *Proc. R. Soc. B Biol. Sci.* **275**, 887–895 (2008).

131. Campbell, F., Strang, C., Ferguson, N., Cori, A. & Jombart, T. When are pathogen genome sequences informative of transmission events? *PLoS Pathog.* **14**, (2018).

132. Wang, W. H. *et al.* Dengue hemorrhagic fever – A systemic literature review of current perspectives on pathogenesis, prevention and control. *J. Microbiol. Immunol. Infect.* **53**, 963–978 (2020).

133. Agoti, C. N. *et al.* Local Evolutionary Patterns of Human Respiratory Syncytial Virus Derived from Whole-Genome Sequencing. *J. Virol.* **89**, 3444 (2015).

134. Malik, J. A. *et al.* The SARS-CoV-2 mutations versus vaccine effectiveness: New opportunities to new challenges. *J. Infect. Public Health* **15**, 228–240 (2022).

135. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352**, 345 (2016).

136. Sebastián-Martín, A., Barrioluengo, V. & Menéndez-Arias, L. Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases. *Sci. Reports 2018 81* **8**, 1–13 (2018).

137. Chiara, M. *et al.* Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief. Bioinform.* **22**, 616–630 (2021).

138. Fitzpatrick, A. H. *et al.* High Throughput Sequencing for the Detection and Characterization of RNA Viruses. *Front. Microbiol.* **12**, 190 (2021).

139. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of

Zika and other virus genomes directly from clinical samples. *bioRxiv* 098913 (2017). doi:10.1101/098913

140. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189 (2009).

141. Metsky, H. C. *et al.* Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat. Biotechnol.* **37**, 160–168 (2019).

142. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).

143. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

144. Williams, R.J., Tutill, H., Roy, S., Romero, E.Y., Williams, C.A., & Breuer, J. *Utilization of Agilent SureSelect Target Enrichment for Whole Genome Sequencing of Viruses and Bacteria.* (2019).

145. Langevin, S. A. *et al.* Peregrine: A rapid and unbiased method to produce strand-specific RNA-Seq libraries from small quantities of starting material. *RNA Biol.* **10**, 502–515 (2013).

146. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

147. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).

148. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*

25, 2078–2079 (2009).

149. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *bioRxiv* 383513 (2018). doi:10.1101/383513

150. Roy, S. K. & Bhattacharjee, S. Dengue virus: Epidemiology, biology, and disease aetiology. *Can. J. Microbiol.* **67**, 687–702 (2021).

151. Davis, J. J. *et al.* Analysis of the ARTIC Version 3 and Version 4 SARS-CoV-2 Primers and Their Impact on the Detection of the G142D Amino Acid Substitution in the Spike Protein. *Microbiol. Spectr.* **9**, (2021).

152. Vogels, C. B. F. *et al.* Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLOS Biol.* **19**, e3001236 (2021).

153. Screaton, G., Mongkolsapaya, J., Yacoub, S. & Roberts, C. New insights into the immunopathology and control of dengue virus infection. *Nat. Rev. Immunol.* **15**, 745–759 (2015).

154. Yung, C. F. *et al.* Dengue Serotype-Specific Differences in Clinical Manifestation, Laboratory Parameters and Risk of Severe Disease in Adults, Singapore. *Am. J. Trop. Med. Hyg.* **92**, 999 (2015).

155. Vicente, C. R. *et al.* Serotype influences on dengue severity: a cross-sectional study on 485 confirmed dengue cases in Vitória, Brazil. *BMC Infect. Dis.* **16**, (2016).

156. Lan, N. T. P. & Hirayama, K. Host genetic susceptibility to severe dengue infection. *Trop. Med. Health* **39**, 73–81 (2011).

157. LaFleur, C. *et al.* HLA-DR antigen frequencies in Mexican patients with dengue virus infection: HLA-DR4 as a possible genetic resistance factor for dengue

hemorrhagic fever. *Hum. Immunol.* **63**, 1039–1044 (2002).

158. Rigau-Pérez, J. G. *et al.* Dengue and dengue haemorrhagic fever. *Lancet* **352**, 971–977 (1998).

159. Tully, D. & Griffiths, C. L. Dengvaxia: the world's first vaccine for prevention of secondary dengue. *Ther. Adv. Vaccines Immunother.* **9**, (2021).

