Molecular Epidemiology of Dengue Viruses from Complete Genome Sequences

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Swee Hoe Ong

aus Malaysia

Basel, 2010

creative commons

Attribution-Noncommercial-No Derivative Works 2.5 Switzerland

---

You are free:

to Share — to copy, distribute and transmit the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Your fair dealing and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license) available in German:
http://creativecommons.org/licenses/by-nc-nd/2.5/ch/legalcode.de

Disclaimer:
The Commons Deed is not a license. It is simply a handy reference for understanding the Legal Code (the full license) — it is a human-readable expression of some of its key terms. Think of it as the user-friendly interface to the Legal Code beneath. This Deed itself has no legal value, and its contents do not appear in the actual license. Creative Commons is not a law firm and does not provide legal services. Distributing of, displaying of, or linking to this Commons Deed does not create an attorney-client relationship.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/deed.en Datum: 3.4.2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Hernn Prof. Dr. Marcel Tanner, Herrn Prof. Dr. Hans-Peter Beck und Dr. Mark J. Schreiber.

Basel, den 28. Mai 2009.

Prof. Dr. Eberhard Parlow
Dekan

Basel, 2010

Table of Contents

Acknowledgements

My sincerest thanks go to Subhash G. Vasudevan for offering me the opportunity to pursue my Ph.D. at the Dengue Unit of the Novartis Institute for Tropical Diseases (NITD). Subhash is the person that started my association with dengue when I first worked under his supervision as an intern back in May 2004.

I would also like to thank Professor Marcel Tanner at the Swiss Tropical Institute for acting as my Faculty Representative and for guiding me in my thesis writing, and Professor Hans-Peter Beck the co-referee for the discussions and suggestions that improved this thesis. My deepest thanks go to Mark J. Schreiber for overseeing my daily activities at NITD and for serving as my mentor throughout the duration of my Ph.D. training.

It has been a pleasure over the past few years working with our collaborators namely Martin L. Hibberd and Pauline Aw Poh Kim (Infectious Diseases, A*STAR Genome Institute of Singapore), Edward C. Holmes (Center for Infectious Disease Dynamics, Penn State University), Ooi Eng Eong (DSO National Laboratories), Harold Soh Soon Hong (A*STAR Institute of High Performance Computing), Tedjo Sasmono (NEHCRI) and Wayne P. Mitchell (A*STAR Experimental Therapeutics Centre).

I would like to thank a number of colleagues at NITD namely Andy Yip Jin Teen, Liu Boping, Liu Wei, Wen Daying, Ivica Res, Celine Nkenfou, Katja Fink, Chen Yen Liang, Gu Feng, Wouter Schul, Wang Qing Yin, Christian Noble and Shi Pei Yong for their guidance and support at various junctures of my work. I would also like to record my gratitude to NITD for their financial support of my study. Last but not least, I thank my beloved wife Khoo Siew Gyan and my family for supporting me throughout the years that culminate in the writing of this thesis that you are now reading.

Basel, Switzerland

May 2009

Summary

Since the dengue virus was first isolated in 1943, studies to understand the pathogenesis of dengue has long been hindered by the lack of suitable experimental models. In light of such crippling deficiency, molecular epidemiological approaches represent a viable route to understanding the role of the virus in dengue pathogenesis. Associations between certain dengue virus genotypes with severe dengue outbreaks may point to strains with increased pathogenicity whereas association between dengue strains with unusual phenotypic traits may help identify the responsible viral genetic determinant.

The availability of the complete genetic blueprint of the dengue virus is essential in order to undertake such molecular epidemiological studies. During the course of this project, over two hundred complete genomes of the dengue virus were generated from clinical samples collected in three dengue-endemic Southeast Asian countries. In addition, a bioinformatics platform integrating a sequence database, sequence retrieval tools, sequence annotation data and a variety of analysis tools was developed for easy management, manipulation and analysis of dengue virus sequence data.

Previous studies have mostly focused on epidemiological events in the Americas and Thailand. In this study, sequence data recovered from dengue epidemics in Indonesia, Malaysia and Singapore have uncovered some of the dengue virus diversity circulating in the region. The three countries appear to share similar pool of dengue viruses, with some viral lineages in sustained circulation since at least the 1970s. Sequencing of historical virus isolates prior to the 2004/2005 epidemics in Indonesia and Singapore revealed that adaptive viral evolution played little or no role in triggering those epidemics. Lastly, a method that utilised all available sequence data from Malaysia was devised to reconstruct the history of dengue virus in that country since the 1960s.

Zusammenfassung

Seit dem erstmaligen Isolieren des Dengue-Viruses in 1943, ist das Fehlen passender experimenteller Modelle ein Hindernis fuer die meisten Studien, welche die Pathogenese von Dengue untersuchen. In Anbetracht derart grundlegender Mängel bietet die molekulare Epidemiologie einen gangbaren Weg zur Untersuchung der Rolle, welche das Virus selbst bei der Pathogenese von Dengue spielt. Verbindungen zwischen gewissen Denguevirus-Genotypen und massiven Ausbrüchen der Krankheit weisen unter Umständen auf Virenstämme mit erhöhter Pathogenizität hin, während Verbindungen zwischen verschiedenen Virenstämmen mit ungewöhnlichen phenotypischen Eigenschaften möglicherweise dabei helfen, den verantwortlichen Faktor des viralen Genoms zu identifizieren.

Für solche molekularepidemiologischen Studien ist es essenziell, ueber den gesamten genetischen Fingerabdruck des Dengue-Virus zu verfuegen. Zu diesem Zweck wurden im Verlaufe dieses Projektes über zweihundert komplette Genome von Dengue-Viren aus klinischen Proben erzeugt, die in drei Ländern Südostasiens gesammelt wurden, in denen Dengue endemisch auftritt. Zusätzlich wurde zur Vereinfachung der Datenverwaltung eine Bioinformatikplattform entwickelt, die eine Datenbank für die Sequenzen, Tools zur Sequenzensuche, zusätzliche Anmerkungen zu den Sequenzen und eine Reihe von Analysetools miteinander integriert.

Der Schwerpunkt bisheriger Studien lag hauptsächlich auf epidemiologischen Ereignissen auf dem amerikanischen Kontinent und in Thailand. In dieser Studie hingegen wurde der Umfang der in Indonesien, Malaysia und Singapur verbreiteten Denguevirus-Vielfalt aus Virensequenzen rekonstruiert, die bei Dengueepidemien in der Region vorgefunden wurden. In den drei Ländern scheint ein ähnliches Reservoir von Viren aufzutreten, wobei einige der Virenstämme mindestens seit den Siebziger Jahren des 20. Jahrhunderts stetig vertreten sind. Die Sequenzierung historischer Einzelproben aus der Zeit vor den Epidemien 2004/2005 in Indonesien und Singapur hat gezeigt, dass adaptive Virenevolution nur eine geringfuegige

oder gar keine Rolle beim Ausbruch einer Epidemie spielt. Ausserdem wurde eine Methode entwickelt, mit der die Entwicklung des Dengue-Virus im Land ab den sechziger Jahren des 20. Jahrhundert, unter Verwendung aller verfügbarer malaysischen Sequenzdaten, rekonstruiert werden kann.

List of Original Publications

This thesis is based on the following papers (listed in chronological order of the date of publication) and other unpublished data.

1. **Swee Hoe Ong**, Jin Teen Yip, Yen Liang Chen, Wei Liu, Syahrial Harun, Erlin Lystiyaningsih, Bambang Heriyanto, Charmagne G Beckett, Wayne P Mitchell, Martin L Hibberd, Agus Suwandono, Subhash G Vasudevan and Mark J Schreiber (2008) Periodic re-emergence of endemic strains with strong epidemic potential-A proposed explanation for the 2004 Indonesian dengue epidemic. *Infect Genet Evol.* 8(2):191–204.

2. Mark J. Schreiber, Edward C. Holmes, **Swee Hoe Ong**, Harold S.H. Soh, Wei Liu, Lukas Tanner, Pauline P.K. Aw, Hwee Cheng Tan, Lee Ching Ng, Yee Sin Leo, Jenny G H. Low, Adrian Ong, Eng Eong Ooi, Subhash G. Vasudevan and Martin L. Hibberd (2009) Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J Virol.* 83(9):4163–4173.

3. **Swee Hoe Ong**, Pauline P.K. Aw, Jin Teen Yip, Jasmin Schmid, Wei Liu, Siew Gyan Khoo, Subhash G. Vasudevan, Pei Yong Shi, Martin L. Hibberd, Shamala Devi, Mark J. Schreiber (2009) Inferring history of dengue virus diversity in Malaysia from sequence data. Manuscript in preparation.

The following papers were published in the course of the Ph.D. study but do not form part of this thesis:

1. Mark J Schreiber, **Swee Hoe Ong**, Richard C G Holland, Martin L Hibberd, Subhash G Vasudevan, Wayne P Mitchell and Edward C Holmes (2007) DengueInfo: A web portal to dengue information resources. *Infect Genet Evol.* 7(4):540–541.

2. In Seok Yang, Chunsun Ryu, Ki Joon Cho, Jin Kwang Kim, **Swee Hoe Ong**, Wayne P. Mitchell, Bong Su Kim, Hee-Bok Oh and Kyung Hyun Kim (2008) IDBD: Infectious Disease Biomarker Database. *Nucleic Acids Res.* 36(Database issue):D455-60.

3. Joseph G. Christenbury, Pauline P.K. Aw, **Swee Hoe Ong**, Mark Schreiber, Duane J. Gubler, Subhash G. Vasudevan, Eng-Eong Ooi and Martin L. Hibberd (2009) A method for sequencing the full genome of all four dengue viruses. Manuscript in preparation.

List of Abbreviations

aa	amino acid
C	Capsid
cDNA	complementary DNA
CDS	coding sequence
CHIKV	Chikungunya Virus
DALYs	Disability-Adjusted Life Years
DENV	Dengue Virus
DENV-1	Dengue Virus type 1
DENV-2	Dengue Virus type 2
DENV-3	Dengue Virus type 3
DENV-4	Dengue Virus type 4
DF	dengue fever
DHF	dengue haemorrhagic fever
DSS	dengue shock syndrome
E	envelope
EDEN	Early DENgue infection and outcome study
ER	endoplasmic reticulum
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
JEV	Japanese Encephalitis Virus
kb	kilobase
kDa	kiloDalton
ICTV	The International Committee on Taxonomy of Viruses
IgA	immunoglobulin A
IgG	immunoglobulin G
IgM	immunoglobulin M
MCMC	Markov Chain Monte Carlo
ML	maximum likelihood

MP	maximum parsimony
MVEV	Murray Valley Encephalitis Virus
NCBI	National Center for Biotechnology Information
NJ	neighbour joining
NS	non-structural
nt	nucleotide(s)
PDB	RCSB Protein Data Bank
RdRp	RNA-dependent RNA polymerase
prM	precursor of membrane
RT-PCR	reverse transcription-polymerase chain reaction
SEA	Southeast Asia
SLEV	St. Louis Encephalitis Virus
TBEV	Tick-Borne Encephalitis Virus
TGN	trans-Golgi network
USUV	Usutu Virus
UTR	untranslated region
WHO	World Health Organization
WNV	West Nile Virus
wt	wild type
YFV	Yellow Fever Virus

Preface

This thesis consists of six chapters. The subject matter is the important tropical arboviral disease *dengue*. Background covering both the dengue disease and its etiological agent, the dengue virus, is provided in the first chapter entitled Literature Review.

Apart from the first and the final sixth chapter, all other chapters are written in manuscript form with their own introduction, material and methods, results, discussion and references. A new section not found in the published manuscripts – Author Contributions – has been added to chapters 2–5 to delineate my contribution to the respective publications.

The main theme of the work covered in this thesis involves generating, organising and analysing nucleotide sequence data of the dengue virus to better understand its molecular epidemiology. In chapters 3 and 4, dengue virus genome sequences obtained from virus samples collected during dengue epidemics in Jakarta, Indonesia in 2004 and in Singapore in 2005 were analysed to answer pertinent questions about the role of virus genetics in causing outbreaks. In Chapter 5, newly-generated complete genome sequence information of the dengue virus were utilised to provide a historical account of the virus diversity in Malaysia since the 1960s, not long after dengue haemorrhagic fever was virologically identified in the region. Chapter 6 summarises the results of the three studies described in chapters 2–5, discusses the implications arising from these results and offers suggestions for future work.

1. Literature Review

1.1 Dengue the Disease

Dengue (pronounced as den'gē) is the most common arboviral (arthropod-transmitted) disease and it also ranks as the most important mosquito-borne viral disease in the world. Some 2.5 billion people living in tropical and sub-tropical regions are at risk of dengue infection, which equates to about two-fifths of humanity (Gubler & Clark, 1995; WHO, 2009). There is an estimated 50-100 million infections occurring globally every year, with 500,000 cases requiring hospitalization and causing 24,000 deaths (Halstead, 1988; WHO, 1997). Furthermore, the number of people living in tropical and sub-tropical regions is set to double by the end of the century (UNEP, 2009; Holden 2009), thus making dengue an unqualified global threat to public health.

1.1.1 Etymology

The term "dengue" is thought to be a Spanish homonym for the Swahili phrase "*ki denga pepo*", meaning a sudden cramp-like seizure by an evil spirit or plague (Christie, 1881). The name "breakbone fever", which is attributed to the excruciating joint pains dengue patients suffer from, is also frequently used in place of dengue.

1.1.2 Symptoms

Dengue is an acute febrile viral disease caused by infection with one of the four serotypes of the dengue virus (DENV1-4). Most dengue infections are asymptomatic while the rest result in a wide spectrum of disease that differs in severity from mild undifferentiated fever, i.e. the classical dengue fever (DF), to the potentially fatal complications known as dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) (Figure 1.1).

Common symptoms of a classic dengue fever patient include fever, fatigue, rash, headache, retro-ocular pain (pain behind the eyes), arthralgia (joint pain) and myalgia (muscle pain), nausea, vomiting and leukopenia

(abnormal deficiency of leukocytes circulating in the blood). More extreme cases could include severe hemorrhage, loss of consciousness and abnormal liver and brain function (Rigau-Pérez *et al*., 1998).

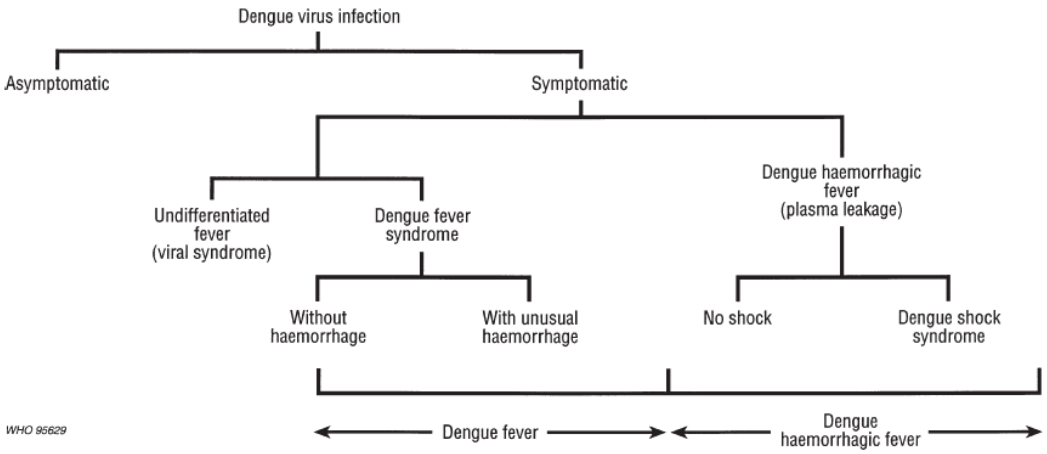


Figure 1.1. Manifestations of dengue virus infection (credit: Figure 2.1 from WHO, 1997).

DHF, the severe form of dengue, is characterised by plasma leakage, thrombocytopenia (low platelet count) and haemorrhagic manifestations. DHF is due to increased vascular permeability believed to be caused by cytokines released when T cells attack dengue-infected cells (Halstead, 2007). The most severe form of dengue disease is DSS which includes all of the symptoms of classic dengue and DHF, with the addition of intense and sustained abdominal pain, persistent vomiting, restlessness or lethargy, a sudden change from fever to hypothermia with sweating and prostration, and shock caused by extremely low blood pressure (Rigau-Pérez *et al*., 1998).

After a patient is infected with dengue virus through the bite of an infected female mosquito, there is an incubation period that can vary between 3 and 14 days. The patient subsequently enters the painful febrile period when viremia is at its peak. Viremia ends 5-7 days after the onset of fever, coincident with defervescence. DHF/DSS usually develops around this time, and intensified observation of the patient is crucial. If DHF develops, the patient may rapidly go into a state of shock and die within 12 to 24 hours if left

untreated. After defervescence, laboratory diagnosis is based on IgG and IgM antibody detection. The disease progression for dengue is presented in schematic form in Figure 1.2.

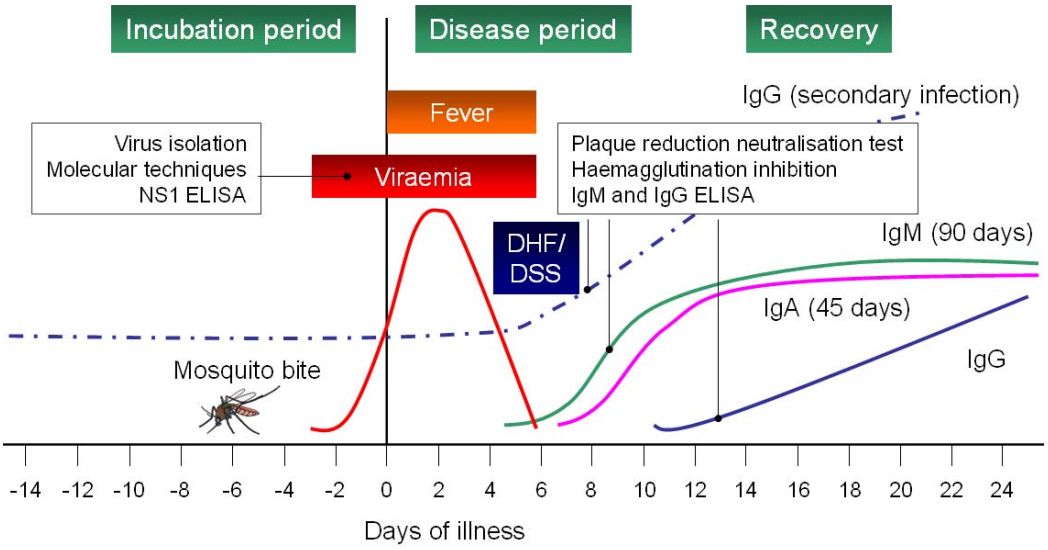


Figure 1.2. Course of dengue infection and the timings and choices of diagnostic methods.

A person could suffer from dengue infection four times throughout his/her lifetime, once for each of the four DENV serotypes. Both primary (first) and secondary (subsequent) infections with any serotype of DENV can result in either the clinically less severe DF or the more severe DHF (Rosen, 1977). A primary dengue infection confers the recovered patient life-long immunity against the infecting serotype and a brief protection against infection by other DENV serotypes (Sabin, 1952). However, epidemiological data and some studies suggest that the immunity thus gained, after the lapse of the temporary cross-serotypic protection, increases the probability of an individual developing DHF when infected by a second heterologous DENV serotype (Halstead *et al*., 1967; Halstead *et al*., 1970). A hypothesis to explain this phenomenon, called antibody-dependent enhancement (ADE), proposes that pre-existing sub-neutralizing antibodies from the primary infection and the second infecting DENV serotype form complexes that bind to cells bearing

Fcγ receptor (FcγR) (monocytes and B cells) leading to increased virus uptake and replication (Figure 1.3) (Halstead, 1988).

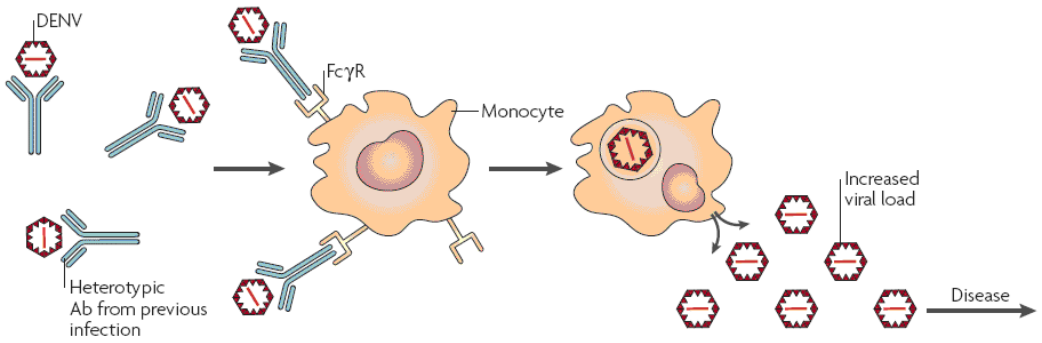


Figure 1.3. Model for antibody-dependent enhancement (ADE) of dengue virus replication (credit: Figure 3 from Whitehead *et al*., 2007).

There is no specific antiviral therapy or vaccine in clinical use for dengue fever. Medical care is supportive in nature and focuses on monitoring and administration of fluids to prevent dehydration and shock, medications to lower fever and reduce pain, and management of bleeding complications. In the late 1960s, DHF fatality has been reported to be as high as 41.3% (Sumarmo, 1987) when healthcare providers understandably were still unfamiliar with the disease. Today, DHF fatality rates can exceed 20% without proper treatment, but can be brought down to 1% with proper medical care (WHO, 1997).

1.1.3 Classification

Although the term “dengue” is commonly used to refer to the entire spectrum of dengue disease, the WHO has devised a formal classification scheme in 1974 that defines dengue as either asymptomatic, DF or DHF/DSS (WHO, 1975). The DHF category is further classified based on the number of haemorrhagic manifestations into four grades of severity (Table 1.1). Grade III and IV of DHF, where profound plasma leakage occurs, are referred to as Dengue Shock Syndrome (DSS).

Table 1.1. WHO case definition for DHF severity.

Classification	Symptoms
Grade I	Fever with other symptoms such as vomiting, headache, muscle and joint pain, skin rash: positive tourniquet test is the only evidence of hemorrhaging.
Grade II	Grade I symptoms and spontaneous bleeding.
Grade III	Failure of circulatory system, clammy skin, rapid and weak pulse, restlessness.
Grade IV	Severe shock, no measurable blood pressure or pulse.

These guidelines were developed based on pediatric cases reported to the Children's Hospital, Bangkok, Thailand in the 1960s. In recent years, clinicians have been reporting difficulties in following these guidelines to classify the disease as dengue has spread globally with a concurrent change in patient demographic profile (Bandyopadhyay *et al*., 2006; Deen *et al*., 206; Rigau-Perez, 2006).