160. De La Cruz-Hernández, S. I. *et al.* Determination of viremia and concentration of circulating nonstructural protein 1 in patients infected with dengue virus in Mexico. *Am. J. Trop. Med. Hyg.* **88**, 446–454 (2013).

161. Vaughn, D. W. *et al.* Dengue Viremia Titer, Antibody Response Pattern, and Virus Serotype Correlate with Disease Severity. *J. Infect. Dis.* **181**, 2–9 (2000).

162. Matangkasombut, P. *et al.* Dengue viremia kinetics in asymptomatic and symptomatic infection. *Int. J. Infect. Dis.* **101**, 90–97 (2020).

163. Recker, M. *et al.* Immunological serotype interactions and their effect on the epidemiological pattern of dengue. *Proc. R. Soc. B Biol. Sci.* **276**, 2541–2548 (2009).

164. Aguiar, M., Kooi, B. & Stollenwerk, N. Epidemiology of Dengue Fever: A Model with Temporary Cross-Immunity and Possible Secondary Infection Shows Bifurcations and Chaotic Behaviour in Wide Parameter Regions. *Math. Model. Nat. Phenom.* **3**, 48–70 (2008).

165. Forshey, B. M., Stoddard, S. T. & Morrison, A. C. Dengue Viruses and Lifelong Immunity: Reevaluating the Conventional Wisdom. *J. Infect. Dis.* **214**, 979–981 (2016).

166. Duong, V. *et al.* Asymptomatic humans transmit dengue virus to mosquitoes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14688–14693 (2015).

167. Cologna, R., Armstrong, P. M. & Rico-Hesse, R. Selection for Virulent Dengue Viruses Occurs in Humans and Mosquitoes. *J. Virol.* **79**, 853 (2005).

168. Aguas, R., Dorigatti, I., Coudeville, L., Luxemburger, C. & Ferguson, N. M. Cross-serotype interactions and disease outcome prediction of dengue infections in Vietnam. *Sci. Reports 2019 91* **9**, 1–12 (2019).

169. Nimmanitya, S., Kliks, S. C., Burke, D. S. & Nisalak, A. Evidence That Maternal Dengue Antibodies Are Important in the Development of Dengue Hemorrhagic Fever in Infants. *Am. J. Trop. Med. Hyg.* **38**, 411–419 (1988).

170. Guzmán, M. G. *et al.* Enhanced severity of secondary dengue-2 infections: death rates in 1981 and 1997 Cuban outbreaks. *Rev. Panam. Salud Pública* **11**, 223–227 (2002).

171. Harrison, S. C. Immunogenic cross-talk between dengue and Zika viruses. *Nat. Immunol.* **17**, 1010–1012 (2016).

172. Carrington, L. B. & Simmons, C. P. Human to mosquito transmission of dengue viruses. *Frontiers in Immunology* **5**, (2014).

173. Anders, K. L. *et al.* Households as Foci for Dengue Transmission in Highly Urban Vietnam. *PLoS Negl. Trop. Dis.* **9**, (2015).

174. Reyes, M. *et al.* Index cluster study of dengue virus infection in Nicaragua. *Am. J. Trop. Med. Hyg.* **83**, 683–689 (2010).

175. Biggs, J. R. *et al.* Estimating the annual dengue force of infection from the age of reporting primary infections across urban centres in endemic countries. *BMC Med.* **19**, 1–14 (2021).

176. Guzman, M. G., Gubler, D. J., Izquierdo, A., Martinez, E. & Halstead, S. B. Dengue infection. *Nat. Rev. Dis. Prim. 2016 21* **2**, 1–25 (2016).

177. Nguyen, N. M. *et al.* Host and viral features of human dengue cases shape the population of infected and infectious Aedes aegypti mosquitoes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9072–9077 (2013).

178. Morsy, S. *et al.* The association between dengue viremia kinetics and dengue severity: A systemic review and meta-analysis. *Rev. Med. Virol.* **30**, 1–10 (2020).