1.1.4 Brief History of Dengue and DHF

DENV-1 was first isolated by Ren Kimura and Susumu Hotta in Japan in 1943 (Kimura and Hotta, 1943). An epidemic of DF involving at least 200,000 cases had occurred between 1942 and 1944 during World War II in Japanese port cities such as Nagasaki, Kobe, and Osaka. The infections originated from persons returning from the tropics, in particular Southeast Asia and the Pacific islands (Hotta, 2000).

A few months after the first isolation of DENV-1 in Japan, Albert Bruce Sabin and Walter Schlesinger isolated DENV-1 from Hawaiian and shortly thereafter, DENV-2 from Papua New Guinean samples (Sabin and Schlesinger, 1945). They demonstrated that these viruses were antigenically related, yet distinct, and they could be distinguished by the hemagglutination inhibition (HI) assay.

Although there were various speculations about the earliest description of *dengue-like* diseases in historical accounts (Halstead, 1980; Henchal and Putnak, 1990), the disease now known as DHF was first recognised in Manila, the Philippines in 1953 (Quinlos *et al.,* 1954). Viruses similar to DENV-1 and DENV-2 were isolated from Manila patients in 1956 by William Hammond and were called DENV-3 and DENV-4. Dengue viruses of multiple serotypes were subsequently isolated from patients of another DHF epidemic in Bangkok, Thailand in 1958 (Hammond, 1960). It is now known all four serotypes of dengue virus can cause DHF.

DHF/DSS outbreaks were mainly restricted to Southeast Asia until the early 1980s (Halstead, 1980). Since then, dengue transmission has intensified and DHF/DSS outbreaks are now frequent in most tropical countries. To this day, DHF/DSS remains a leading cause of hospitalisation and death among children in Southeast Asia. Outside the region, the disease burden of dengue is most acutely felt in Central and South America where 24 countries have reported laboratory-confirmed DHF between 1981 and 1997 (Monath, 1994; Gubler and Clark, 1995; Gubler, 1998).

1.1.5 Vectors and Transmission Cycles

Dengue is transmitted from person to person through the bites of infected female mosquitoes. The etiological agent, the DENV, is believed to have been maintained in sylvatic/enzootic transmission cycles involving non-human primate hosts and vector species living in forests. The virus was transmitted to humans when the two come into contact and thereafter was maintained in continuous human-mosquito cycles in and/or around human population centers (Figure 1.4).

Many species from the genus *Aedes* of the family *Culicidae* are known to transmit DENV, but the principal vector is *Aedes aegypti* which is also the vector of the yellow fever virus (YFV). Moreover, this species transmits a third arboviral disease, chikungunya, which is caused by the chikungunya virus (CHIKV), an alphavirus of the family *Togaviridae*. Chikungunya has similar symptoms as dengue which often made accurate diagnosis difficult. The

Australian naturalist Thomas Lane Bancroft first suggested *Ae. aegypti* as a carrier of dengue fever in 1906 based on epidemiological grounds, and this was confirmed in 1916 by John Burton Cleland (Cleland *et al*., 1916). *Ae. aegypti* is known to be a day-biting mosquito that prefers to breed in domestic and peridomestic water containers. Its adaptation to human habitats and its desiccation-resistant eggs have allowed it to flourish in urban centers.

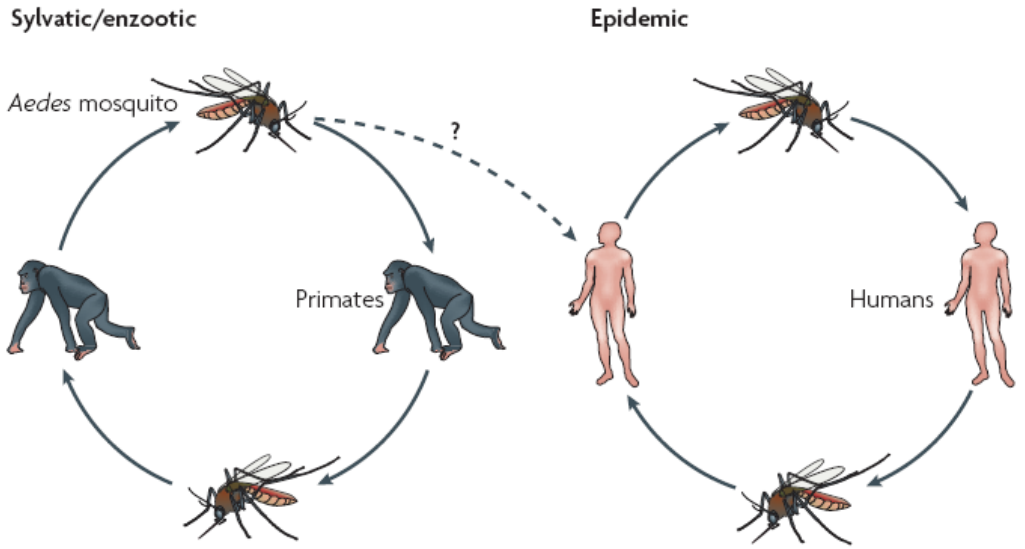


Figure 1.4. Transmission of dengue viruses (credit: Figure 2 from Whitehead *et al*., 2007).

The secondary vector for dengue is *Aedes albopictus* which is commonly known as the Asian tiger mosquito. Its role as dengue vector in semi-tropical regions was first identified by Koizumi *et al*. in Taiwan in 1917 (Kuno, 2007). *Ae. albopictus* serves as the primary vector for dengue in countries where *Ae. aegypti* is absent and as a maintenance vector in rural areas where both species coexist (Smith, 1956; Gratz, 2004). In the Pacific islands *Ae. polynesiensis* has been suggested as the primary dengue vector (Rosen *et al*., 1954; Freier and Rosen, 1987) whereas *Ae. scutellaris* was identified as the 'jungle' vector for dengue (Mackerras, 1946). Similar to *Ae. aegypti*, *Ae. albopictus* is also an efficient vector for CHIKV and it has been implicated in causing major chikungunya epidemics in recent years (Bessaud *et al*., 2006; Bonilauri *et al.*, 2008).

Ae. aegypti and *Ae. albopictus* have both been shown to be anthropophilic, i.e. prefer to feed on humans (Ponlawat and Harrington, 2005) and are widely distributed in both urban and semi-urban areas in the tropics and subtropics. Both species have also been demonstrated to possess high vector competence for the dengue virus (Moncayo *et al*., 2004). In the continued absence of vaccines and specific treatment, effective vector control (either though fogging that kills adult mosquitoes, application of larvicides that target the aquatic stage of mosquitoes, or source reduction that reduces their breeding habitat) is currently the only practical method available for reducing the incidence of dengue disease.

1.1.6 Geographical Distribution

DENV is the world's most geographically widespread arthropod-borne virus and its geographical distribution is inherently tied to the range and habitat of its principal vector mosquitoes (Figure 1.5). Dengue infections are reported in more than one hundred tropical and sub-tropical countries worldwide, mostly in urban and semi-urban areas where the vectors are widely found. Dengue is hyperendemic in many of these urban centers with co-circulation of multiple dengue virus serotypes. In non-tropical regions, dengue is usually the result of infection of international travelers that have visited dengue-endemic areas.

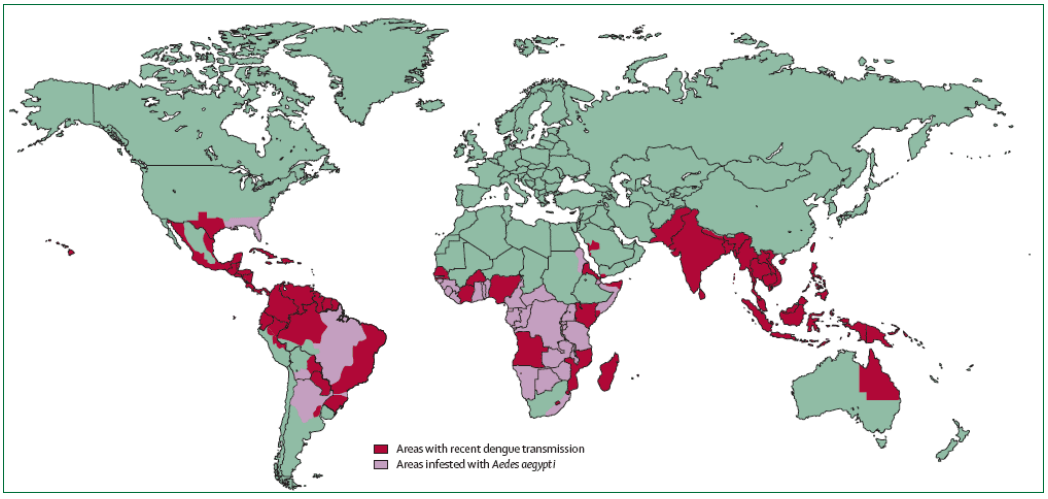


Figure 1.5. Approximate global distribution of dengue and *Aedes aegypti* in 2005 (credit: Figure 1 from Halstead, 2007).

The larvae of the principal vector *Ae. agypti* under naturally changing temperature are capable of developing into adults in conditions lower than 10°C, whereas those of *Ae. albopictus* can survive even lower temperatures (Tsuda and Takagi, 2001). Consequently the two species can be found between latitudes 35°N and 35°S, approximately corresponding to a winter isotherm of 10°C (WHO, 1997). As shown in Figure 1.6, the southern parts of the United States and Europe, and major parts of Australia and Africa are among areas at risk of future dengue transmissions. A dengue outbreak reported in Buenos Aires, Argentina (34°36'S) in early 2009 is very close to this isotherm and is the furthest south dengue has spread.

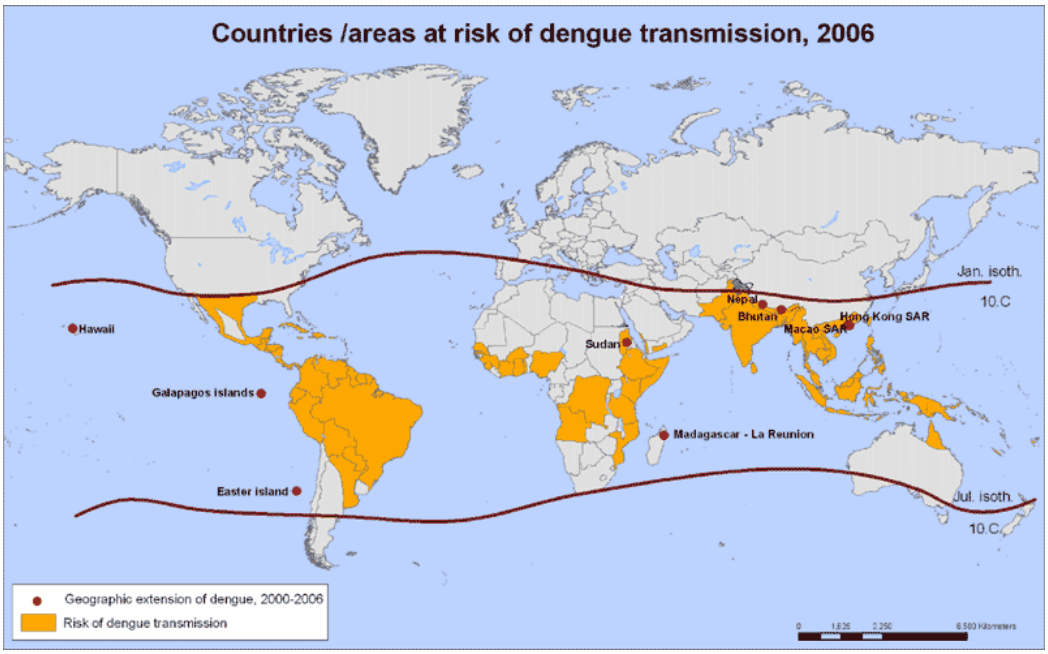


Figure 1.6. Areas at risk of dengue transmission, 2006 (credit: WHO).

1.1.7 Factors Influencing Transmission

Since the etiology of dengue and dengue haemorrhagic fever were virologically described in the mid-1950s, the incidence of dengue worldwide has increased tremendously (Figure 1.7). There is a plethora of inter-related factors that contributed to the prevalence of dengue around the globe. For vector-borne diseases such as dengue, these factors can be categorised into

three obvious components – virus, vector and host – and a less-clear fourth – the surrounding ecology for the three components. All four play important roles in the continued spread and transmission of dengue.

Uncontrolled urbanization, expanding urban population, poverty, ineffective public health infrastructure, faster modes of transportation, globalization of trade and increased international travel have all been implicated as factors leading to the spread of dengue around the world (Gubler and Clark, 1995). Rapid urbanization is probably the single most important contributing factor – the resulting population centers tend to lack public piped water and residents have to resort to using containers to store water which often ended up as breeding sites for the *Ae. aegypti* vector. The lack of adequate sewage systems often leads to the same result.

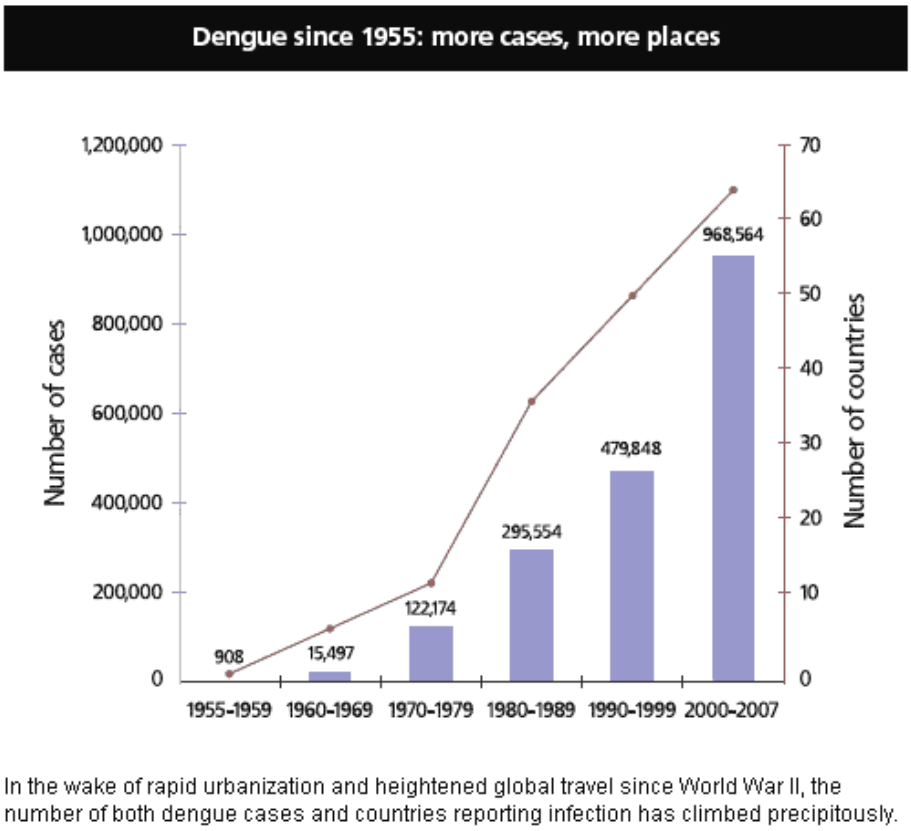


Figure 1.7. Average annual number of dengue (DF) and dengue haemorrhagic fever (DHF) cases reported to WHO and average number of countries reporting dengue (credit: WHO). The figure is accessible at http://www.who.int/disease/dengue/impact/en/index.html.

Inherent differences in the virulence of the introduced DENV strains have also been suggested as being a contributing factor in causing outbreaks and in the emergence of the severe form of dengue disease (Rosen, 1977). An often-cited example is the replacement of the indigenous American genotype of DENV-2 in the Western Hemisphere with one originating from Southeast Asia (Rico-Hesse *et al*., 1997). Viruses of the Southeast Asian genotype have been shown in the laboratory to be better adapted to transmission by the vector *Ae. aegypti* by causing higher viremia in both human dendritic cells and mosquito cells (Armstrong and Rico-Hesse, 2001; Cologna *et al*., 2005; Anderson and Rico-Hesse, 2006).

Relaxation of vector control efforts, expansion of the vector range, and the build-up of vector resistance to insecticides (Gubler and Clark, 1995; Kawada *et al*., 2009) are some of the recognised factors affecting the contribution of the mosquito vector. The impact of environmental factors on the transmission and spread of mosquito-borne diseases - as exemplified by effects of temperature, rainfall and humidity on vector transmission cycles - are also well known (Watts *et al*., 1987). Beside the effect of generalised climatic factors (global warming, for example) the local ecology probably plays an equal, if not more important, role in a disease as complex as dengue (Kuno, 1995; Reiter, 2008; Johansson *et al*., 2009).

1.1.8 Economic impact

Apart from physical pain, dengue also causes economic hardship to recovered individuals in the form of hospitalization costs and disruption of earning potential. The DALYs (disability-adjusted life years) lost to dengue in Southeast Asia were estimated to be comparable to the burden caused by diseases such as HIV, malaria and tuberculosis (Gubler and Meltzer, 1999). At the governmental level, vast amounts of money have to be allocated for public awareness campaigns, medical services and vector eradication efforts. Another indirect cost comes in the form of loss of revenue through reduced tourism (WHO, 1997).

1.2 Dengue the Virus

The causative agent of the dengue disease is the dengue virus (DENV), a group of four flaviviruses that are closely related but antigenically distinct. They are hypothesised to have evolved independently from ancestral sylvatic viruses between 100-1,500 years ago (Wang *et al*., 2000). The four groups are known as *serotypes* and denoted as dengue virus type 1 (DENV-1), dengue virus type 2 (DENV-2), dengue virus type 3 (DENV-3) and dengue virus type 4 (DENV-4).

1.2.1 Taxonomy

There are three genera in the *Flaviviridae* family (formerly known as group B arboviruses) namely *Flavivirus*, *Pestivirus* and *Hepacivirus*. The dengue virus is a member of the genus *Flavivirus* which consists of 55 identified virus species (ICTVdB, 2006). The word *Flavi* is a derivation from the Latin "*flavus*" which means "yellow" and the type species of the genus is the yellow fever virus (YFV). The flaviviruses are thus named due to the jaundice observed in yellow fever patients. Many flaviviruses are important human pathogens, most notably the dengue viruses, yellow fever virus, Japanese encephalitis virus (JEV), West Nile virus (WNV) and tick-borne encephalitis virus (TBEV). The flaviviruses are predominantly transmitted by mosquitoes and ticks, whereas some have no known vector.

Dengue was one of the groups classified when early researchers divided the flaviviruses serologically into eight antigenic complexes using cross-neutralization tests. However, many viruses, for example the prototype of the genus YFV, could not be affiliated with any complexes (Calisher *et al*., 1989). When sequence data became available, phylogenetic inference from molecular data showed agreement with the antigenic complex classification. In addition, it revealed the clear clustering of the *Flavivirus* genus into non-vector and vector-borne virus clusters, with the latter splitting into mosquito-borne and tick-borne virus clusters (Kuno *et al*., 1998). As shown in Figure 1.8, the mosquito-borne virus cluster has been shown to further diverge into YFV, JEV, and dengue viruses, in that order (Zanotto *et al*., 1996).

The dengue virus was divided into four groups called *serotypes* based on antigenic properties. Subsequent evidence from molecular data reaffirmed this classification and also provided a clearer understanding of the phylogeny of the four serotypes: among the dengue viruses, DENV-4 diverged first from the common ancestor, followed by DENV-2, and finally DENV-1 and DENV-3 (Zanotto *et al*., 1996).

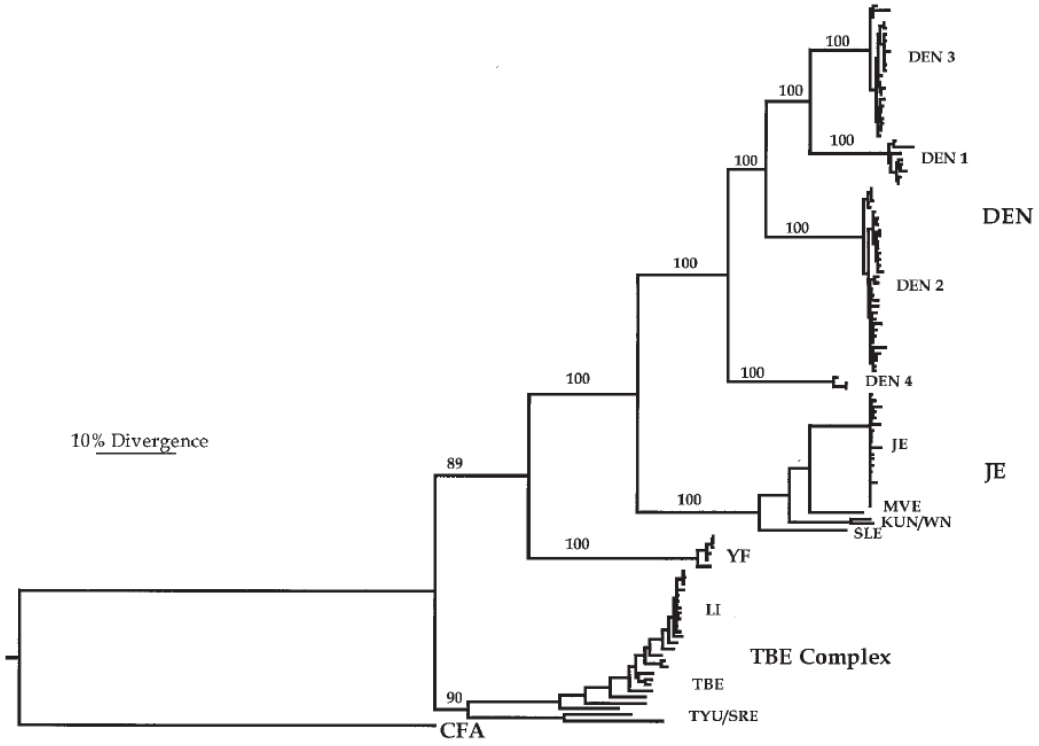


Figure 1.8. Maximum likelihood tree for the E gene from 123 flaviviruses. The tree is rooted by the sequence from *Aedes albopictus* cell fusion agent (CFA) virus (credit: Figure 1 from Zanotto *et al*., 1996).

1.2.2 Virion Morphology

The dengue virus virion, like those of other flaviviruses, is spherical and 40-50 nm in diameter. It is comprised of a nucleocapsid about 30 nm in diameter that is enclosed in a lipid envelope. The nucleocapsid contains the viral capsid and RNA genome. The lipid-containing envelope consists of a lipid bilayer, an envelope protein between 51,000 and 59,000 daltons that mediates attachment, fusion, and penetration, and a small non-glycosylated

internal matrix protein of approximately 8,500 daltons. The envelope protein is glycosylated in most flaviviruses and is exposed on the virion surface. Electron microscopy studies have shown that mature dengue virions are characterised by a relatively smooth surface, as shown in Figure 1.9, with 180 copies of the envelope protein forming the icosahedral scaffold (Kuhn *et al*., 2002).