179. Sangkaew, S. *et al.* Risk predictors of progression to severe disease during the febrile phase of dengue: a systematic review and meta-analysis. *Lancet Infect. Dis.* **21**, 1014–1026 (2021).

180. Hang, V. T. *et al.* Diagnostic Accuracy of NS1 ELISA and Lateral Flow Rapid Tests for Dengue Sensitivity, Specificity and Relationship to Viraemia and Antibody Responses. *PLoS Negl. Trop. Dis.* **3**, (2009).

181. Ahmad, M. H. *et al.* The Sensitivity, Specificity and Accuracy of Warning Signs in Predicting Severe Dengue, the Severe Dengue Prevalence and Its Associated Factors. *Int. J. Environ. Res. Public Health* **15**, (2018).

182. Sharp, T. M. *et al.* Knowledge gaps in the epidemiology of severe dengue impede vaccine evaluation. *Lancet Infect. Dis.* **22**, e42–e51 (2022).

183. Tsheten, T. *et al.* Clinical predictors of severe dengue: a systematic review and meta-analysis. *Infect. Dis. Poverty* **10**, 1–10 (2021).

184. Leung, N. H. L. Transmissibility and transmission of respiratory viruses. *Nature Reviews Microbiology* 1–18 (2021). doi:10.1038/s41579-021-00535-6

185. Bannister-Tyrrell, M. *et al.* Households or Hotspots? Defining Intervention Targets for Malaria Elimination in Ratanakiri Province, Eastern Cambodia. *J. Infect. Dis.* **220**, (2019).

186. Stoddard, S. T. *et al.* House-to-house human movement drives dengue virus transmission. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 994–999 (2013).

187. Chuang, Y. C., Wang, S. Y., Lin, Y. S., Chen, H. R. & Yeh, T. M. Re-evaluation of the pathogenic roles of nonstructural protein 1 and its antibodies during dengue virus infection. *J. Biomed. Sci.* **20**, (2013).

188. Srikiatkhachorn, A. & Kelley, J. F. Endothelial Cells in Dengue Hemorrhagic Fever. *Antiviral Res.* **0**, 160 (2014).

189. Cardoso, C. W. *et al.* Outbreak of Exanthematous Illness Associated with Zika, Chikungunya, and Dengue Viruses, Salvador, Brazil. *Emerg. Infect. Dis.* **21**, 2274–2276 (2015).

190. Schmidt, W.-P. *et al.* Population Density, Water Supply, and the Risk of Dengue Fever in Vietnam: Cohort Study and Spatial Analysis. *Sciences (New. York).* doi:10.1371/journal.pmed.1001082

191. Morrison, A. C. *et al.* Aedes aegypti (Diptera: Culicidae) production from nonresidential sites in the Amazonian city of Iquitos, Peru. *Ann. Trop. Med. Parasitol.* **100**, (2006).

192. Buhler, C., Winkler, V., Runge-Ranzinger, S., Boyce, R. & Horstick, O. Environmental methods for dengue vector control – A systematic review and meta-analysis. *PLoS Negl. Trop. Dis.* **13**, e0007420 (2019).

193. Alphey, L. *et al.* Sterile-Insect Methods for Control of Mosquito-Borne Diseases: An Analysis. *https://home.liebertpub.com/vbz* **10**, 295–311 (2010).

194. Caragata, E. P., Dutra, H. L. C., Sucupira, P. H. F., Ferreira, A. G. A. & Moreira, L. A. Wolbachia as translational science: controlling mosquito-borne pathogens. *Trends in Parasitology* **37**, (2021).

195. Fongwen, N. *et al.* Implementation strategies for the first licensed dengue vaccine: A meeting report. *Vaccine* **39**, 4759–4765 (2021).

196. Brady, O. J. *et al.* Case-area targeted interventions (CATI) for reactive dengue control: Modelling effectiveness of vector control and prophylactic drugs in Singapore. **15**, e0009562 (2021).

197. Nikin-Beers, R. & Ciupe, S. M. The role of antibody in enhancing dengue virus infection. *Math. Biosci.* **263**, 83–92 (2015).

198. Yuan, H. Y. *et al.* The effects of seasonal climate variability on dengue annual incidence in Hong Kong: A modelling study. *Sci. Reports 2020 101* **10**, 1–10 (2020).

199. Takemura, T. *et al.* The 2017 Dengue virus 1 outbreak in northern Vietnam was caused by a locally circulating virus group. *Trop. Med. Health* **50**, 3 (2022).

200. Galarion, M. J. *et al.* Genotypic persistence of dengue virus in the Philippines. *Infect. Genet. Evol.* **69**, 134–141 (2019).

201. Schreiber, M. J. *et al.* Genomic Epidemiology of a Dengue Virus Epidemic in Urban Singapore. *J. Virol.* **83**, 4163–4173 (2009).

202. Holmes, E. C. The phylogeography of human viruses. *Mol. Ecol.* **13**, 745–756 (2004).

203. EC, H. RNA virus genomics: a world of possibilities. *J. Clin. Invest.* **119**, 2488–2495 (2009).

204. Pybus, O. G., Fraser, C. & Rambaut, A. Evolutionary epidemiology: preparing for an age of genomic plenty. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, (2013).

205. Pollett, S. *et al.* Understanding dengue virus evolution to support epidemic surveillance and counter-measure development. *Infection, Genetics and*

Evolution **62**, 279–295 (2018).

206. Williams, R. J. *et al. Utilization of Agilent SureSelect Target Enrichment for Whole Genome Sequencing of Viruses and Bacteria*.

207. Virus Variation - NCBI. Available at: https://www.ncbi.nlm.nih.gov/genomes/VirusVariation/Database/nph-select.cgi. (Accessed: 3rd October 2021)

208. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).

209. Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268 (2015).

210. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, (2018).

211. Murrell, B. *et al.* Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLOS Genet.* **8**, e1002764 (2012).

212. Weaver, S. *et al.* Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **35**, 773–777 (2018).

213. World Health Organization. *Western Pacific Regional Action Plan for Dengue Prevention and Control (2016)*. (2017).

214. Cuong, H. Q. *et al.* Spatiotemporal Dynamics of Dengue Epidemics, Southern Vietnam. *Emerg. Infect. Dis.* **19**, 945 (2013).

215. Dengue Situation Update Number 584. *World Health Organisation - Western Pacific Region* (2019). Available at:

https://iris.wpro.who.int/bitstream/handle/10665.1/14329/Dengue-20191219.pdf. (Accessed: 8th December 2021)

216. Rico-Hesse, R. Dengue virus virulence and transmission determinants. *Current Topics in Microbiology and Immunology* **338**, (2009).

217. Nguyen-Tien, T., Probandari, A. & Ahmad, R. A. Barriers to Engaging Communities in a Dengue Vector Control Program: An Implementation Research in an Urban Area in Hanoi City, Vietnam. *Am. J. Trop. Med. Hyg.* **100**, 964 (2019).

218. CHADEE, D. D. Landing periodicity of the mosquito Aedes aegypti in Trinidad in relation to the timing of insecticidal space-spraying. *Med. Vet. Entomol.* **2**, 189–192 (1988).

219. McDonald, P. T. Population characteristics of domestic Aedes aegypti (diptera: culicidae) in villages on the Kenya coast. II. Dispersal within and between villages. *J. Med. Entomol.* **14**, 49–53 (1977).

220. Reiter, P., Amador, M. A., Anderson, R. A. & Clark, G. G. Short report: Dispersal of Aedes aegypti in an urban area after blood feeding as demonstrated by rubidium-marked eggs. *Am. J. Trop. Med. Hyg.* **52**, 177–179 (1995).

221. Trpis, M. & Hausermann, W. Dispersal and other population parameters of Aedes aegypti in an African village and their possible significance in epidemiology of vector-borne diseases. *Am. J. Trop. Med. Hyg.* **35**, 1263–1279 (1986).

222. Halstead, S. Recent advances in understanding dengue [version 1; peer review: 2 approved]. *F1000Research* **8**, (2019).