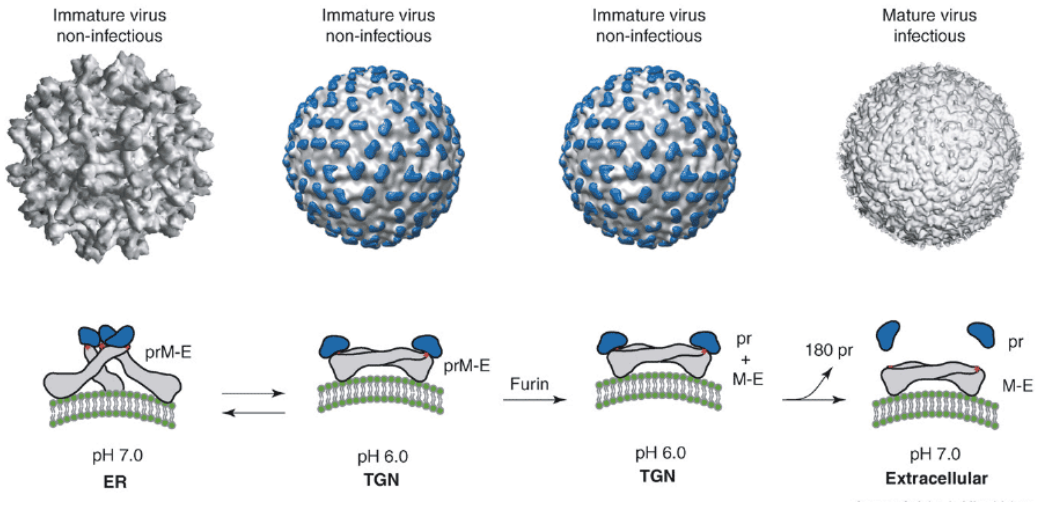


Figure 1.9. Structure of the dengue virion and conformations of the E protein (credit: Figure 2 from Perera and Kuhn, 2008). ER: endoplasmic reticulum; TGN: trans-Golgi network; prM: precursor of membrane.

1.2.3 Genomic Organization

The genomic organization of the dengue virus, and by extension all flaviviruses, is relatively simple compared to other arboviral families such as the *Togaviridae* (formerly known as group A arboviruses), *Bunyaviridae* or *Rhabdoviridae*. The DENV genome consists of a single-stranded, positive-sense RNA molecule roughly 10.7 kb in size. It contains a single translated open reading frame (ORF) that encodes a precursor polypeptide of around 3390 amino acids which is processed catalytically into ten viral proteins (Table 1.2). There is no evidence of alternative or overlapping reading frames that are translated and there is also no hyper-variable region in the DENV genome like those reported in the HCV genome.

Table 1.2. Typical lengths of the ten DENV proteins determined from multiple sequence alignments of deduced amino acid sequences derived from complete genome sequences in GenBank.

Proteins	**DENV-1**	**DENV-2**	**DENV-3**	**DENV-4**
C	114	114	113	113
prM/M	166	166	166	166
E	495	495	493	495
NS1	352	352	352	352
NS2A	218	218	218	218
NS2B	130	130	130	130
NS3	619	618	619	618
NS4A	150	150	150	150
NS4B	249	248	248	245
NS5	899	900	900	900
Length of CDS	**3392**	**3391**	**3390**	**3387**

The DENV ORF is flanked at its 5' terminus by an untranslated region (UTR) of about 100 nucleotides and a longer UTR of about 500 nucleotides at its 3' terminus. The 5' terminus of the genome has a type I cap (m^7GpppAmp) and there is no polyadenylation of the 3' terminus (reviewed in Chambers *et al*., 1990). The translated polyprotein is cleaved co- and post-translationally by viral and host proteases into ten viral proteins: three structural proteins (C, capsid; prM/M, precursor of membrane; E, envelope) encoded at the 5' end of the ORF, and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) encoded at the 3' end (Figure 1.10).

The three structural proteins constitute the DENV virion: the capsid protein surrounds the viral RNA genome to form the nucleocapsid, whereas the prM and E proteins are embedded in the lipid bilayer that forms the viral envelope. Cleavage of the prM into the membrane (M) protein by furin during viral release has been shown to be a prerequisite for the production of mature infectious virions. Of the three structural proteins, the E protein is the most

studied as it is the major constituent of the virus envelope. It is glycosylated at two sites (Asn-67 and Asn-153) and is responsible for virus attachment to receptors of susceptible host cells and for fusion with cell membranes. The E glycoprotein also contains the main epitopes recognised by neutralizing antibodies (reviewed in Chambers *et al*., 1990). Such epitopes are also found to a lesser extent on the M glycoprotein (Kaufman *et al*., 1989).

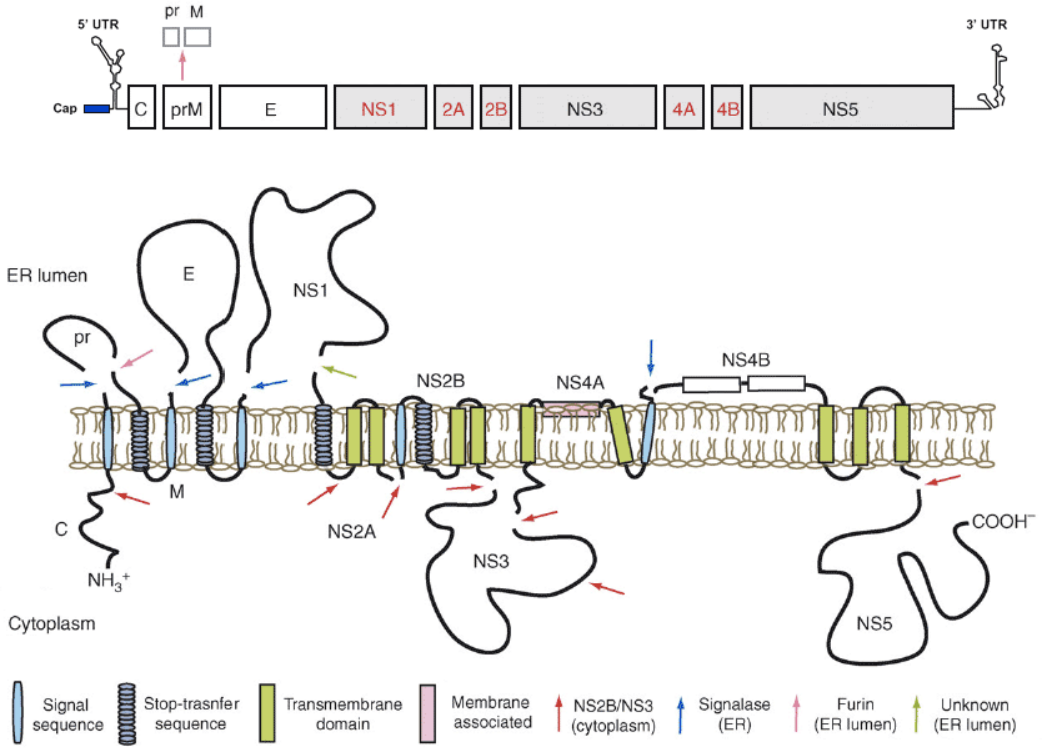


Figure 1.10. Schematic diagram showing: (top) gene organisation in the dengue virus RNA genome, (bottom) the membrane topology and proteolytic cleavage sites of the transcribed polyprotein. Cellular and viral proteases, which are denoted by arrows, process the immature polyprotein into ten separate proteins (credit: adaptation of Figure 1 from Perera and Kuhn, 2008).

The 3' end of the DENV genome encodes seven non-structural (NS) proteins of various sizes in the order: NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. Some non-structural proteins are known to be multi-functional while little is known about NS1, NS2A and NS4A/4B. Functions of the non-structural proteins are summarised in Table 1.3.

Table 1.3. Known and possible functions of dengue non-structural proteins (reviewed in Perera and Kuhn, 2008).

NS proteins	Description of known functions
NS1	Plays a role in viral RNA replication complex; acts as soluble complement-fixing antigen.
NS2A	Forms part of the RNA replication complex.
NS2B	Co-factor for NS3 protease.
NS3	Serine protease, RNA helicase and RTPase/NTPase.
NS4A	Possibly induces membrane alterations important for virus replication.
NS4B	Possibly blocks IFN α/β-induced signal transduction.
NS5	Methyltransferase (MTase) and RNA-dependent RNA polymerase (RdRp).

1.2.4 Genetic Diversity

Based on available molecular data it is well known that there is great genetic diversity among the dengue viruses. The factors that contributed to this are many fold, so are the epidemiological implications arising from this diversity, as illustrated in Figure 1.11.

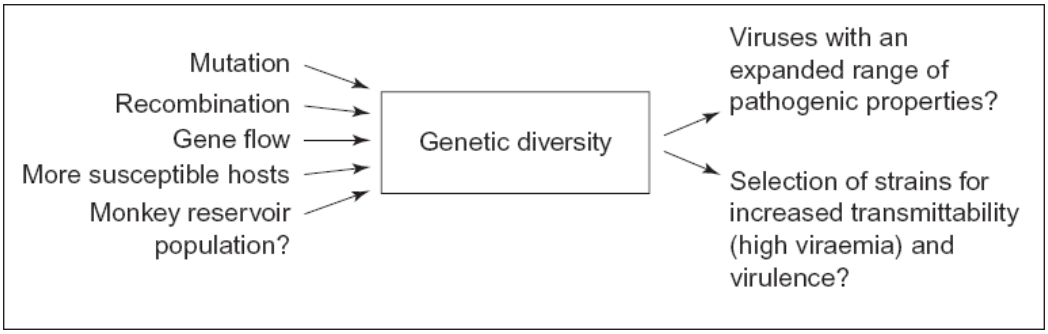


Figure 1.11. The processes that have caused an increase in the genetic diversity of dengue virus and two possible evolutionary consequences of this increase (credit: Figure 1 from Holmes and Burch, 2000).

For most purposes the four dengue serotypes are generally treated as the same virus and the diseases they cause are considered as the same disease. However, the genetic distances between the four serotypes are greater than the distances between many of the recognised virus species in the genus, for example between the Japanese Encephalitis virus (JEV), West Nile virus (WNV), Murray Valley Encephalitis virus (MVEV), Usutu virus (USUV) and St Louis Encephalitis virus (SLEV) (Figure 1.12). Based on this observation, others have argued that the four dengue serotypes warrant the rank of species on their own right (Kuno *et al*., 1998; Holmes and Burch, 2000).

Each serotype of the dengue virus can be further classified into several genetic groups called *genotypes* (the term *subtype* is used interchangeably) based on sequence diversity. Rico-Hesse (1990) initially defined a dengue genotype as a group of dengue viruses having no more than 6% sequence divergence within a 240-nucleotide region of the DENV-1 and DENV-2 E/NS1 junction. Since then, both the length and region of virus genome selected for sequencing varied greatly depending on research groups, ranging from the complete sequence of single genes to the complete genome of the DENV. Assignment of genotypes now relies on phylogenetic analysis rather than arbitrary cut-off values in sequence diversity.

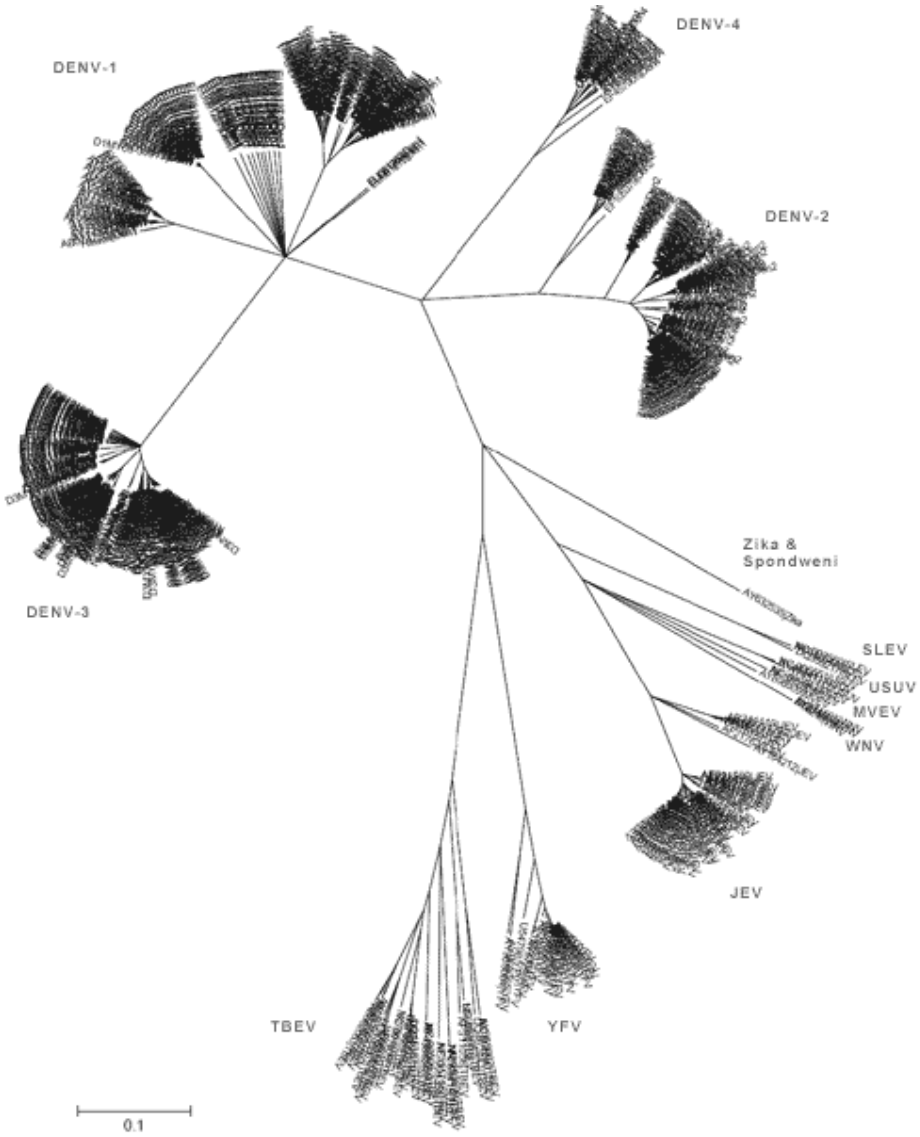


Figure 1.12. Unrooted minimum evolution tree for the complete E gene nucleotide sequences from 554 flaviviruses (unpublished data). Genetic distances between the four dengue serotypes are greater than the distances between the many species within the tick-borne encephalitis and Japanese encephalitis antigenic complexes.

Rico-Hesse (2003) and Vasilakis and Weaver (2008) have published excellent and detailed descriptions of the genotype classification for all four dengue serotypes. The following paragraphs describe only the essential points of the subject matter.

DENV-1 can be divided into five genotypes based on the complete E gene sequence as described by Goncalvez *et al*. (2002). Earlier work by Rico-Hesse (1990) also classified DENV-1 into five groups based on the 240-nucleotide E/NS1 junction sequences, but with some minor differences from the newer scheme which is listed in Table 1.4.

Table 1.4 DENV-1 genotypes according to Goncalvez *et al*. (2002).

Genotypes	**Original known distribution**
I	Japan, Hawaii in the 1940s (the prototype strains), China, Taiwan and Southeast Asia.
II	Thailand in the 1950s and 1960s.
III	Sylvatic source in Malaysia.
IV	Nauru, Australia, Indonesia and the Philippines.
V	Africa, Southeast Asia and the Americas.

The DENV-1 genotypes all have a wide area of distribution apart from genotype III (sylvatic) and genotype II which consists of Thai strains from the 1950s and 1960s. Viruses of genotype I and IV have recently been implicated as causing epidemics in the Pacific between 2000 and 2004 (A-Nuegoonpipat *et al*., 2004) and genotype V viruses are frequently isolated during epidemics in the Americas (Avilés *et al*., 2003). However, it is still inconclusive whether any of these three DENV-1 genotypes can be consistently associated with causing more severe dengue (Rico-Hesse, 2003).

DENV-2 is the most studied serotype among the dengue viruses. Twiddy *et al*. (2002) proposed the existence of six genotypes of DENV-2 (Table 1.5) based on the complete E gene sequence following earlier work by Rico-Hesse (1990) and Lewis *et al*. (1993). Sylvatic DENV-2 strains that are closely related have been isolated from several countries in West Africa and Malaysia, two locations that are far apart, leading Wang *et al.* (2000) to hypothesise that the DENV sylvatic ancestor arose in the Asian-Oceanic region before diverging into today's four DENV serotypes.

The first DHF epidemic in the Americas occurred after an introduction of the Asian II genotype to Cuba in 1981 (Guzman *et al*., 1995). Likewise, the America/Asian genotype (genotype III) has been reported to have replaced the pre-existing American genotype (genotype V) in the Western Hemisphere (Rico-Hesse *et al*., 1997) and is considered to be the DENV-2 genotype with the highest epidemiological impact (Rico-Hesse, 2003).

Table 1.5 DENV-2 genotypes according to Twiddy *et al*. (2002).

Genotypes	**Original known distribution**
American	Formerly known as subtype V. Found in Latin America, old strains from India (1957), the Caribbean, and the Pacific islands between 1950 and 1970s.
American/Asian	Formerly known as subtype III. Found in China, Vietnam, Thailand and in Latin America since the 1980s.
Asian I	Thailand, Myanmar and Malaysia.
Asian II	Formerly known as subtype I and II. Found in China, the Philippines, Sri Lanka, Taiwan and Vietnam. Includes the New Guinea C prototype strain.
Cosmopolitan	Formerly known as genotype IV. Wide distribution including Australia, the Pacific islands, Southeast Asia, the Indian subcontinent, Indian Ocean islands, Middle East, and both East and West Africa.
Sylvatic	Isolated from non-human primates in West Africa and Malaysia.

The current genotype classification for DENV-3 follows the nomenclature proposed by Lanciotti *et al.* (1994) which recognised four DENV-3 genotypes based on prM/E sequences (Table 1.6). These four genotypes are similar to the four groups described by Chungue *et al*. (1993) using a 195-nucleotide region at the 5' terminus of the E gene.

Introduced to the Americas via Nicaragua in 1994, genotype III DENV-3 is now widely found in Central and Southern America (Balmaseda *et al*., 1999; Usuku *et al*., 2001; Messer *et al*., 2003) and is considered as the most virulent of the four DENV-3 genotypes. It is worthy of note that genotype IV has never been associated with any DHF epidemics (Lanciotti *et al.*, 1994). Although their existence is anticipated through the presence of DENV-3 antibodies in non-human canopy-dwelling primates, no sylvatic lineage of DENV-3 has been found thus far (Rudnick, 1984).

Table 1.6. DENV-3 genotypes according to Lanciotti *et al*. (1994) and the known distribution of the genotypes prior to 1993.

Genotypes	Original known distribution
I	Indonesia, Malaysia, Thailand, Burma, Vietnam, the Philippines and the South Pacific islands (French Polynesia, Fiji and New Caledonia). Includes the H87 prototype strain.
II	Thailand, Vietnam and Bangladesh.
III	Singapore, Indonesia, South Pacific islands, Sri Lanka, India, Africa and Samoa.
IV	Puerto Rico and French Polynesia (Tahiti).

Lanciotti *et al*. (1997) initially separated DENV-4 into two genotypes, I and II, based on the complete E gene sequence. A further two genotypes were subsequently described (Table 1.7), with one found only in non-human primates in Malaysia and another, genotype III, found only in Bangkok, Thailand (Klungthong *et al*., 2004). Genotype II DENV-4 is the most widespread of the four following an introduction to the Western hemisphere in 1981, possibly via the Pacific islands (Lanciotti *et al*., 1997; Foster *et al*.,

2003). Although DENV-4 is the least frequently sampled serotype, it is often associated with haemorrhagic fever during secondary infection (Vaughn, 2000).

Table 1.7. DENV-4 genotypes and their known distribution.

Genotypes	Original known distribution
I	Thailand, Malaysia, the Philippines and Sri Lanka. Includes the H241 prototype strain.
II	Indonesia, Malaysia, Tahiti, the Caribbean islands (Puerto Rico and Dominica) and the Americas.
III	Thailand (Bangkok, specifically).
Sylvatic	Isolated from non-human primates in Malaysia.

1.2.5 Role of virus genetics and evolution

Except for the sylvatic genotypes, genotype classification can often unveil the geographical origin of the dengue virus strains. This has enabled tracking the route of virus transmissions across distant time and place, and has served as the basis of molecular epidemiological studies that can determine whether dengue epidemics are caused by introduction of new viruses or the result of re-emergence of endemic strains.

Introduction of new viruses inevitably leads to the question whether particular genotypes of DENV are associated with higher virulence or severe disease. To date, severe disease has often been associated with several DENV genotypes originating in Southeast Asia (Guzman *et al*., 1995; Rico-Hesse *et al*., 1997). The lack of a suitable animal model for the dengue disease, however, means such hypotheses cannot be easily verified (Rico-Hesse, 2003). On the other hand, association of re-emergence of endemic strains with outbreaks leads to a different question that can only be answered by a combination of classic epidemiology and comparative genomics: whether the viruses re-emerged due to environmental, population immunity and/or

vectorial factors, or whether outbreaks were triggered by adaptive evolution of the virus that endowed it with an increase in fitness and virulence?

Possession of the complete genetic blueprint of the dengue viruses is a prerequisite to answering these crucial questions. This thesis describes efforts to sequence and then examine the complete genome sequences of DENV isolates from two recent epidemics in Indonesia and Singapore. The results from these two studies showed that the re-emergence of endemic strains is likely to be the main cause of most dengue outbreaks in Southeast Asian countries. No evidence of mutational signatures that could serve as a trigger of epidemics was found in isolates collected up to six years prior to the epidemic. Lastly, a new way of reconstructing the history of dengue virus diversity from all existing sequence data was introduced. The result showed that Malaysia, a country occupying a central position in Southeast Asia, has experienced both frequent importations of DENV strains from neighbouring countries and maintenance of endemic viral lineages which have been in sustained transmission for many decades.

1.3 References

1. A-Nuegoonpipat A, Berlioz-Arthaud A, Chow V, Endy T, Lowry K, Mai le Q, Ninh TU, Pyke A, Reid M, Reynes JM, Su Yun ST, Thu HM, Wong SS, Holmes EC and Aaskov J (2004) Sustained transmission of dengue virus type 1 in the Pacific due to repeated introductions of different Asian strains. *Virology*. 329:505–512.
2. Anderson JR and Rico-Hesse R (2006) *Aedes aegypti* vectorial capacity is determined by the infecting genotype of dengue virus. *Am J Trop Med Hyg*. 75(5):886–892.
3. Armstrong PM and Rico-Hesse R (2001) Differential susceptibility of *Aedes aegypti* to infection by the American and Southeast Asian genotypes of dengue type 2 virus. *Vector Borne Zoonotic Dis*. 1(2):159–168.