223. Halstead, S. B. Pathogenesis of dengue: challenges to molecular biology. *Science* **239**, 476–81 (1988).

224. Rudolph, K. E., Lessler, J., Moloney, R. M., Kmush, B. & Cummings, D. A. T. Review article: Incubation periods of mosquito-borne viral infections: a systematic review. *American Journal of Tropical Medicine and Hygiene* **90**, 882–891 (2014).

225. Clements, A. N. & Paterson, G. D. The Analysis of Mortality and Survival Rates in Wild Populations of Mosquitoes. *J. Appl. Ecol.* **18**, 373 (1981).

226. Chan, M. & Johansson, M. A. The Incubation Periods of Dengue Viruses. *PLoS One* **7**, 50972 (2012).

227. Salje, H. *et al.* Reconstructing unseen transmission events to infer dengue dynamics from viral sequences. *Nat. Commun.* **12**, 1–10 (2021).

228. Biggs, J. R. *et al.* A serological framework to investigate acute primary and post-primary dengue cases reporting across the Philippines. *BMC Med.* **18**, (2020).

229. Balaban, M., Moshiri, N., Mai, U., Jia, X. & Mirarab, S. TreeCluster: Clustering biological sequences using phylogenetic trees. *PLoS One* **14**, e0221068 (2019).

230. Campbell, F. *et al.* outbreaker2: a modular platform for outbreak reconstruction. *BMC Bioinforma. 2018 1911* **19**, 1–8 (2018).

231. Ong, J., Aik, J. & Ng, L. C. Short Report: Adult Aedes abundance and risk of dengue transmission. *PLoS Negl. Trop. Dis.* **15**, e0009475 (2021).

232. Priyamvada, L., Hudson, W., Ahmed, R. & Wrammert, J. Humoral cross-reactivity between Zika and dengue viruses: implications for protection and pathology. *Emerg. Microbes Infect.* **6**, e33 (2017).

233. Harrington, L. C. *et al.* Heterogeneous Feeding Patterns of the Dengue Vector, Aedes aegypti, on Individual Human Hosts in Rural Thailand. *PLoS Negl. Trop. Dis.* **8**, (2014).

234. Grantz, K. H. *et al.* The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat. Commun. 2020 111* **11**, 1–8 (2020).

235. Priyamvada, L. *et al.* Human antibody responses after dengue virus infection are highly cross-reactive to Zika virus. *Proc. Natl. Acad. Sci.* **113**, 7852–7857 (2016).

236. Flipse, J. *et al.* Dengue tropism for macrophages and dendritic cells: the host cell effect. *J. Gen. Virol.* **97**, 1531–1536 (2016).

237. Sasmono, R. T. *et al.* Genomic analysis and growth characteristic of dengue viruses from Makassar, Indonesia. *Infect. Genet. Evol.* **32**, 165–177 (2015).

238. Zhang, L. *et al.* Genetic and pathogenicity diversity of dengue virus type 2 strains circulating in Guangdong, China. *Biosaf. Heal.* **3**, 333–342 (2021).

239. Phan, D. Q. *et al.* The Distribution of Dengue Virus Serotype in Quang Nam Province (Vietnam) during the Outbreak in 2018. *Int. J. Environ. Res. Public Heal. 2022, Vol. 19, Page 1285* **19**, 1285 (2022).

240. Raghwani, J. *et al.* Endemic Dengue Associated with the Co-Circulation of Multiple Viral Lineages and Localized Density-Dependent Transmission. *PLOS Pathog.* **7**, e1002064 (2011).

241. Chew, M. H., Rahman, M. M. & Hussin, S. Molecular epidemiology and phylogenetic analysis of Dengue virus type-1 and 2 isolated in Malaysia. *Pakistan J. Med. Sci.* **31**, 615 (2015).

242. Stoddard, S. T. *et al.* The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl. Trop. Dis.* **3**, (2009).

243. Siqueira, J. B. *et al.* Household survey of dengue infection in central Brazil: Spatial point pattern analysis and risk factors assessment. *American Journal of*

Tropical Medicine and Hygiene **71**, 646–651 (2004).

244. Scott, T. W. & Morrison, A. C. Vector Dynamics and Transmission of Dengue Virus: Implications for Dengue Surveillance and Prevention Strategies. in 115–128 (2010). doi:10.1007/978-3-642-02215-9_9