4. Avilés G, Meissner J, Mantovani R and St Jeor S (2003) Complete coding sequences of dengue-1 viruses from Paraguay and Argentina. *Virus Res*. 98(1):75-82.
5. Balmaseda A, Sandoval E, Pérez L, Gutiérrez CM and Harris E (1999) Application of molecular typing techniques in the 1998 dengue epidemic in Nicaragua. *Am J Trop Med Hyg*. 61(6):893–897.
6. Bessaud M, Peyrefitte CN, Pastorino BAM, Tock F, Merle O, Colpart J, Dehecq J, Girod R, Jaffar-Bandjee M, Glass PJ, Parker M, Tolou H and Grandadam M (2006) Chikungunya virus strains, Reunion Island outbreak. *Emer Infec Dis.* 12(10):1604–1606.
7. Bonilauri P, Bellini R, Calzolari M, Angelini R, Venturi L, Fallacara F, Cardioli P, Angelini P, Venturelli C, Merialdi G and Dottori M (2008) Chikungunya virus in *Aedes albopictus*, Italy. *Emer Infec Dis.* 14(5):852–854.
8. Calisher CH, Karabatsos N, Dalrymple JM, Shope RE, Porterfield JS, Westaway EG and Brandt WE (1989) Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. *J Gen Virol.* 70:37–43.
9. Chambers TJ, Hahn CS, Galler R and Rice CM (1990) Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol.* 44:649–688.
10. Christie, J. (1881). On epidemics of dengue fever: Their diffusion and etiology. *Glasgow Med. J*. 16:161–176.
11. Chungue E, Deubel V, Cassar O, Laille M and Martin PM (1993) Molecular epidemiology of dengue 3 viruses and genetic relatedness among dengue 3 strains isolated from patients with mild or severe form of dengue fever in French Polynesia. *J Gen Virol.* 74:2765–2770.
12. Chungue E, Cassar O, Drouet MT, Guzman MG, Laille M, Rosen L and Deubel V (1995) Molecular epidemiology of dengue-1 and dengue-4 viruses. *J Gen Virol*. 76:1877–1884.
13. Cleland JB, Bradley B and McDonald W (1916) On the transmission of Australian dengue by the mosquito *Stegomyia fasciata*. *Med J Aust*. 2:179–205.

14. Cologna R, Armstrong PM and Rico-Hesse R (2005) Selection for virulent dengue viruses occurs in humans and mosquitoes. *J Virol*. 79(2):853–859.
15. Deen JL, Harris E, Wills B, Balmaseda A, Hammond SN, Rocha C, Dung NM, Hung NT, Hien TT and Farrar JJ (2006) The WHO dengue classification and case definitions: time for a reassessment. *Lancet*. 368:170–173.
16. Freier JE and Rosen L (1987) Vertical transmission of dengue viruses by mosquitoes of the *Aedes scutellaris* group. *Am J Trop Med Hyg*. 37(3):640–647.
17. Foster JE, Bennett SN, Vaughan H, Vorndam V, McMillan WO and Carrington CV (2003) Molecular evolution and phylogeny of dengue type 4 virus in the Caribbean. *Virology*. 306(1):126–34.
18. Goncalvez AP, Escalante AA, Pujol FH, Ludert JE, Tovar D, Salas RA and Liprandi F (2002) Diversity and evolution of the envelope gene of dengue virus type 1. *Virology*. 303:110–119.
19. Gratz NG (2004) Critical review of the vector status of *Aedes albopictus*. *Med Vet Entomol*. 18(3):215–227.
20. Gubler DJ (1998) Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev.* 11:480–496.
21. Gubler DJ and Clark GG (1995) Dengue/Dengue Hemorrhagic Fever: the emergence of a global health problem. *Emer Infec Dis.* 1(2):55–57.
22. Gubler DJ and Meltzer M (1999) Impact of dengue/dengue hemorrhage fever on the developing world. *Adv Virus Res*. 53:35–70.
23. Guzman MG, Deubel V, Pelegrino JL, Rosario D, Marrero M, Sariol C and Kouri G (1995) Partial nucleotide and amino acid sequences of the envelope and the envelope/nonstructural protein-1 gene junction of four dengue-2 virus strains isolated during the 1981 Cuban epidemic. *Am J Trop Med Hyg*. 52(3):241–246.
24. Halstead SB (1980) Dengue haemorrhagic fever – a public health problem and a field for research. *Bull World Health Organ*. 58(1): 1–21.
25. Halstead SB (1988) Pathogenesis of dengue: challenges to molecular biology. *Science* 239: 476–481.
26. Halstead SB (2007) Dengue. *Lancet*. 370(9599):1644–1652.

27. Halstead SB, Nimmannitya S, Yamarat C and Russell PK (1967) Hemorrhagic fever in Thailand: newer knowledge regarding etiology. *Jpn J Med Sci Biol*. 20(suppl):96–103.
28. Halstead SB, Nimmannitya S and Cohen SN (1970) Observations related to pathogenesis of dengue hemorrhagic fever. IV. Relation of disease severity to antibody response and virus recovered. *Yale J Biol Med*. 42:311–328.
29. Hammon WM, Rudnick A and Sather GE (1960) Viruses associated with epidemic hemorrhagic fevers of the Philippines and Thailand. *Science.* 131:1102–1103.
30. Henchal ER and Putnak R (1990) The dengue viruses. *Clin Microbiol Rev*. 3(4):376–379.
31. Holden C (2009) Higher temperatures seen reducing global harvests. *Science.* 323(5911):240–244.
32. Holmes EC and Burch SS (2000) The causes and consequences of genetic variation in dengue virus. *Trends Microbiol*. 8(2):74–77.
33. Hotta S (2000) Dengue fever and dengue virus – a challenge to tropical medicine. *Jpn J Trop Med Hyg.* 28:369–381.
34. ICTVdB - The Universal Virus Database, version 4 (2006) Büchen-Osmond, C (Ed), Columbia University, New York, USA. (http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_flavi.htm).
35. Johansson MA, Dominici F and Glass GE (2009) Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Negl Trop Dis*. 3(2):e382.
36. Kawada H, Higa Y, Nguyen YT, Tran SH, Nguyen HT and Takagi M (2009) Nationwide investigation of the pyrethroid susceptibility of mosquito larvae collected from used tires in Vietnam. *PLoS Negl Trop Dis*. 3(3):e391.
37. Kaufman BM, Summers PL, Dubois DR, Cohen WH, Gentry MK, Timchak RL, Burke DS and Eckels KH (1989) Monoclonal antibodies for dengue virus prM glycoprotein protect mice against lethal dengue infection. *Am J Trop Med Hyg*. 41:576–580.
38. Kimura R and Hotta S (1943) Experimental inoculation of dengue virus into mice. *Nippon Igaku*. 3344:1378–1379.

39. Klungthong C, Zhang C, Mammen MP Jr, Ubol S and Holmes EC (2004) The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. *Virology*. 329:168–179.
40. Koizumi T, Yamaguchi K and Tonomura K (1917) [Research on dengue]. *Taiwan-Igakukai-Zasshi* [Taiwan J Med] 176:369–392.
41. Kuhn RJ, Zhang W, Rossman MG, Pletnev SV, Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR and Strauss JH (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell*. 108:717–725.
42. Kuno G (1995) Review of the factors modulating dengue transmission. *Epidemiol Rev*. 17(2):321–335.
43. Kuno G (2007) Research on dengue and dengue-like illness in East Asia and the Western Pacific during the first half of the 20th century. *Rev Med Virol*. 17:327–341.
44. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N and Cropp CB (1998) Phylogeny of the genus Flavivirus. *J Virol.* 72:73–83.
45. Lanciotti RS, Lewis JG, Gubler DJ and Trent DW (1994) Molecular evolution and epidemiology of dengue-3 viruses. *J Gen Virol*. 75:65–75.
46. Lanciotti RS, Gubler DJ and Trent DW (1997) Molecular evolution and phylogeny of dengue-4 viruses. *J Gen Virol*. 78:2279–2286.
47. Lewis JA, Chang GJ, Lanciotti R, Kinney RM, Mayer LW and Trent DW (1993) Phylogenetic relationships of dengue-2 viruses. *Virology*. 197:216–224.
48. Mackerras IM (1946) Transmission of dengue fever by *Aedes (Stegomyia) scutellaris* Walk. in New Guinea. *Trans Roy Soc Trop Med Hyg*. 40:295–312.
49. Messer WB, Gubler DJ, Harris E, Sivananthan K and de Silva A (2003) Emergence and global spread of a dengue serotype 3, subtype III virus. *Emer Infec Dis*. 9:800–809.
50. Monath TP (1994) Dengue: The risk to developed and developing countries. *Proc Natl Acad Sci USA*. 91:2395–2400.
51. Moncayo AC, Fernandez Z, Ortiz D, Diallo M, Sall A, Hartman S, Davis CT, Coffey L, Mathiot CC, Tesh RB and Weaver SC (2004) Dengue

emergence and adaptation to peridomestic mosquitoes. *Emerg Infect Dis*. 10(10):1790–1796.

52. Perera R and Kuhn RJ (2008) Structural proteomics of dengue virus. *Annu Rev Microbiol*. 11:369–377.
53. Ponlawat A and Harringon LC (2005) Blood feeding patterns of *Aedes aegypti* and *Aedes albopictus* in Thailand. *J Med Entomol*. 42(5):844–849.
54. Quinlos FN, Lim LE, Juliano A, Reyes A and Lacson P (1954) Haemorrhagic fever observed among children in the Philippines. *Philip J Pediatrica*. 3:1–19.
55. Reiter P (2008) Climate change and mosquito-borne disease: knowing the horse before hitching the cart. *Rev Sci Tech*. 27(2):383–398.
56. Rico-Hesse (1990) Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology*. 174:479–493.
57. Rico-Hesse R, Harrison LM, Salas RA, Tovar D, Nisalak A, Ramos C, Boshell J, de Mesa MTR, Nogueira RMR and da Roasa AT (1997) Origins of dengue type-2 viruses associated with increased pathogenicity in the Americas. *Virology*. 230:244–251.
58. Rico-Hesse (2003) Microevolution and virulence of dengue viruses. *Adv Virus Res*. 59:315–341.
59. Rigau-Pérez JG (2006) Severe dengue: the need for new case definitions. *Lancet Infect Dis*. 6:297–302.
60. Rigau-Pérez JG, Clark GG, Gubler DJ, Reiter P, Sanders EJ and Vorndam AV (1998) Dengue and dengue haemorrhagic fever. *Lancet*. 352:971–977.
61. Rosen L (1977) The emperor's new clothes revisited, or reflections on the pathogenesis of dengue hemorrhagic fever. *Am J Trop Med Hyg*. 26(3):337–343.
62. Rosen L, Rozeboom LE, Sweet BH and Sabin AB (1954) The transmission of dengue by *Aedes polynesiensis* Marks. *Am J Trop Med Hyg*. 3(5):878–882.
63. Rudnick A (1984) The ecology of the dengue virus complex in peninsular Malaysia, p. 7. In T. Pang and R. Pathmanathan (ed.), Proceedings of

the International Conference on Dengue/DHF. University of Malaysia Press, Kuala Lumpur, Malaysia.

64. Sabin AB (1952) Research on dengue during World War II. *Am J Trop Med Hyg*. 1(1):30–50.
65. Sabin AB and Schlesinger RW (1945) Production of immunity to dengue with virus modified by propagation in mice. *Science*. 101:640–642.
66. Smith CEG (1956) The history of dengue in tropical Asia and its probable relationship to the mosquito *Aedes aegypti*. *J Trop Med Hyg*. 59(10):243–251.
67. Sumarmo (1987) Dengue haemorrhagic fever in Indonesia. *Southeast Asian J Trop Med Pub Hlth*. 18(3):269–274.
68. Tsuda Y and Takagi M (2001) Survival and development of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) larvae under a seasonally changing environment in Nagasaki, Japan. *Environ Entomol*. 30(5):855–860.
69. Twiddy SS, Farrar JJ, Nguyen VC, Wills B, Gould, EA, Gritsun T, Lloyd G and Holmes EC (2002) Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology*. 298:63–72.
70. United Nations Environment Programme (2009) The Environment in the News. (http://www.unep.org/cpi/briefs/2009Jan09.doc, January 9, 2009).
71. Usuku S, Castillo L, Sugimoto C, Noguchi Y, Yogo Y and Kobayashi N (2001) Phylogenetic analysis of dengue-3 viruses prevalent in Guatemala during 1996–1998. *Arch Virol*. 146:1381–1390.
72. Vasilakis N and Weaver SC (2008) The history and evolution of human dengue emergence. *Adv Virus Res*. 72:1–76.
73. Vaughn DW (2000) Dengue lessons from Cuba. *Am J Epidemiol*. 125(9):800–803.
74. Wang E, Ni H, Xu R, Barrett AD, Watowich SJ, Gubler DJ and Weaver SC (2000) Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J Virol*. 74(7):3227–3234.
75. Watts DM, Burke DS, Harrison BA, Whitmire RE and Nisalak A (1987) Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. *Am J Trop Med Hyg*. 36(1):143–152.

76. Whitehead SS, Blaney JE, Durbin AP and Murphy BR (2007) Prospects for a dengue virus vaccine. *Nat Rev Microbiol*. 5:518–528.
77. WHO Technical Advisory Committee (TAC) on DHF for SE Asia and Western Pacific regions (1975) Guides for Diagnosis, Treatment, Surveillance, Prevention and Control of Dengue Haemorrhagic Fever. Technical Report Series. World Health Organization, Geneva.
78. World Health Organization (1997) Dengue haemorrhagic fever: diagnosis, treatment, prevention and control. 2nd ed. Geneva: World Health Organization.
79. World Health Organization (2009) Dengue and dengue haemorrhagic fever. (http://www.who.int/mediacentre/factsheets/fs117/en/, accessed 1 May 2009).
80. Zanotto PM, Gould EA, Gao GF, Harvey PH and Holmes EC (1996) Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci USA.* 93:548–553.

2. Development of a Dengue Virus Genome Database

Swee Hoe Ong[1,2]

1. Novartis Institute for Tropical Diseases, 10 Biopolis Road, Chromos #05-01, Singapore 138670.
2. Swiss Tropical Institute, Socinstrasse 57, P.O. Box CH-4002 Basel, Switzerland.

Working paper.

2.1 Abstract

In order to study the distribution, molecular epidemiology, and the phenotype/genotype relationships of the dengue virus (DENV), 230 complete DENV genomes from the Southeast Asia region have been sequenced since 2005. Along with DENV genome sequence data collected from public domain sources, all sequences have been meticulously annotated and stored in a dengue genome database. An informatics platform in the form of a Web application has been constructed around the database to provide easy sequence retrieval and as a common platform for sequence analyses. There was a total of 1523 DENV genomes in this database as of January 1, 2009.

2.2 Introduction

The dengue virus (DENV), the causative agent of dengue fever (DF) and the potentially fatal dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS), is a group of four closely-related but antigenically distinct viruses from the genus *Flavivirus* denoted as DENV-1, DENV-2, DENV-3 and DENV-4. The genomic material of the DENV is contained in a positive-sense, single-stranded RNA molecule approximately 10.7 kb in length. It contains a single open reading frame that encodes a polyprotein of between 3387 and 3392 amino acids. The first complete genome sequences for all four serotypes of DENV have been reported since the late 1980s (Zhao *et al*., 1986; Hahn *et al*., 1988; Osatomi and Sumiyoshi, 1990; Fu *et al*., 1992).

Recent advances in sequencing technology and a coincidental drop in sequencing cost have resulted in a well-documented exponential growth in sequence data. A lot about an organism can be learned through comparative genomics with the availability of sequence data (Hardison, 2003). The field of dengue research has also benefited from this trend–there has been a huge increase in the number of DENV genomes published on GenBank since 2005. This growth in sequence data has created not only a need for models of data storage, but also data retrieval systems that are user-friendly, and provide the data in a format that is amenable to further computational analyses.

The DengueInfo Web portal and database (Schreiber *et al*., 2007), Flavitrack (Misra and Schein, 2007) and the NCBI's Virus Variation Resources (Resch *et al*., 2009) represent attempts to fulfill some of these needs. Nevertheless, the frequency of update, choice of annotation data, accuracy of annotation data, data retrieval options and the format of the retrieved data understandably relied entirely on the resource providers. Moreover, inclusion of pre-released sequences generated from in-house sequencing efforts and the addition of new annotation fields according to personal and evolving needs when using these external informatics resources is impractical, further underlining the need for a bioinformatics application on which creative freedom can be fully exercised.

In this study, a Web application was built on top of a database that stores manually-curated annotation and DENV genome sequence information generated from in-house sequencing efforts and from public-domain sources. The Web application was then extended to include a collection of independently-acting but compatible sequence analysis tools to provide a common bioinformatics platform for easy sequence retrieval and analysis of DENV genomic data.

2.3 Material and Methods

2.3.1 Complete genome sequencing of DENV

The two-step strategy and experimental protocol for doing complete genome sequencing of dengue virus have been described in detail elsewhere (Ong *et al*., 2008; Schreiber *et al*., 2009). Briefly, cDNA templates were generated from viral RNA with priming primers using the SuperScript III reverse transcriptase (Invitrogen). The cDNA template was then amplified using *Pfu* Turbo DNA polymerase (Stratagene) in five separate PCR reactions to generate five slightly overlapping fragments (Figure 2.1). Optimum PCR conditions can be variable for different batches of virus samples depending on their source, and the parameters would have to be optimised anew.

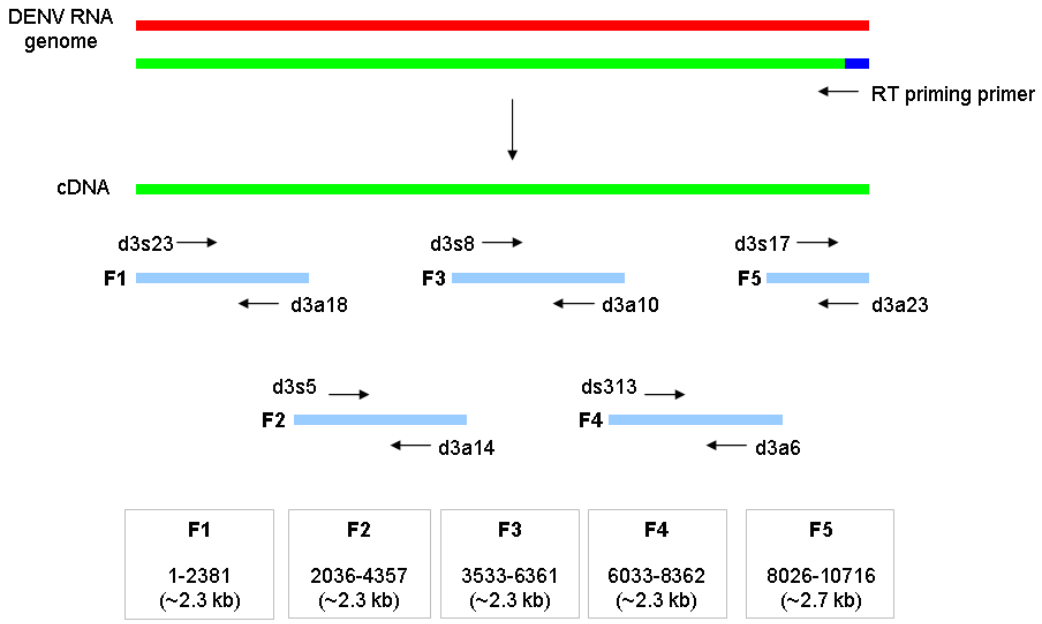


Figure 2.1. Schematic representation of the strategy employed in sequencing the complete genomes of DENV. The example depicted is for DENV-3.

The PCR products were quantified by spectrophotometry (NanoDrop ND-1000 Spectrophotometer) and sent for automated capillary sequencing using 50 serotype-specific sequencing primers. Sequencing was done with a 3730xl DNA Analyzer (Applied Biosystems) using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems).

The chromatograms from capillary sequencing were assembled using SeqScape version 2.5 (Applied Biosystems) to produce a *consensus sequence* of the virus sample. The consensus sequences were first aligned to the sequences of other known genomes of the same serotype. These consensus sequences then went through a round of manual editing to ensure its accuracy and to weed out artifacts from the base-calling algorithm of the automated sequencer. Whenever possible, manual editing was carried out independently by two individuals in a “buddy system”, and the two sets of results are then compared and merged to produce the final version of the genome sequence.

2.3.2 Inclusion criteria of public domain DENV sequences

Complete and near-complete DENV genomes in the public domain were retrieved by searching the NCBI Nucleotide database - GenBank (http://www.ncbi.nlm.nih.gov/) using the query string: *txid11052[orgn] AND 10000[SLEN]:13000[SLEN]*. The number 11052 is the taxonomy ID for the Dengue Virus Group as defined by the NCBI Taxonomy Browser; the balance of the query string specified that sequences with length between 10,000 and 13,000 nucleotides be retrieved. Typical DENV genomes are around 10,700 nucleotides so the query string will retrieve near-complete genome sequences while leaving out almost all of the longer-than-usual sequences of engineered constructs and whole-virus replicon vectors.

2.3.3 DENV genome sequence annotation

Primary annotation of the genome sequences retrieved from GenBank was achieved by parsing the associated GenBank records. The only information that was readily available were the accession number and serotype, whereas other information such as strain name, severity (DF, DHF, DSS), country of isolation and year of isolation is not uniformly recorded and often required manual curation and correction. The missing information is filled by following the “paper trail” of cited scientific literature, searching using Internet search engines or through personal communication.

A second round of annotation assigns information for fields such as the nature of virus (either as Wild Type, Engineered or Patent), category of virus based on sequencing project (EDEN, NEHCRI, Universiti Malaya, sylvatic *et cetera*), subtype classification (based on phylogenetic trees, see section 2.3.4) and other useful miscellaneous remarks. Coordinates for the boundaries of the viral genes were calculated from a massive multiple sequence alignment of the 4 serotypes of DENV to allow extraction of sequence fragment of interest. A detailed description of the fields can be found in Table 2.1.

Table 2.1. The description of the fields used in the annotations of the dengue genome sequences.

Fields	Sample Data	Sources of Annotation Data
Accession	EU482478	GenBank.
Serotype	DENV-1 to DENV-4	GenBank or based on phylogenetic trees.
Strain name	NGC, MON501 etc.	GenBank, literature, or assigned for in-house strains.
Nature	Wild Type, Engineered, Patent	Assigned based on literature.
Severity	DF, DHF	GenBank, literature, or personal communication.
Country of isolation	Singapore, India etc.	GenBank, literature, or personal communication.
Year of Isolation	1946, 2005 etc.	GenBank, literature, or personal communication.
Collection	EDEN, Makassar etc.	GenBank, literature, or personal communication.
Subtype	Cosmopolitan, Asian/American etc.	Assigned based on phylogenetic trees.
First seen on GenBank	8/2/1993 etc.	GenBank.
Remarks	"From Hainan province" etc.	GenBank, literature, personal communication, or assigned.

2.3.4 DENV subtype classification

Phylogenetic trees for each of the four DENV serotypes were generated from aligned complete genome sequences using the Minimum Evolution tree inference method implemented in MEGA version 4.0 (Tamura *et al*., 2007). The substitution model used was the Tamura-Nei model and the obtained tree topology was tested using 1000 bootstrap replications. The subtype classification scheme follows that of Goncalvez *et al*. (2002) for DENV-1, Twiddy *et al*. (2002) for DENV-2, Lanciotti *et al*. (1994) for DENV-3 and Lanciotti *et al*. (1997) for DENV-4.

2.3.5 Technical specifications of DENV genome database

Nucleotide sequences were retrieved from public domain sources and generated from sequencing efforts, and their associated annotation data were stored in a MySQL relational database management system. The data retrieval function was achieved through the implementation of a MySQL/PHP Web application. The Web application was hosted on Microsoft Internet Information Server (IIS) version 5.1. The database was implemented on MySQL Server version 5.0.67 whereas the scripting engine used was PHP version 5.2.6.

Various analysis tools, totaling 30 at the time of writing, were implemented in the form of PERL scripts that are accessed via a Web browser. The PERL scripting engine used was ActivePerl version 5.8.8.817 from ActiveState Software. The analysis results are also provided via the browser graphical user interface (GUI), thereby guaranteeing ease of use. The GUI design uses a standardised CSS-driven menu for easy access to the various sections of the Web application and the analysis tools. The application uses client-side JavaScripts, DHTML and AJAX snippets, where appropriate, to enhance its usability and aesthetic value. A screenshot of the front page of the application is shown in Figure 2.4.

The application has dependencies on a number of PERL modules namely: GD (version 2.5), DBD-mysql (version 4.008), CGI (version 3.41), DBI (version 1.607), Math::Trig (version 1.18) and Tie-IxHash (version 1.21). The SQL file of the MySQL database generated by the *mysqldump* tool (visualised in graphical form in Figure 2.2), can be found in Appendix A.

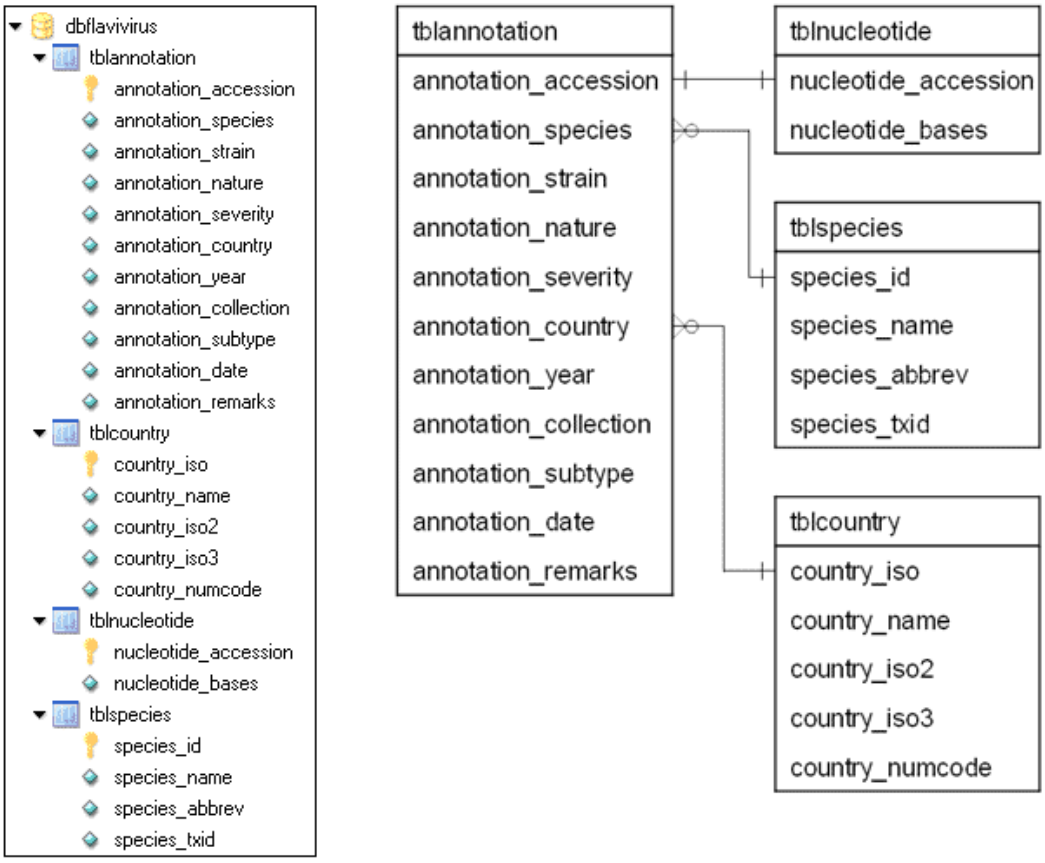


Figure 2.2. Graphical representation of the database structure as visualised by MySQL AB's Query Browser version 1.2.13 (left) and its entity-relationship diagram (right).

2.4 Results and Discussion

2.4.1 Complete genome sequencing of DENV

As of January 1, 2009 a total of 230 complete DENV genomes have been generated since the middle of 2004 through the joint efforts of the Novartis Institute for Tropical Diseases (NITD), the Genome Institute of Singapore (GIS), and the NITD-Eijkman Institute-Hasanuddin University Clinical Research Initiative (NEHCRI). The sequenced viruses can be divided into 5 groups based on the time and location the viruses were sampled (Table 2.2). The first four groups of viruses were selected for sequencing to address different questions and these sequences were treated as standalone data sets and analysed independently (Ong *et al*., 2008; Schreiber *et al*., 2009).

The group labeled as Miscellaneous contains laboratory strains, prototype strains and *ad hoc* sequencing of a small number of interesting viral strains.

Table 2.2. The number of completed DENV genomes generated through sequencing efforts by the end of 2008, grouped by the time and location of virus sampling.

Place of collection	Year of collection	Sequenced genomes
Jakarta, Indonesia	1985-2003	16
Kuala Lumpur, Malaysia	1995-2005	46
Singapore	2005-2006	106
Makassar, Indonesia	2006-2007	54
Miscellaneous	Not applicable	8
	TOTAL	**230**

The 5-fragment RT-PCR genome sequencing protocol has undergone various rounds of optimization since 2004, in particular the continued refinement of the primer sequences used for PCR amplification and capillary sequencing. As more sequences become available, primers can now be designed to cater to virus isolates from a particular subtype or from a particular clade within a subtype. Still, protocol optimization often has to be started anew for each set of “virus collections”, and this has proved to be the main bottleneck in the entire genome sequencing effort. New sequencing strategies that employ “next-generation” sequencing technology should be explored to speed up future DENV genome sequencing efforts.

2.4.2 DENV genome sequence annotation

The primary source of the annotation for the publicly-available DENV genome sequences was their respective accompanying GenBank record. Automated annotation could only be done partially because much of the desired metadata was not uniformly recorded. In the examples highlighted in Figure 2.3, a sequence submitter could use either “strain” or “isolate” to refer to the strain name of a virus isolate, or the information is simply absent. Many

GenBank records, like in the third example, are devoid of any useful information.

```
FEATURES             Location/Qualifiers
     source          1..10690
                     /organism="Dengue virus 1"
                     /mol_type="genomic RNA"
                     /isolate="BID V793"
                     /isolation_source="cell supernatant"
                     /host="Homo sapiens"
                     /db_xref="taxon:11053"
                     /country="Viet Nam: south"
                     /collection_date="2006"
                     /note="passage history: C6/36 4; cohort
                     population: DENCO type: 1"
```

```
FEATURES             Location/Qualifiers
     source          1..10707
                     /organism="Dengue virus 3"
                     /mol_type="genomic RNA"
                     /strain="D3BR/RP1/2003"
                     /isolation_source="dengue fever patient"
                     /host="Homo sapiens"
                     /db_xref="taxon:11069"
                     /country="Brazil"
                     /note="genotype: III"
```

```
FEATURES             Location/Qualifiers
     source          1..10723
                     /organism="Dengue virus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:12637"
```

Figure 2.3. Examples of GenBank records from which primary annotation for publicly-available DENV genomes were extracted. Red boxes highlight the lack of convention in field names; entries underlined in blue show the use of non-uniform country names; entry underlined in green shows information that is useful but seldom provided. The third example is devoid of any useful annotation data.

The solution to fill the numerous gaps was to manually annotate information not extracted by the initial round of automated annotation. Fields that are useful in subsequent analyses are added in addition to those that can be readily found in GenBank records. Subtype information, for example, can be used to help design primers for specific subtypes of viruses and to follow the microevolution of viruses. Knowing the nature of viruses (whether Wild type, Engineered or Patent) would be useful when deciding which genomes to

include when doing specific kinds of sequence analysis. The “Patent” sequences were invariably for engineered or attenuated isolates used in vaccine development efforts. Detailed description of the fields used for genome annotation is provided in Table 2.1.

2.4.3 DENV genome database Web application

The Web application for the dengue genome database was named the Flavivirus Toolkit (FVT), as can be seen in the screenshot of the application’s front page (Figure 2.4). The name alludes to the initial objective of including in the database genome sequences of all flaviviruses. It is well known that many flaviviruses are important human pathogens. In addition, the flaviviruses all share the same genomic organization and their genomes are of more or less similar size. As such, they can utilise similar schemes in annotation. These factors made housing all flavivirus genome sequences together in a single repository scientifically useful and technically feasible. Currently the FVT only contains the genome sequences of DENV, but the infrastructure is already in place to integrate sequences of other flaviviruses in the future.

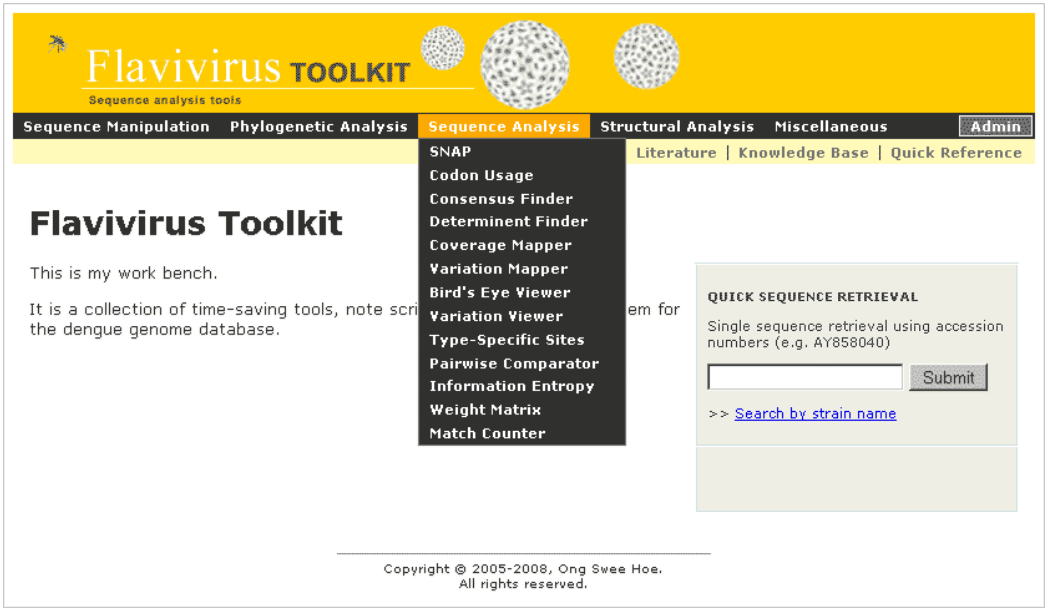


Figure 2.4. A screenshot of the front page of the browser-based user interface to the dengue genome database Web application (nicknamed the Flavivirus Toolkit). The CSS-driven dropdown menu allows easy access to the available analysis tools.

From the user's perspective the Web application can be functionally divided into three main parts: sequence retrieval, sequence analysis tools, and informative. The 3-step sequence retrieval function is most fundamental to the database and it is comprised of two tools that work one after another: the Sequence Selector and Sequence Extractor. The Sequence Selector tool allows the user to specify the selection criteria of sequences to be retrieved. The range of available selection criteria (which resembles the fields of annotation in Table 2.1) is shown in Figure 2.5, and the output of the Sequence Selector is shown in Figure 2.6.

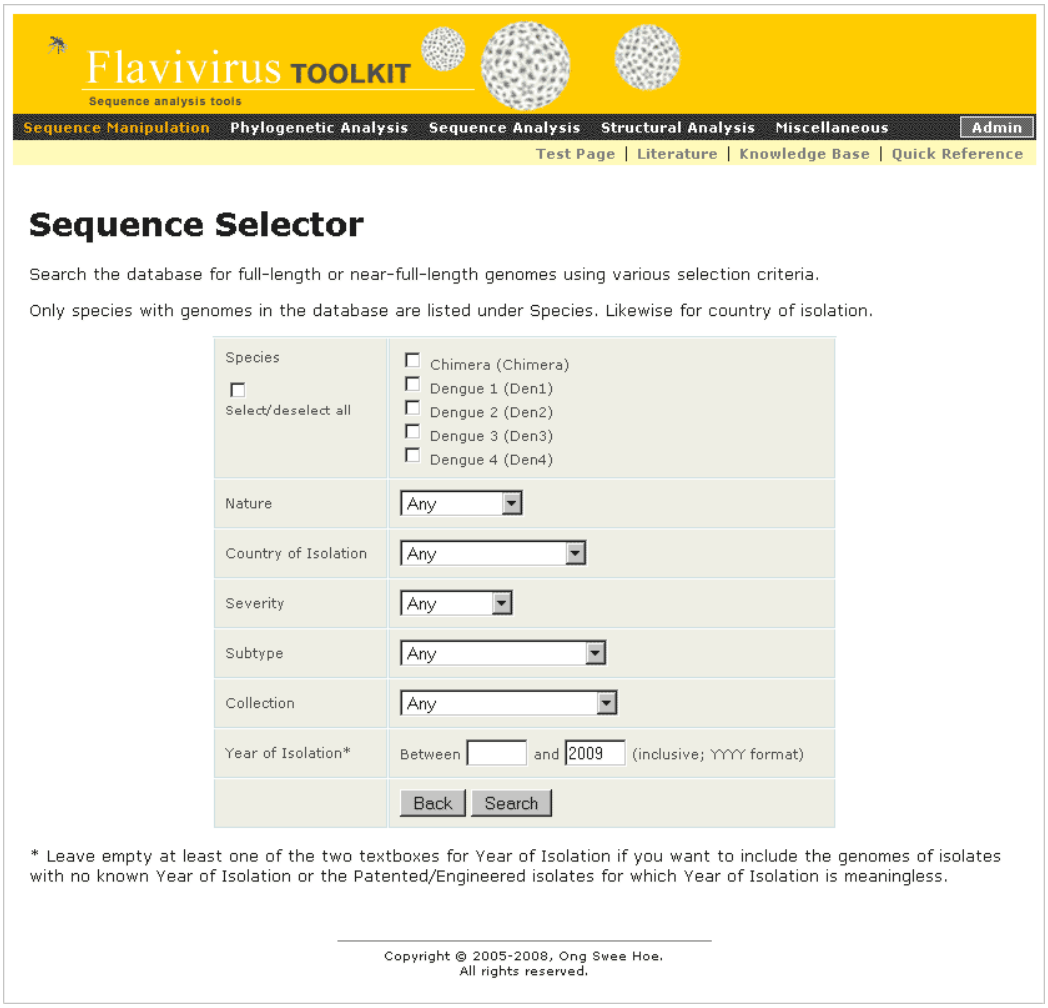


Figure 2.5. Step 1 of 3 to retrieve sequences stored in the dengue genome database. The Sequence Selector allows the user to specify the selection criteria of sequences to be retrieved.

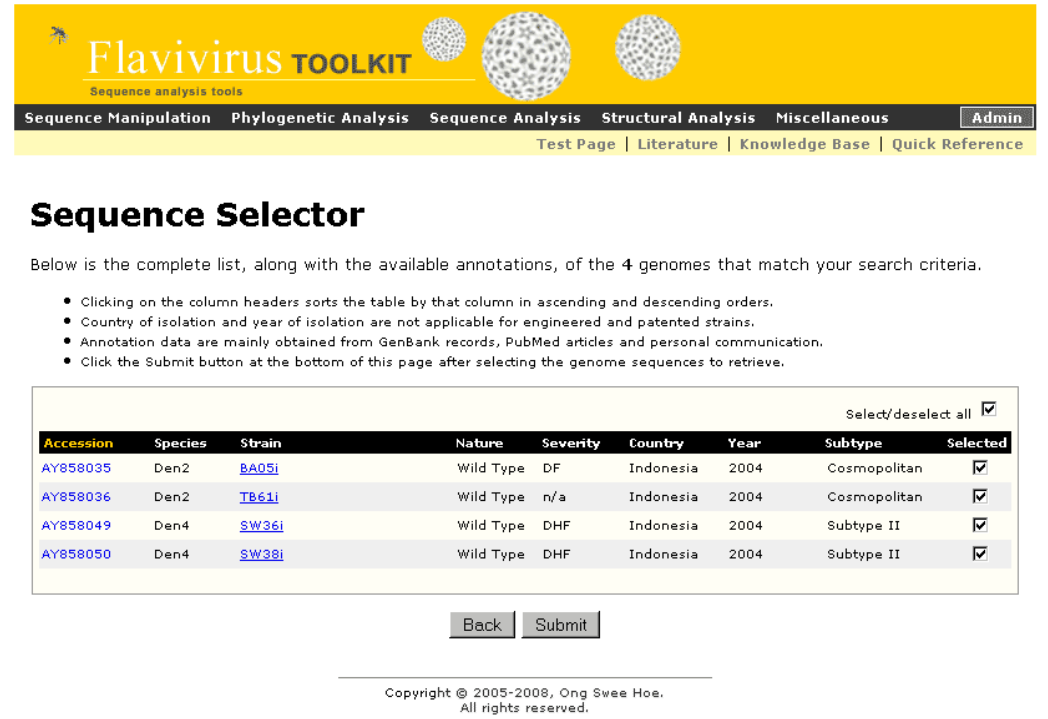


Figure 2.6. Step 2 of 3 to retrieve sequences stored in the dengue genome database. The Sequence Selector produces a list of sequences that match the selection criteria, along with some pertinent information of the matched sequences. The user can remove sequences from the list before proceeding to the next step.

The user is brought to the Sequence Extractor tool once the selection list is finalised by clicking the “Submit” button (Figure 2.6). DENV isolates that matched the selection criteria are displayed in the Sequence Extractor which is a tool that allows the user to control which part of the genome is to be retrieved. Retrieved sequences can be sorted by multiple parameters befitting the kind of analysis to be done subsequently. For example, DENV sequences from a particular subtype found in a country can be ordered by Year (of isolation) so that a chronological view of the mutations that have accrued over years of transmission can be examined.

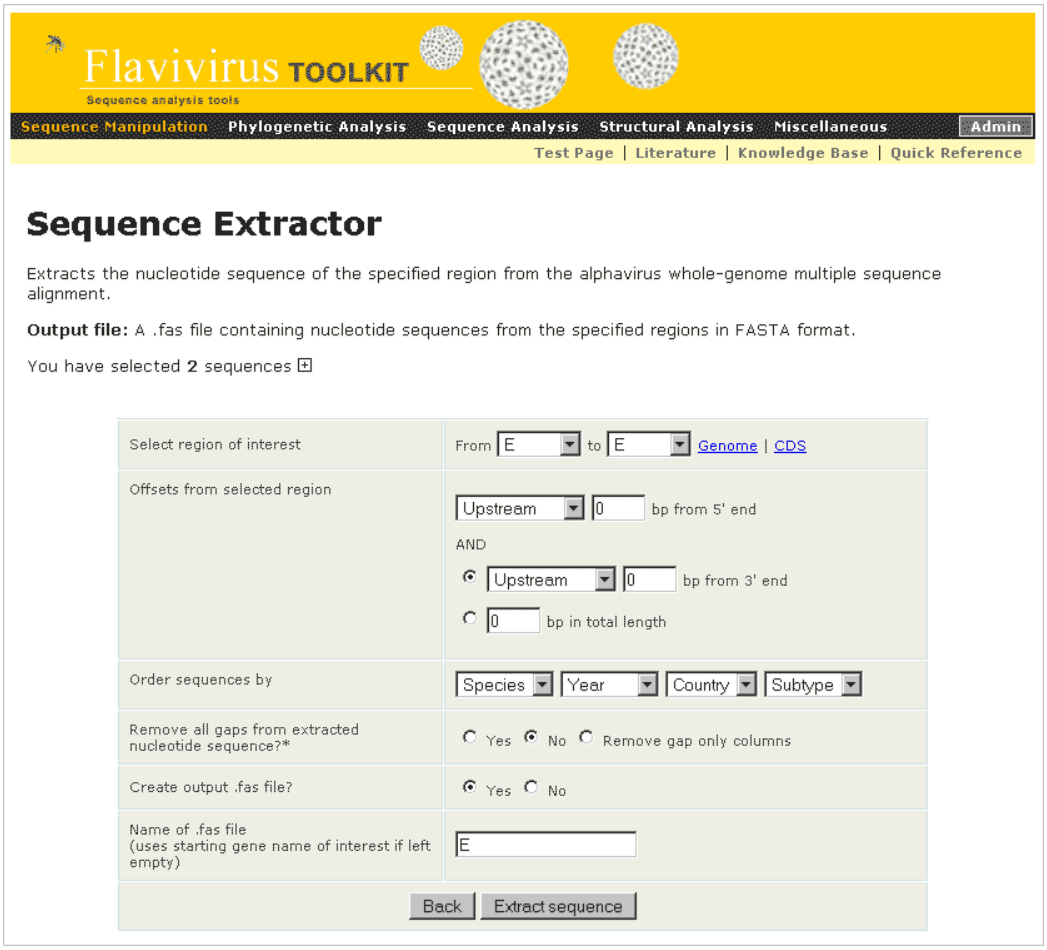


Figure 2.7. Step 3 of 3 to retrieve sequences stored in the dengue genome database. The Sequence Extractor allows the user to decide which part of the selected genomes to extract, and how those sequences should be ordered in the output file.

The output of the three sequence retrieval steps is a single FASTA format text file containing the selected (complete or partial) genome sequences ordered in the defined manner. This file can then be used externally with other software, or with the analysis tools that form the second functional part of the Flavivirus Toolkit. A complete description of the various tools here would be too lengthy, but suffice to say each of the tools are designed for specific tasks and they can be used serially to perform various kinds of analysis. A few examples of the kinds of analyses that can be performed with the tools are outlined in Figure 2.8.

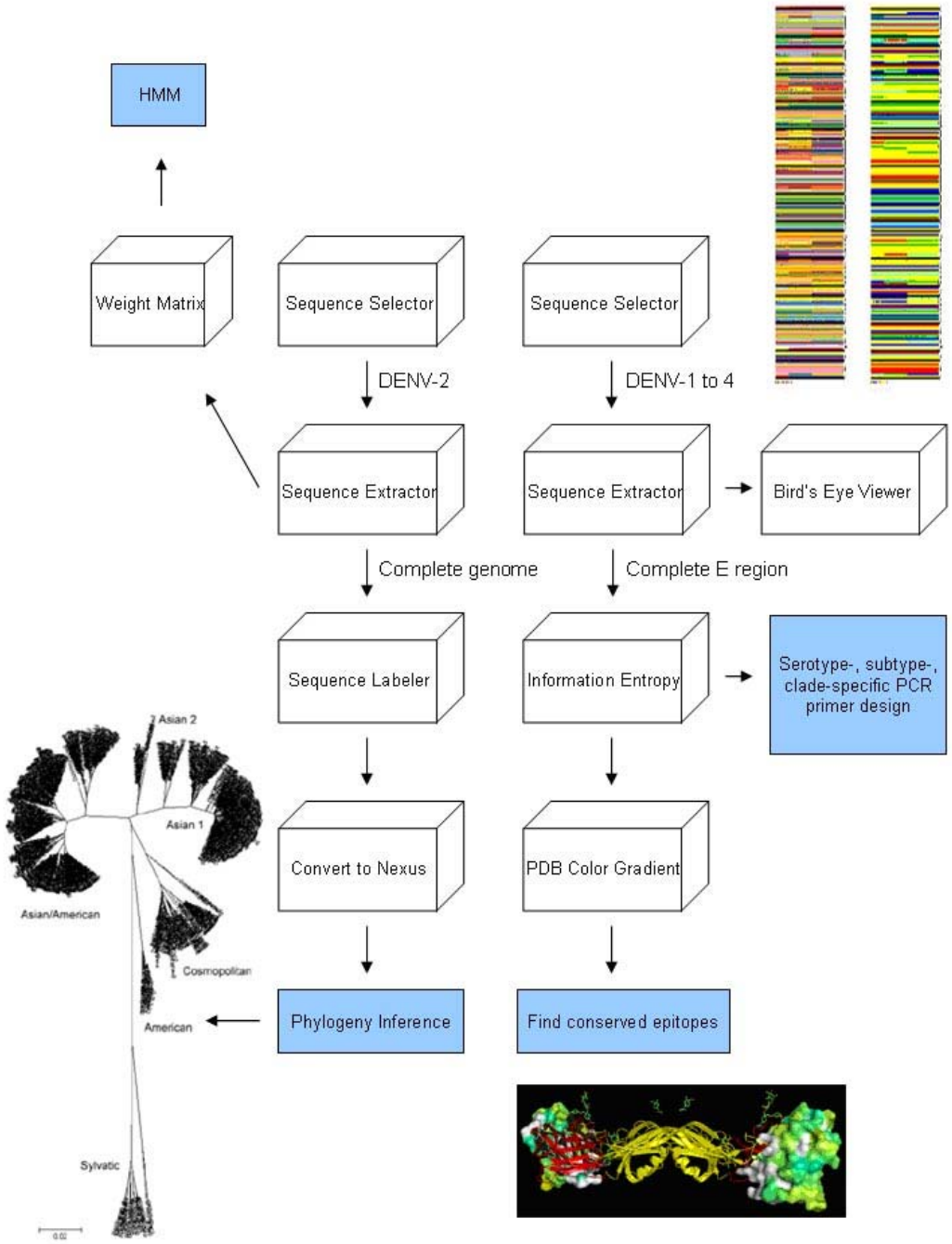


Figure 2.8. Schematic diagram showing a few examples of sequence-based analyses that can be done by combining the various sequence analysis tools available in the Flavivirus Toolkit.

The third and final part of the Flavivirus Toolkit consists of a series of Web pages that provide various kinds of information on the state of the database. These are either displayed as charts and tables for statistics, maps and schematic diagrams, or in the form of phylogenetic trees. Web pages for administrative and maintenance tasks will need to be added prior to releasing the FVT for public use.

2.4.4 State of the DENV genome database

Currently there are 1523 DENV genomes in the database consisting of 616 DENV-2 genomes (41%), 583 DENV-1 genomes (38%), 274 DENV-3 genomes (18%) and 48 DENV-4 genomes (3%). One thing that should be pointed out here is that the numbers do not reflect the frequency the serotypes are encountered in the wild, but are merely the results of the choices made by the various genome sequencing projects. There are also two engineered (and patented) chimeric genomes that are half DENV-1 and half DENV-2 (Figure 2.9).

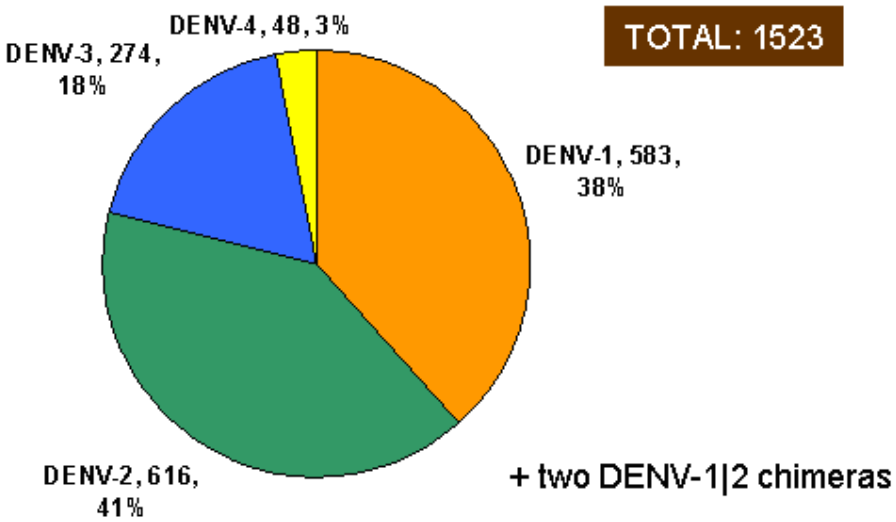


Figure 2.9. Breakdown of dengue genome sequences in the database by serotype as of January 1, 2009.

Of the 1523 genomes, 1394 genomes (91.53%) are available publicly on GenBank as of January 1, 2009, with 122 genomes of these (8.75%) being the result of the genome sequencing effort by NITD and its collaborators. Another 129 genomes are in the process of being completed and will be released soon together with a publication (Figure 2.10).

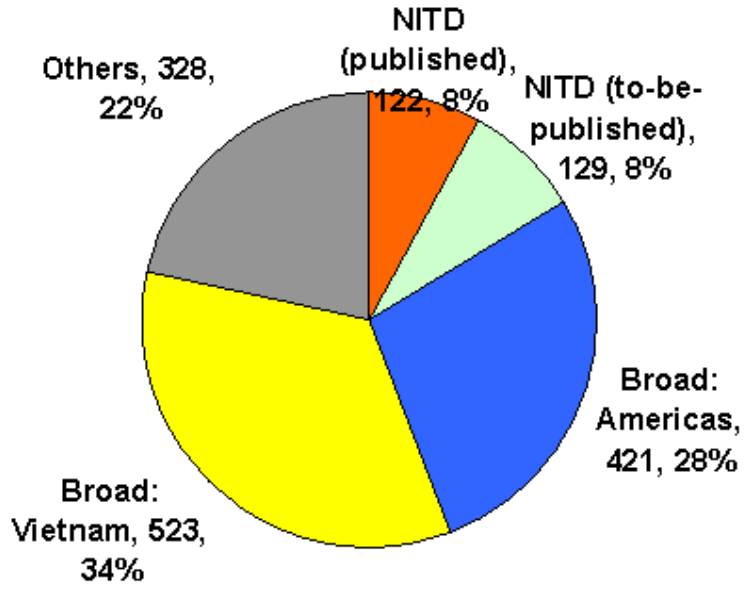


Figure 2.10. Breakdown of dengue genome sequences in the database by main contributing institutions as of January 1, 2009. Contribution by NITD and collaborators are divided into two segments: Published and To-be-published.

Apart from the NITD, the Broad Institute in Massachusetts, US is the other big contributor of DENV genomic data. The project Web site can be access at http://www.broad.mit.edu/annotation/viral/Dengue/ProjectInfo.html. Starting from March 2008 and as of January 1, 2009 the Broad Institute have released 523 genomes of dengue isolates from Vietnam and 421 genomes of isolates from the Americas. This has resulted in a massive spike in the number of publicly available dengue genomes in GenBank in the year 2008 (Figure 2.11). New kinds of analysis have been made possible with the increased amount of sequence data and it is hoped this trend will continue for the foreseeable future.

Year	Available Genomes
1993	8
1994	8
1995	8
1996	8
1997	15
1998	23
1999	38
2000	56
2001	70
2002	83
2003	88
2004	113
2005	201
2006	241
2007	405
2008	1394

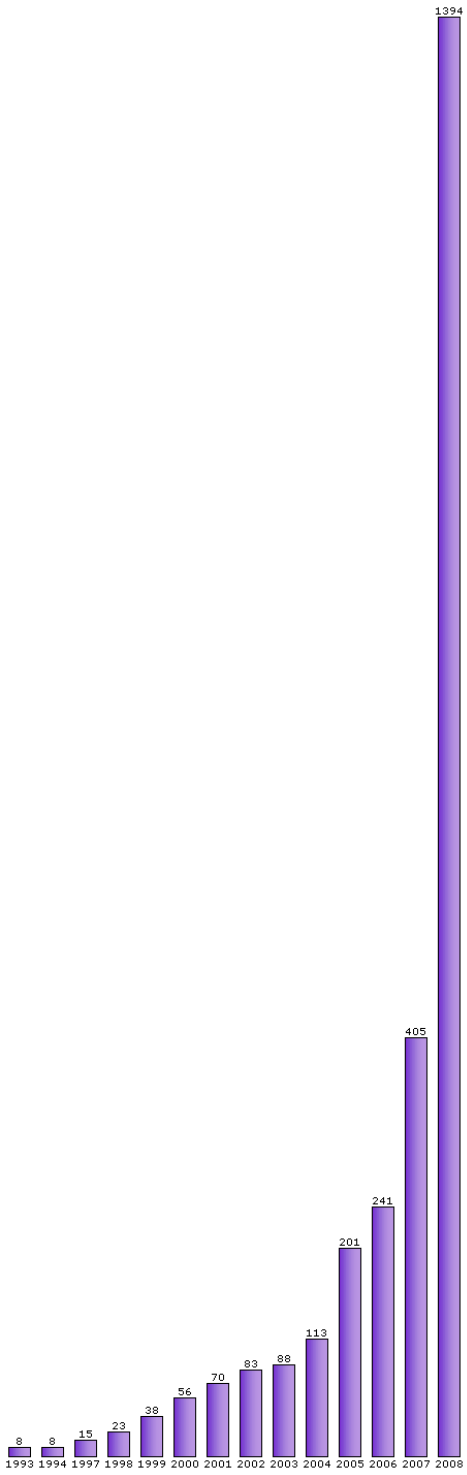


Figure 2.11. Bar chart and table showing the total number of publicly available dengue genome sequences since 1993. NITD started releasing DENV genome sequences in 2005 which accounted for the first non-linear increase in the chart. Contributions from the Broad Institute in 2008 resulted in a huge spike in the number of dengue genome sequences.

2.5 Conclusion

The DENV genome sequencing activities by NITD have contributed a large amount of valuable data to the research community since 2005. Prior to that, public repositories of genetic information for dengue virus were heavily skewed toward structural genes, in particular the E (envelope). In order to understand the molecular epidemiology of the DENV and to identify molecular markers of pathogenecity, among other things, it is crucial to obtain the complete genetic blueprint of the infectious agent.

By combining the genome sequencing and annotation efforts we have developed a valuable bioinformatics resource of dengue genome sequences that far exceeds the usefulness of a general-purpose sequence repository such as the GenBank. The dengue genome database with its related analysis tools have made analysis and comparison of viral genomes manageable and can serve as a framework for data sharing and analysis in the future. However, as the quantity and complexity of sequence data continue to grow, so will the challenge of organizing and maintaining databases such as this.

In the following chapters, whole-genome comparative sequence analyses of DENV strains isolated from three countries and at different time points were used to answer pertinent molecular epidemiological questions about the diversity, distribution, and molecular evolution of DENV serotypes and subtypes in Southeast Asian countries - the region where DHF was first identified more than five decades ago.

2.6 Acknowledgements

The contributions by several co-workers and external collaborators to the DENV genome sequencing effort is hereby acknowledged: Martin L. Hibberd, Khoo Chen Ai and Pauline Aw from the A*STAR Genome Institute of Singapore (GIS); Tedjo Sasmono from NEHCRI; Andy Yip Jin Teng, Liu Wei, Chen Yen Liang and Mark J. Schreiber from NITD; Lukas Tanner and Jasmin Schmid – attachment students at NITD; and Subhash G. Vasudevan, former Unit Head of the Dengue Unit, NITD.

2.7 Author Contributions

SH Ong assembled the sequences, annotated the DENV genome sequences, designed and coded the database and Web application. Genome sequencing involved contributions from a number of people as shown in the chart below. Wei Liu propagated all virus samples prior to sequencing.

Data sets	RT-PCR	Sequencing	Assembly
Jakarta, Indonesia	SH Ong Andy Yip	A*STAR GIS	SH Ong
Kuala Lumpur, Malaysia	SH Ong Jasmin Schmid	A*STAR GIS	SH Ong Pauline Aw
Singapore	Lukas Tanner	A*STAR GIS	SH Ong Pauline Aw
Makassar, Indonesia	NEHCRI	NEHCRI	SH Ong
Miscellaneous	Miscellaneous	Miscellaneous	SH Ong

2.8 References

1. Fu JL, Tan BH, Yap EH, Chan YC and Tan YH (1992) Full-length cDNA sequence of dengue type 1 virus (Singapore strain S275/90). *Virology*. 188:953–958.
2. Goncalvez AP, Escalante AA, Pujol FH, Ludert JE, Tovar D, Salas RA and Liprandi F (2002) Diversity and evolution of the envelope gene of dengue virus type 1. *Virology*. 303:110–119.
3. Hardison RC (2003) Comparative genomics. *PLoS Bio*. 1(2):156–160.
4. Hahn YS, Galler R, Hunkapiller T, Dalrymple JM, Strauss JH and Strauss EG (1988) Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. *Virology*. 162:167–180.
5. Lanciotti RS, Lewis JG, Gubler DJ and Trent DW (1994) Molecular evolution and epidemiology of dengue-3 viruses. *J Gen Virol*. 75:65–75.

6. Lanciotti RS, Gubler DJ and Trent DW (1997) Molecular evolution and phylogeny of dengue-4 viruses. *J Gen Virol*. 78:2279–2286.
7. Misra M and Schein CH (2007) Flavitrack: an annotated database of flavivirus sequences. *Bioinformatics*. 23(19):2645–2647.
8. Ong SH, Yip JT, Chen YL, Liu W, Harun S, Lystiyaningsih E, Heriyanto B, Beckett CG, Mitchell WP, Hibberd ML, Suwandono A, Vasudevan SG, and Schreiber MJ (2008) Periodic re-emergence of endemic strains with strong epidemic potential-a proposed explanation for the 2004 Indonesian dengue epidemic. *Infect Genet Evol*. 8:191–204.
9. Osatomi K and Sumiyoshi H (1990) Complete nucleotide sequence of dengue type 3 virus genome RNA. *Virology*. 176:643–647.
10. Resch W, Zaslavsky L, Kiryutin B, Rozanov M, Bao Y and Tatusova TA (2009) Virus variation resources at the National Center for Biotechnology Information: dengue virus. *BMC Microbiol*. 9(1):65.
11. Schreiber MJ, Ong SH, Holland RCG, Hibberd ML, Vasudevan SG, Mitchell WP and Holmes EC (2007) DengueInfo: A web portal to dengue information resources. *Infect Genet Evol.* 7(4):540–541.
12. Schreiber MJ, Holmes EC, Ong SH, Soh HS, Liu W, Tanner L, Aw PP, Tan HC, Ng LC, Leo YS, Low JG, Ong A, Ooi EE, Vasudevan SG and Hibberd ML (2009) Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J Virol.* 83(9):4163–4173.
13. Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol and Evol*. 24:1596–1599.
14. Twiddy SS, Farrar JJ, Nguyen VC, Wills B, Gould, EA, Gritsun T, Lloyd G and Holmes EC (2002) Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology*. 298:63–72.
15. Zhao B, Mackow E, Buckler-White A, Markoff L, Chanock RM, Lai CJ and Makino Y (1986) Cloning full-length dengue type 4 viral DNA sequences: analysis of genes coding for structural proteins. *Virology*. 155:7–88

3. Periodic re-emergence of endemic strains with strong epidemic potential—A proposed explanation for the 2004 Indonesian dengue epidemic

Swee Hoe Ong[1,4], Jin Teen Yip[1], Yen Liang Chen[1], Wei Liu[1], Syahrial Harun[2], Erlin Lystiyaningsih[2,3], Bambang Heriyanto[2], Charmagne G. Beckett[3], Wayne P. Mitchell[4], Martin L. Hibberd[4], Agus Suwandono[2], Subhash G. Vasudevan[1], Mark J. Schreiber[1]

1. Novartis Institute for Tropical Diseases Pte. Ltd., 10 Biopolis Road, #05-01 Chromos, Singapore 138670.
2. Center for Research and Program Development on Disease Control, Ministry of Health, Jl. Percetakan Negara No. 29, Jakarta Pusat 10560, Indonesia.
3. US Naval Medical Research Unit No. 2 (US NAMRU-2), Kompleks P2M-PLP/LITBANGKES, Jl. Percetakan Negara No. 29, Jakarta Pusat 10560, Indonesia.
4. Genome Institute of Singapore, 60 Biopolis Street, #02-01 Genome, Singapore 138672.

This article has been published in:

Infection, Genetics and Evolution (2008) 8: 191–204

3.1 Abstract

Indonesia experienced a severe dengue epidemic in the first quarter of 2004 with 58,301 cases and 658 deaths reported to the WHO. All four dengue virus (DENV) serotypes were detected, with DENV-3 the predominant strain. To ascertain the molecular epidemiology of the DENV associated with the epidemic, complete genomes of 15 isolates were sequenced from patient serum collected in Jakarta during the epidemic, and two historical DENV-3 isolates from previous epidemics in 1988 and 1998 were selectively sequenced for comparative studies. Phylogenetic trees for all four serotypes indicate the viruses are endemic strains that have been circulating in Indonesia for a few decades. Whole-genome phylogeny showed the 2004 DENV-3 isolates share high similarity with those isolated in 1998 during a major epidemic in Sumatra. Together these subtype I DENV-3 strains form a Sumatran-Javan clade with demonstrated epidemic potential. No newly-acquired amino acid mutations were found while comparing genomes from the two epidemics. This suggests re-emergence of little-changed endemic strains as causative agents of the epidemic in 2004. Notably, the molecular evidence rules out change in the viral genomes as the trigger of the epidemic.

3.2 Introduction

Dengue fever is a mosquito-born disease with high public health impact that is estimated to affect nearly 50 million people worldwide each year (Monath, 1994; Gubler and Clark, 1995; WHO, 2002). Dengue fever is endemic throughout most of the tropical areas of the world, coincident with the distribution pattern of its mosquito vectors. Dengue disease can manifest in the form of the mild dengue fever (DF), or the more severe and potentially fatal dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS) which has a fatality rate as high as 10–15% depending on the availability of healthcare (Gubler, 2002). Currently there is no vaccine or therapeutic agent available against dengue fever.

The causative agent of dengue fever, the dengue virus (DENV), is transmitted to humans by infected females of the mosquito vectors *Aedes aegypti* and *Aedes albopictus*. DENV is a single-stranded positive-sense RNA virus of the genus *Flavivirus*. The ~10.7 kb DENV genome encodes three structural (capsid, pre/membrane and envelope) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) in a single open reading frame. DENV is divided into four antigenically-related serotypes denoted as DENV-1, DENV-2, DENV-3 and DENV-4. Each serotype is sufficiently different that infection with one does not provide complete cross protection for the other three. In a scheme first proposed by Rico-Hesse (1990), DENV can be further divided into intra-serotypic categories called interchangeably as subtypes or genotypes based on their nucleotide sequence data. Subtype determination via phylogenetic means are often used to infer the phylogeny and to monitor the spread of virus strains (Chungue *et al*., 1995; Rico-Hesse *et al*., 1997; Messer *et al*., 2003).

South East Asia has been a focal point of dengue activity since 1950s when DHF was first described. Indonesia, the largest country in the region, has experienced periodic outbreaks of dengue since 1968 with increasing numbers of infections and severity (Sumarmo, 1987). Major dengue epidemics occurred in Indonesia in 1998, during which 72,133 cases and 1414 deaths were reported, and again in 2004 with more than 58,301 cases and 658 deaths in the first 4 months of the year (WHO, May 11, 2004). In the 2004 epidemic, running from the start of January until early May, cases were reported in all provinces across the Indonesian archipelago. The densely populated capital city of Jakarta, with a population of over 16 million people, was the hardest hit in terms of reported cases and deaths.

Availability of viral genome sequence data will no doubt contribute to a better understanding of the molecular evolution and epidemiology of DENV, especially in a country with a long history of dengue infections such as Indonesia. To this end, we sequenced the genome of 15 DENV isolates collected from hospitals around Jakarta during the 2004 epidemic. In addition, two samples collected during previous epidemics in Jakarta in 1988 and 1998

were also sequenced to ascertain the ancestry of the DENV associated with the 2004 epidemic. Surprisingly, Indonesia does not seem to experience importation of DENV strains despite its proximity to busy international waterway that is the Malacca Strait, and to Thailand, the other country with a long history of dengue infections. Rather, our results point to periodic re-emergence of endemic strains with a demonstrated epidemic potential as the most likely cause of the 2004 epidemic.

3.3 Material and Methods

3.3.1 Virus sample collection and preparation

Sixty nine patient serum samples were collected at eight hospitals in the Greater Jakarta area during the epidemic. Blood samples were taken from patients showing symptoms of dengue fever, as part of routine surveillance, after administering a consent form designed by the Center for Research and Program Development on Disease Control in Jakarta. The presence of acute dengue infection was confirmed by serological tests at the CDC in Jakarta and the serum samples were stored at -80°C until use. The two historical DENV-3 samples collected in Jakarta from 1988 (den3_88) and 1998 (den3_98) were provided by the US Naval Medical Research Unit No 2 (NAMRU-2).

3.3.2 Virus propagation, RNA extraction and virus typing via RT-PCR

Virus samples were propagated in C6/36 *Aedes albopictus* gut cells and viral RNA was extracted using QIAamp Viral RNA Mini Kit (QIAGEN) according to manufacturer's protocol. Serotype of the isolates was inferred via RT-PCR prior to sequencing. Dengue viral RNA was first reverse transcribed into cDNA using SuperScript III reverse transcriptase (Invitrogen) and random hexamers, followed by amplification using *Taq* DNA polymerase (Roche). Serotyping was performed using a slightly modified version of the multiplex RT-PCR protocol (Lanciotti *et al*., 1992) and uses primers described by Seah *et al*. (1995) that distinguishes the four dengue serotypes by PCR product size.

3.3.3 Primer design

Four consensus sequences, one for each serotype, were derived from alignments of published dengue genomes in GenBank. Known recombinant and artificially-mutated strains were excluded. For DENV-4, partial sequences were included in the alignment as only two complete genomes were publicly available at the time. Forward and reverse primers were designed using Vector NTI Suite 9 (Invitrogen Bioinformatics) to give overlapping sequence traces. Criteria of primers design include low number of degenerate positions, no degeneracy in the final three bases, a T_m between 55°C and 65°C and a GC content of 35–60%. The binding positions for the primers on the consensus sequence are shown in Appendix B and the sequences of the sequencing and amplification primers are shown in Appendix C.

3.3.4 Viral cDNA amplification, sequencing, assembly and annotation

cDNA templates were generated from viral RNA using five serotype-specific priming primers at 2 pmol each with the SuperScript III reverse transcriptase. Due to the linear nature of the viral genome the 5' and 3' extremes of the cDNA are primer sequence. The cDNA template was then amplified using *Pfu* Turbo DNA polymerase (Stratagene) in five separate reactions to generate five slightly overlapping PCR fragments. The PCR fragments were separated on 1% agarose gels, excised, and the DNA purified using the QIAquick® PCR Purification Kit (QIAGEN). The purified PCR fragments, ranging from 2.1 to 2.8 kb, were then sent for automated capillary sequencing using 50 serotype-specific sequencing primers. Sequencing was done with a 3730xl DNA Analyzer (Applied Biosystems) using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). Contigs were assembled using SeqScape version 2.5 (Applied Biosystems) and the assembled genome sequences were aligned to other relevant DENV sequences using ClustalW-MPI (Li, 2003), followed by manual editing of the alignments.

3.3.5 Strain nomenclature

Details of the genomes sequenced in this study are listed in Table 3.1. The first two alphabets of the strain names refer to the identity of the hospitals from which the samples were collected. The eight hospitals were abbreviated as follows: BA for Budi Asih and PH for Persahabatan, two hospitals in the city (kota) of East Jakarta. Similarly, SW (Sumber Waras) is in West Jakarta, KJ (Kodja) and PI (Infectious Disease Sulianti Saroso) in North Jakarta, SC (St Carolus) in Central Jakarta, while both FW (Fatmawati) and TB (Tebet) are in South Jakarta. The genome sequences were deposited in the GenBank database with the accession numbers AY858035–AY858050 and AY858983. Both AY662691, a 2003 DF-causing strain from Singapore, and AY947539, the H241 prototype DENV-4, were previously sequenced in 2003 at NITD. Sequences and annotation of the reported genomes are also available at the DengueInfo database (http://www.dengueinfo.org/) (Schreiber *et al*., 2007).

3.3.6 Phylogenetic analysis of the DENV genomes

All phylogenetic trees were constructed from aligned nucleotide sequence data using the maximum likelihood (ML) method implemented in PAUP* (Swofford, 2002), with the best-fit model of nucleotide substitution (GTR+G) selected by Akaike Information Criterion (AIC) implemented in ModelTest (Posada and Crandall, 1998). Branch topology was verified by generating 1000 bootstraps using TBR branch swapping and the scores on tree nodes represent the number of bootstrap replicates (presented in percentage) supporting each node. The length of the tree branches is proportional to the number of nucleotide changes. All trees are midpoint rooted and strains sequenced in this study are underlined for clarity.

Strain datasets used to construct the phylogenetic trees for the purpose of subtype determination were selected to provide good coverage of the known diversity of the four dengue serotypes. Subsequently, published complete genomes of DENV were collected from the DengueInfo database (Schreiber *et al*., 2007) to build genome trees for each of the serotypes. Information on the year of isolation, country of origin, and clinical outcome of

the strains (DF, DHF or DSS) was collected, when available, from GenBank records or through personal communication with submitters of the records.

3.3.7 Clade-specific mutations of the DENV-3 isolates

Clades which are phylogenetically relevant to the Indonesian strains and contain sufficient amount of genome sequence data were then identified and selected for further comparative sequence analysis. Deduced amino acid sequences for subtype I DENV-3 genomes were divided into two groups: those belonging to the Sumatran-Javan lineage (including the slightly more distantly-related Timor Leste strains) and those that are not. The χ^2 test was used to identify residues that are significantly different between the two groups at the 0.05 significance level. The identified residues were submitted to SIFT (http://blocks.fhcrc.org/sift/SIFT.html), a tool that uses sequence homology to predict the effects of amino acid substitutions on protein function (Ng and Henikoff, 2001). The database used by SIFT was SWISS-PROT 51.3 and TREMBL 34.3, with the median conservation of sequences set at 3.00. Conservative and non-conservative amino acid substitutions were defined according to the BLOSUM62 matrix (Henikoff and Henikoff, 1992) with changes having a positive or neutral value in the matrix considered as conservative whereas those with a negative value considered as non-conservative. Similar analyses for the other DENV serotypes were not pursued due to a lack of closely related genome sequences for comparison (data not shown).

3.3.8 Site-specific selection pressures

Site-specific selection pressures acting on the DENV-3 coding sequences were determined using the HyPhy package, implemented as a Web application at http://www.datamonkey.org, using the single likelihood ancestor counting (SLAC) method (Pond and Frost, 2005a, b) and the General Reversible Model nucleotide substitution model.

Table 3.1. The GenBank Accession, strain name, serotype, year of isolation, sequence length and patient-related information of dengue genomes sequenced in this study. The disease severity for seven cases was not evaluated by the admitting hospitals.

GenBank Accession	Strain Name	Serotype	Year Isolated	Length	Severity	Patient Age	Patient Sex
AY858035	BA05i	DENV-2	2004	10723	DF	2	M
AY858036	TB61i	DENV-2	2004	10723	-	-	-
AY858037	BA51	DENV-3	2004	10707	DF	14	M
AY858038	den3_88	DENV-3	1988	10707	-	-	-
AY858039	den3_98	DENV-3	1998	10707	-	-	-
AY858040	FW01	DENV-3	2004	10706	DF + HM	33	M
AY858041	FW06	DENV-3	2004	10707	-	-	-
AY858042	KJ30i	DENV-3	2004	10707	-	-	-
AY858043	KJ46	DENV-3	2004	10706	DHF-I	18	F
AY858044	KJ71	DENV-3	2004	10707	-	-	-
AY858045	PH86	DENV-3	2004	10707	DF	15	M
AY858046	PI64	DENV-3	2004	10707	DHF-I	31	M
AY858047	TB16	DENV-3	2004	10707	DF	59	M
AY858048	TB55i	DENV-3	2004	10673	-	-	-
AY858049	SW36i	DENV-4	2004	10114	DHF-I	12	F
AY858050	SW38i	DENV-4	2004	10516	DHF-I	30	F
AY858983	SC01	DENV-1	2004	7455	DF	7	M

Key: DF- dengue fever; HM – haemorrhagic manifestations; DHF-I dengue haemorrhage fever grade I according to WHO guidelines, i.e. patients with positive tourniquet test, thrombocytopenia and plasma leakage.

3.4 Results

3.4.1 Genome sequencing of dengue isolates

The genomes of 15 DENV isolates were successfully sequenced out of the 69 serum samples collected in Jakarta, Indonesia during the 2004 epidemic. Ten of the genomes were found to be DENV-3, along with two DENV-2, two DENV-4 and one DENV-1. In the remaining cases, there was insufficient virus in the serum samples to allow viral propagation or RNA extraction. Two historical Indonesian dengue samples collected in 1988 and 1998 (den3_88 and den3_98) were similarly sequenced. The two samples were selected to represent the two other major dengue epidemics reported in Indonesia in 1988 and 1998 (Corwin *et al*., 2001; WHO, May 11, 2004; Suwandono *et al*., 2006). The final length obtained for each genome and other relevant details are listed in Table 3.1. The solitary DENV-1 genome was sequenced apart from the last 2.7 kb which were recalcitrant to repeated amplification attempts. Similarly, approximately 550 bases at the 50 end for SW36i could not be obtained. Due to the small number of published DENV-4 genomes in GenBank, it was not possible to design redundant primers that encompass the variability typically seen within a dengue serotype. For this reason, the two DENV-4 isolates were subsequently sequenced using a primer walking approach.

3.4.2 Phylogeny of the DENV-1, DENV-2 and DENV-4 isolates

Phylogeny of the sequenced DENV-1, DENV-2 and DENV-4 isolates were inferred based on the nucleotide sequences of the complete envelope (E) gene. This choice is due to the availability of a greater diversity of E gene sequences for these serotypes compared to genome sequences. The E gene phylogenetic tree of DENV-1 places the solitary DENV-1 isolate in this study, SC01 (AY858983), in subtype IV as described by Goncalvez *et al*. (2002) (Figure 3.1). This group primarily contains isolates from Australia and the West Pacific islands (Indonesia and the Philippines included) – a result that suggests SC01 is an endemic strain in Indonesia.

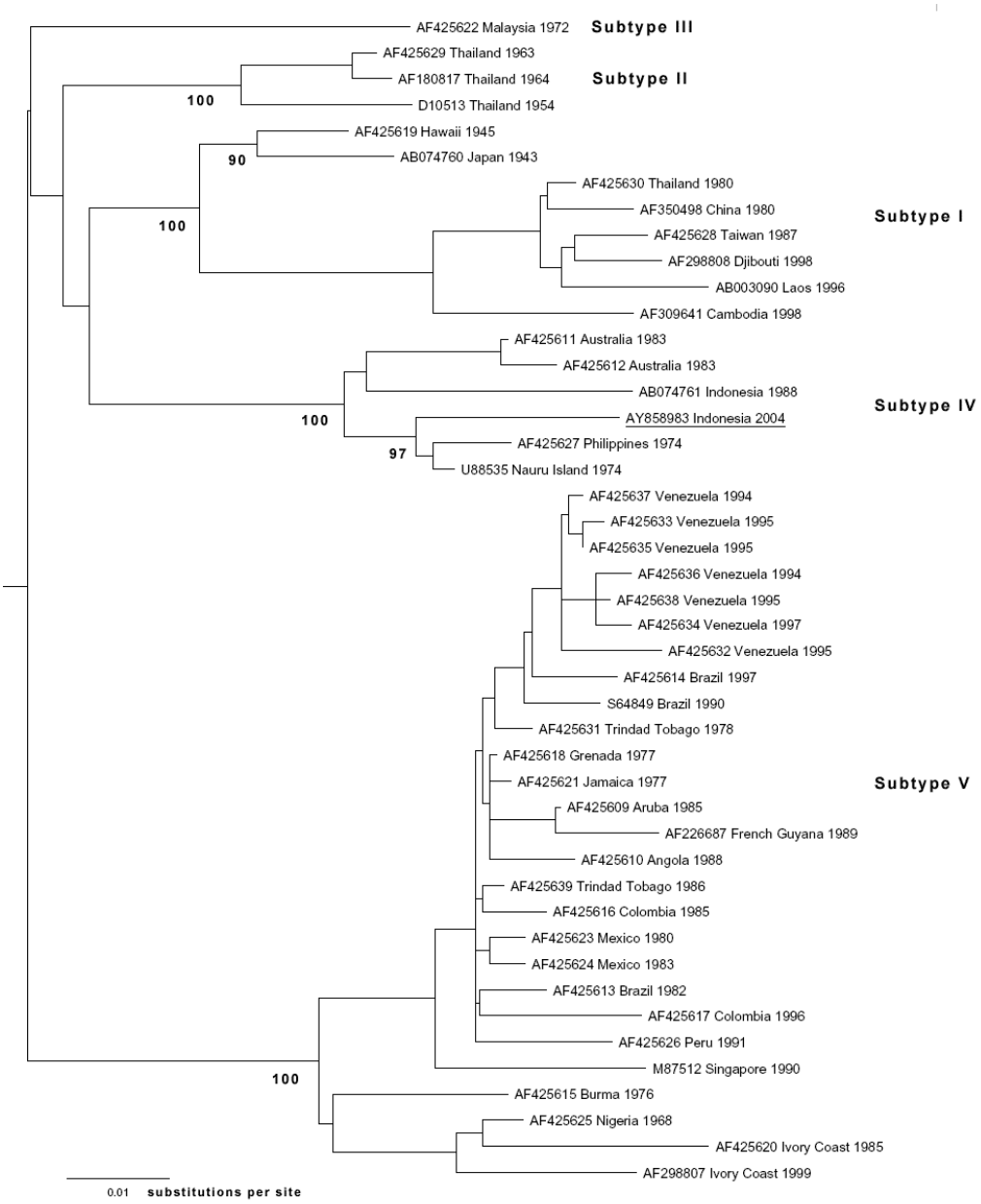


Figure 3.1. Phylogenetic tree of DENV-1 based on the complete nucleotide sequences of the E gene (1485 bases). The leaves are labeled with the GenBank accession number, country of isolation, and year of isolation. The dataset used is identical to the one used by Goncalvez *et al*. (2002) with the addition of AY858983, which fall within subtype IV of DENV-1.

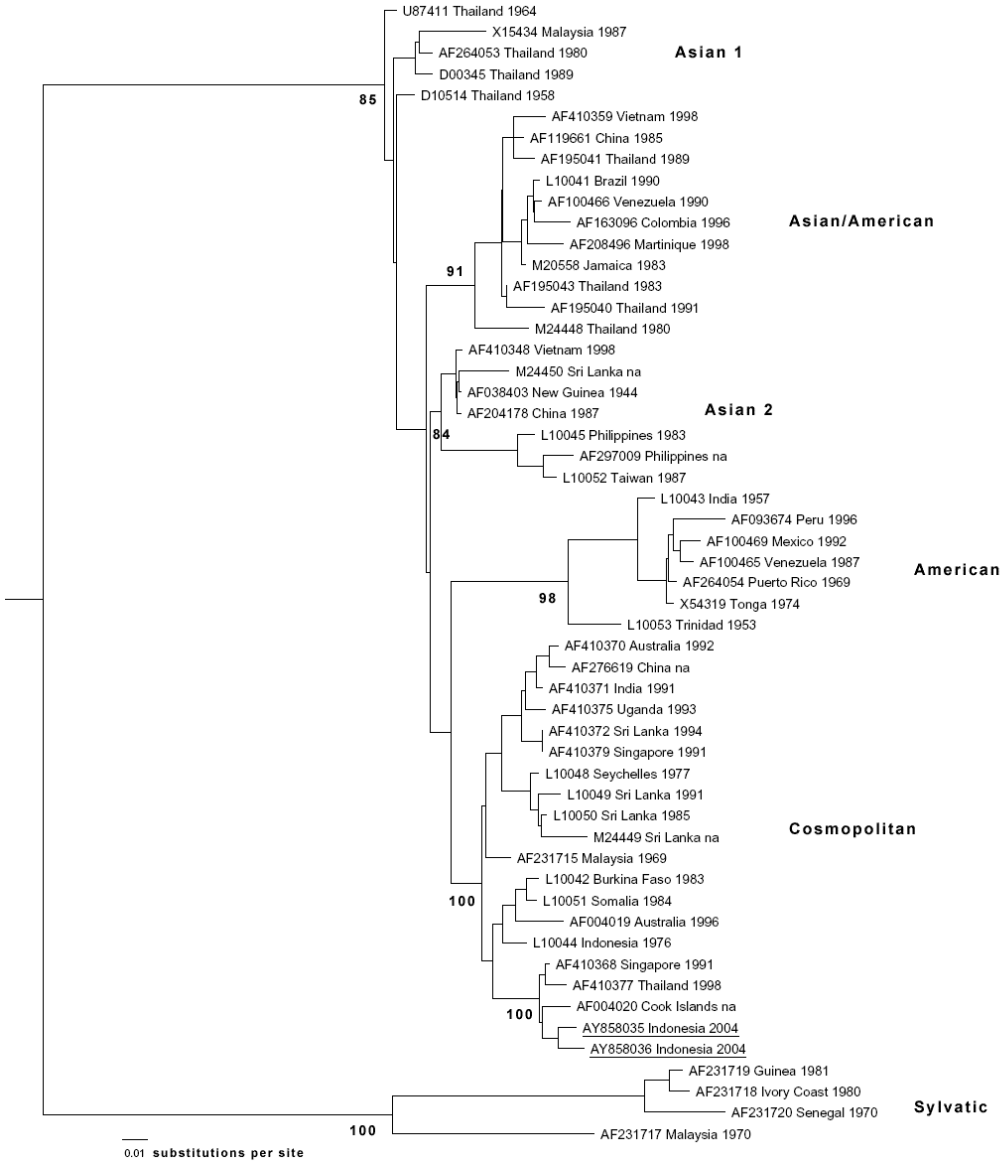


Figure 3.2. Phylogenetic tree of DENV-2 based on the complete nucleotide sequences of the E gene (1485 bases). The leaves are labeled with the GenBank accession number, country of isolation, and year of isolation. The dataset used is a subset of the one used by Twiddy *et al*. (2002), comprising just 52 of the original 147 strains, and with the addition of AY858035 and AY858036. The two 2004 Indonesian strains sequenced in this study fall within the ‘‘Cosmopolitan’’ subtype of DENV-2.

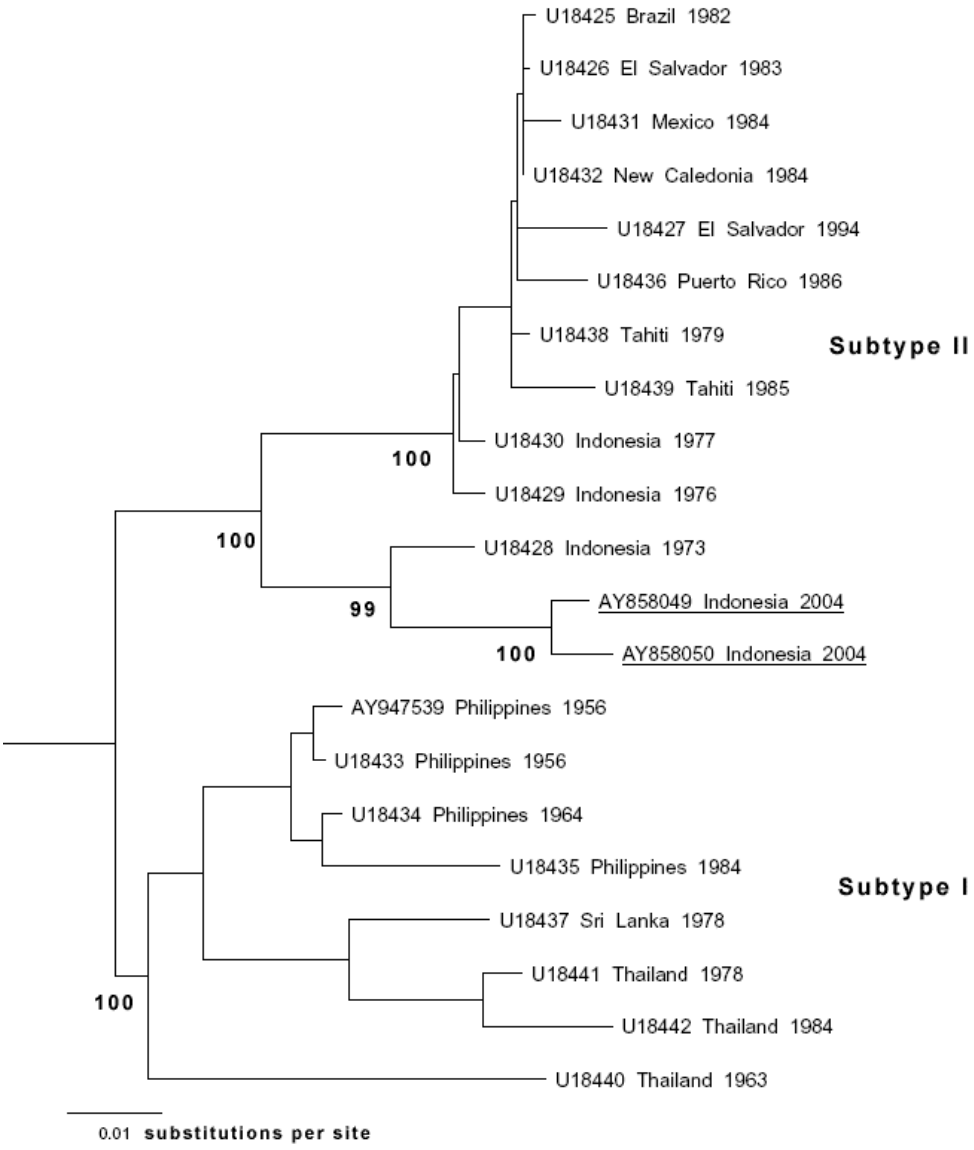


Figure 3.3. Phylogenetic tree of DENV-4 based on the complete nucleotide sequences of the E gene (1485 bases). The leaves are labeled with the GenBank accession number, country of isolation, and year of isolation. The dataset used is identical to the one used by Lanciotti *et al*. (1997) with the addition of AY858049, AY858050 and AY947539. The two 2004 Indonesian strains sequenced in this study fall within subtype II of DENV4.

The E gene phylogenetic tree for DENV-2 (Figure 3.2) places BA05i (AY858035) and TB61i (AY858036) in the Cosmopolitan subtype according to the classification scheme proposed by Twiddy *et al*. (2002). This subtype has previously been associated with DHF/DSS (Leitmeyer *et al*., 1999) and as the name suggests, is a diverse lineage that contains viruses from India, Southeast Asia, Africa, the Middle East and Australia. This means the two DENV-2 strains are likely to be regionally endemic strains.

The E gene phylogenetic tree for DENV-4 (Figure 3.3) shows both SW36i (AY858049) and SW38i (AY858050) as belonging to subtype II, a genetic lineage that contains viruses from Indonesia, the South Pacific and the Western hemisphere (Chungue *et al*., 1995; Lanciotti *et al*., 1997). Another Indonesian strain isolated in 1973 is also found in the clade that contains the two 2004 strains, thereby confirming the two as being endemic in Indonesia.

3.4.3 Diversity and phylogeny of the DENV-3 isolates

Figure 3.4 shows the phylogenetic tree constructed from a dataset containing the 40 DENV-3 isolates described by Messer *et al*. (2003), but with additional sequences from a DF-causing strain isolated in 2003 in Singapore (AY662691) and the newly sequenced Indonesian DENV-3 isolates. This tree is based on a 705-base segment covering pre-M/M and a portion of the E gene. All DENV-3 strains from Indonesia fall into subtype I, according to the classification described by Lanciotti *et al*. (1994), a lineage that can be traced all the way back to strains isolated in the same region in the early 1970s. In contrast, the strain from nearby Singapore clusters with subtype III strains which are commonly associated with viruses found in Eastern Africa and South Asia. The 2004 Indonesian strains can be further divided into two distinct clades, but there is no spatial clustering by the location of the admitting hospitals. The amount of data available is insufficient to attempt any correlation between genetic variation and the reported disease severity.

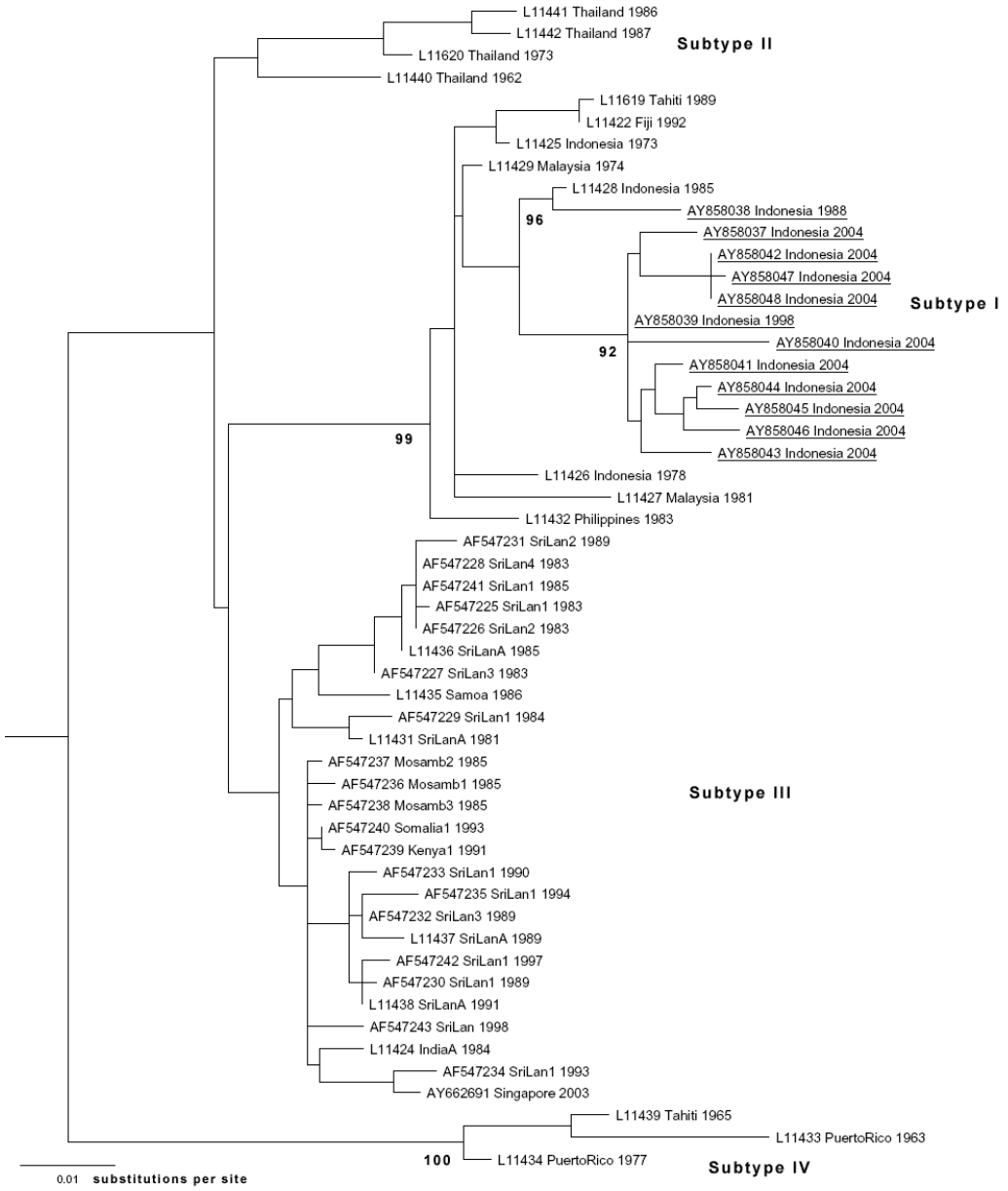


Figure 3.4. Phylogenetic tree of DENV-3 based on a 705-base segment covering the pre-M/M and a part of the E gene (from position 437 to 1141 of the DENV-3 genome). The leaves are labeled with the GenBank accession number, strain name/country of isolation, and year of isolation. The dataset used is identical to the one used by Messer *et al*. (2003) with the addition of AY858037-AY858048 and AY662691. All Indonesian strains sequenced in this study fall within subtype I of DENV-3 whereas the AY662691 strain from Singapore fall within subtype III.

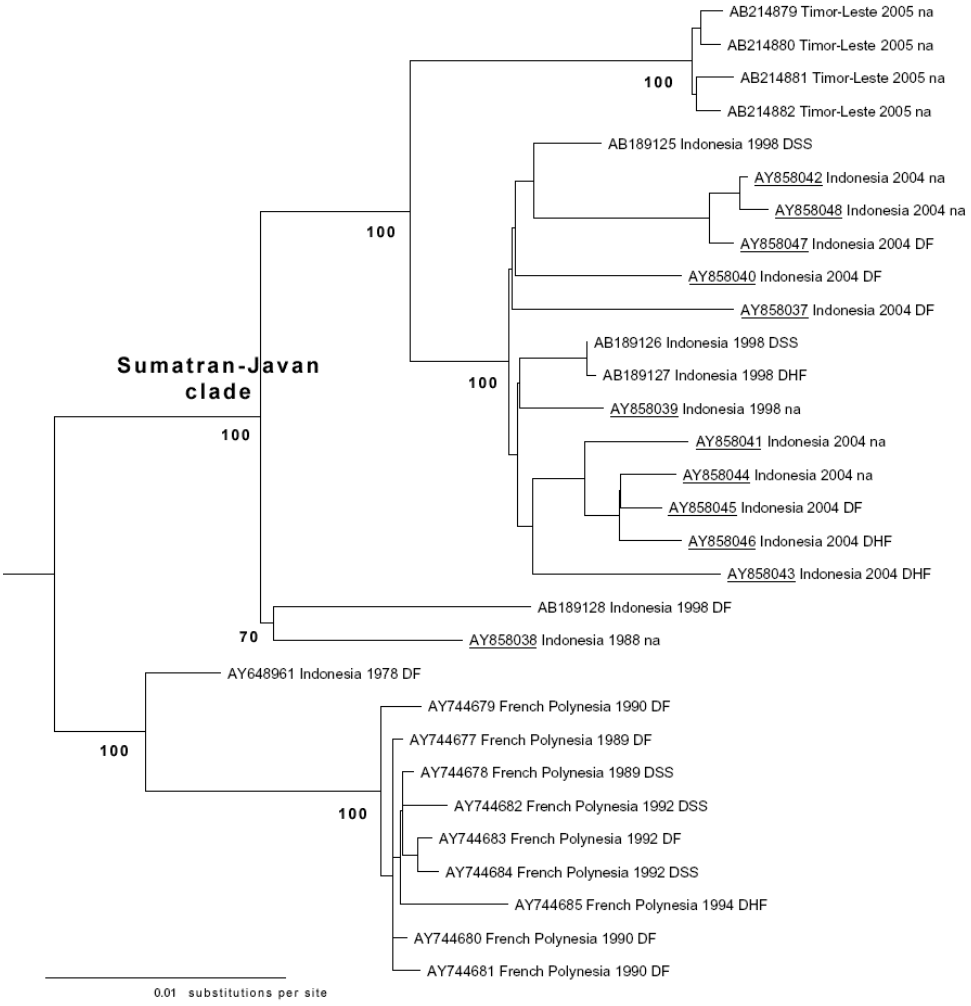


Figure 3.5. Phylogenetic tree based on 30 subtype I DENV-3 *whole-genome* nucleotide sequences. The tree leaves are labeled with the GenBank accession number, country of isolation, year of isolation and known clinical severity. The branch topology for the Indonesian strains sequenced in this study is similar to the prM-E tree in Figure 3.4.

After the intra-serotypic classification of the DENV-3 strains had been established, all published DENV-3 complete genomic sequences were collected to construct a whole-genome phylogeny. The whole-genome phylogenetic tree for 30 subtype I genomes (Figure 3.5) shows a distinct cluster that groups strains isolated from 1988, 1998 and 2004 in Indonesia, as well as strains isolated in a subsequent outbreak in Timor Leste in 2005 (WHO, 2005). This Sumatran-Javan clade is potentially a viral lineage that

possesses a superior level of evolutionary fitness and epidemic potential, as evident by its sustained transmission since 1970s and being implicated in major dengue epidemics in 1988, 1998 and 2004. In contrast, the other clade within subtype I does not have a similar epidemic-causing history, and has not been implicated in any epidemic since the early 1990s. The ability to cause DHF/DSS is not restricted to the Sumatran-Javan lineage and is observed in both lineages within subtype I (Figure 3.5), suggesting that disease severity and epidemic potential are likely to be governed by separate discrete factors.

3.4.4 Clade-specific mutations of the DENV-3 isolates

The availability of complete genome sequences made it possible to identify amino acid mutations that are specific to the Sumatran-Javan lineage, which could be the viral genetic elements that contribute to its continuing circulation in the region and its epidemic potential. Comparative analysis identified 24 amino acid residues that are significantly different between the Sumatran-Javan lineage and the other subtype I DENV-3 strains at the 0.05 significance level (Figure 3.6).

The impact of amino acid changes on protein function were subsequently predicted using SIFT (Ng and Henikoff, 2001) to identify residues that could be conferring competitive fitness advantage to the Sumatran-Javan lineage. Within the prM protein an A107T substitution located downstream of the furin cleavage site and retained in the mature virus was identified. Two non-conservative mutations were found in the E protein: a leucine typically found at position 124 of the E protein in strains within the Sumatran-Javan clade is replaced by a serine (L124S) outside the clade; the second mutation (S301L) is located within domain III of the E protein which is believed to be involved in antigenicity (Roehrig, 2003). There were also mutations within the non-structural proteins but none that are chemically non-conservative and predicted as not tolerated by SIFT (Table 3.2).

Using the single likelihood ancestor counting (SLAC) method (Pond and Frost, 2005a, b) we determined site-specific selection pressures acting on all the available DENV-3 coding sequences. While many sites seem to be

under a negative selection pressure, no evidence of positive selection was found. Unfortunately the Sumatran-Javan clade is too small to confidently detect any evidence of positive selection within this group.

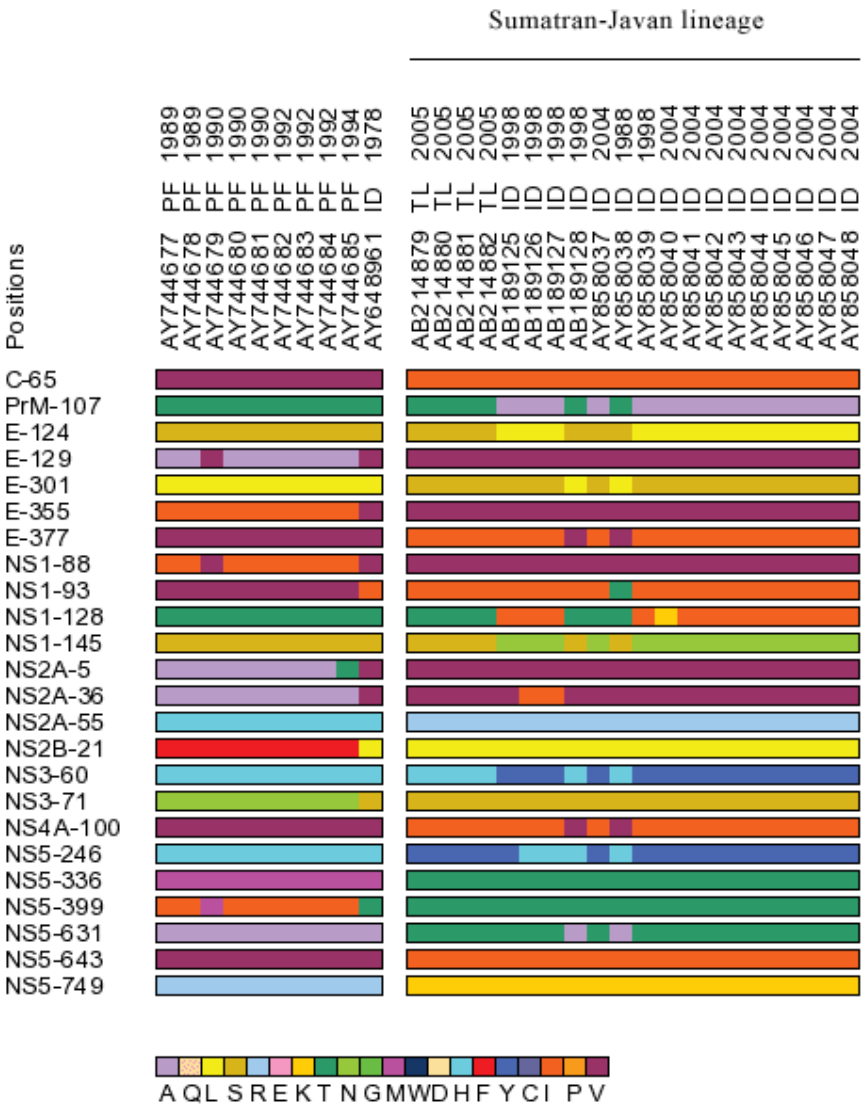


Figure 3.6. Amino acid residues that are significantly different between the Sumatran-Javan lineage and other subtype I DENV-3 strains ($p < 0.05$, chi-square). Sequences AY858037-AY858048 are the strains sequenced in this study. Amino acids are color-coded as shown in the key beneath the main diagram.

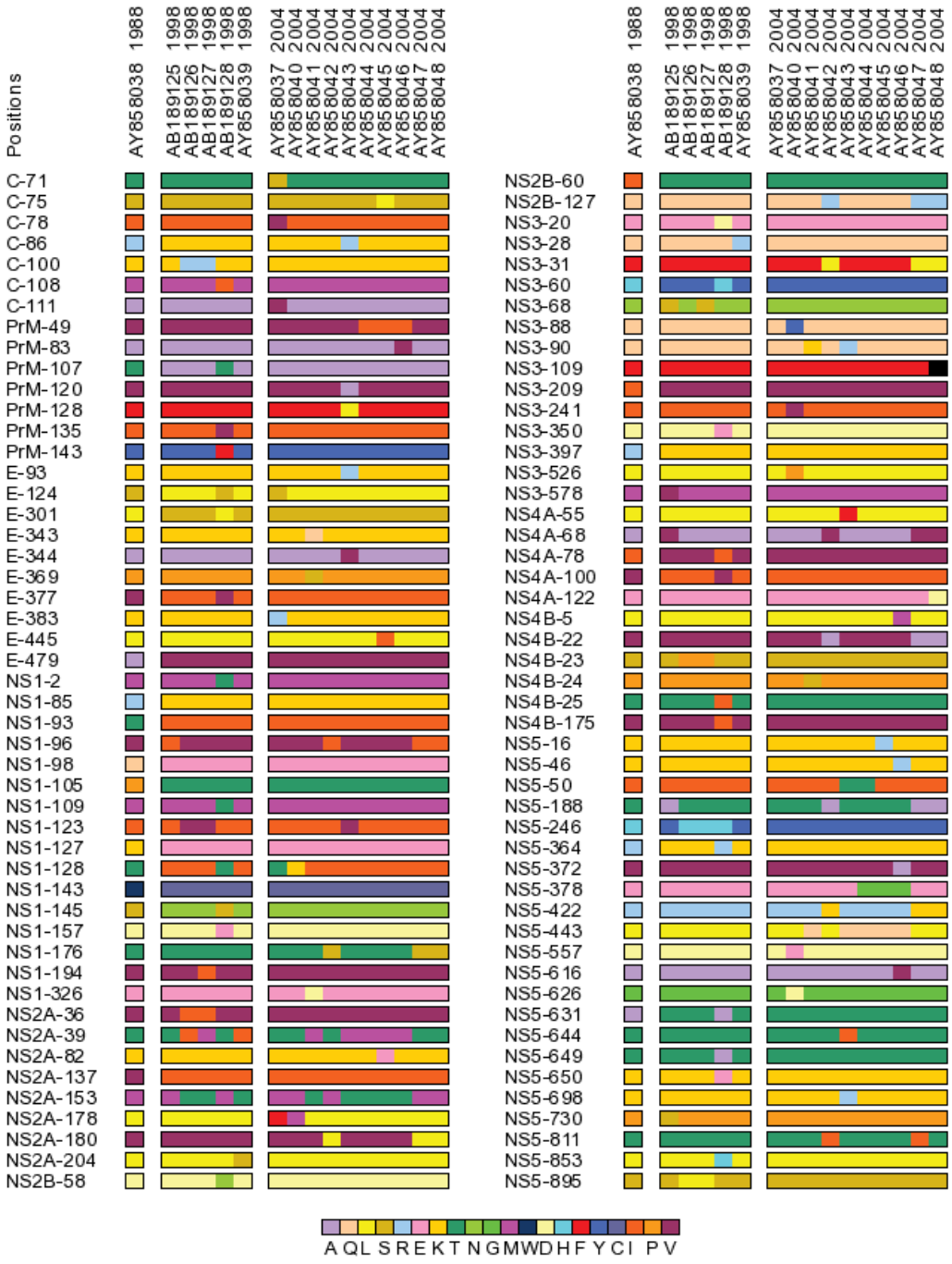


Figure 3.7. Amino acid differences among the Indonesian DENV-3 strains isolated in 1988, 1998 and 2004. The sequences are sorted in chronological order followed by accession number, starting from the left. Only 98 positions out of 3390 of the DENV-3 polyprotein reported any mutations, and none could be suggested as the trigger for the epidemic in 2004. Amino acids are color-coded as shown in the key beneath the main diagram.

A similar comparison of the deduced amino acid sequences involving only the 16 Indonesian DENV-3 strains from the epidemics in 1988, 1998 and 2004 did not detect any specific mutation that could have served as the trigger of the epidemic in 2004 (Figure 3.7). Only 98 sites out of 3390 (2.89%) of the DENV-3 polyprotein reported any mutation, and this number drops to 87 if den3_88 from 1988 is removed from the analysis. From the pattern of mutations observed these differences are likely to be the result of random mutations that confers little or no evolutionary advantage.

3.5 Discussion

The genome sequences of three serotypes of DENV were successfully obtained using a genome sequencing strategy that can be used on patient serum samples. Success with DENV-4 sequencing using this approach is expected in the future as more DENV-4 genomes become available for better RT-PCR primer design. Based on the obtained sequences, we have determined the molecular characteristics and phylogenetic relationships of the DENV strains isolated during the 2004 epidemic in Jakarta, Indonesia.

The phylogenetic trees for each of the four serotypes show that the viruses isolated in the 2004 epidemic cluster within subtypes that have been circulating in South East Asia for at least 30 years. The strongest evidence of these being endemic strains comes from the phylogenetic data for DENV-3. All 12 Indonesian DENV-3 strains sequenced in this study fall into subtype I which is recognized as comprising strains from Southeast Asia and the South Pacific islands. The close phylogenetic relationship observed between the strains isolated from Sumatra in 1998 and those from Jakarta in 2004 indicates that the 2004 epidemic might have had its origins in strains derived from viruses circulating in Sumatra in 1998, or possibly in a common ancestor of the Sumatran strains (Figure 3.5). Sumatra is one of the largest Indonesian islands and Jakarta on the nearby island of Java is the capital city of Indonesia, therefore frequent transmission by travelers is likely.

The identified Sumatran-Javan clade links the newly sequenced DENV-3 strains directly to the strains implicated in the epidemic on the island of Java in 1988, on the island of Sumatra 6 years earlier and those that subsequently caused an outbreak in Timor Leste in 2005. This clearly suggests these viruses possess an inherent epidemic potential. Since viruses from this particular lineage have been implicated in causing four epidemics in the past two decades, it is very likely these viruses could yet cause another dengue epidemic in Indonesia in the near future if they remained in circulation. We then examined the clade-specific mutations of the Sumatran-Javan DENV-3 isolates for potential viral genetic markers that could be the trigger of the epidemics. However, comparative study of the deduced polyprotein sequences established that re-emergence of little-changed endemic strains, and not newly-acquired amino acid mutations by the DENV-3 strains, as the most likely cause of the dengue epidemic in 2004 (Figure 3.7).

According to a parallel serological study involving 272 hospitalised patients, all four dengue serotypes were detected during the 2004 epidemic with DENV-3 being the predominant circulating serotype (Suwandono *et al*., 2006). Similar serological result was reported for the 1998 epidemic in south Sumatra, and the predominance of DENV-3 as a result of inherent sampling bias (on the premise that DENV-3 causes more severe illness and therefore causes more hospitalisation that facilitates sampling) has been proposed (Corwin *et al*., 2001). The fact that all four serotypes were found to be circulating in successive epidemics adds credence to the suggestion that advantageous amino acid mutations is not the trigger of the epidemics, otherwise this hypothetical fitter form would then dominate and become the sole serotype detected in subsequent outbreaks.

Table 3.2. Amino acid residues that potentially confer competitive fitness advantage to the Sumatran-Javan DENV-3 lineage when compared to other subtype I DENV-3 genomes. The two most likely candidates are shaded in grey. Mutations predicted by SIFT as "Tolerated" are less likely to have an impact on the fitness level of the virus. Threshold of intolerance used by SIFT is 0.05.

Viral Protein	Mutation	Functional Domain	Amino Acid Variation	SIFT Prediction
C	I65V	Helix α3	Conservative	Not tolerated
prM/M	A107T	-	Conservative	Tolerated
E	L124S	Domain II	Non-conservative	Not tolerated
E	V129A	Domain II	Conservative	Tolerated
E	S301L	Domain III	Non-conservative	Not tolerated
E	V355I	Domain III	Conservative	Not tolerated
E	I377V	Domain III	Conservative	Tolerated
NS1	V88I	-	Conservative	Tolerated
NS1	I93V	-	Conservative	Tolerated
NS1	I128T	-	Non-conservative	Tolerated
NS1	N145S	-	Conservative	Tolerated
NS2A	V5A	-	Conservative	Not tolerated
NS2A	V36A	-	Conservative	Not tolerated
NS2A	R55H	-	Conservative	Not tolerated
NS2B	L21F	-	Conservative	Not tolerated
NS3	Y60H	Protease	Conservative	Tolerated
NS3	S71N	Protease	Conservative	Tolerated
NS4A	I100V	Transmembrane Segment 2	Conservative	Tolerated
NS5	Y246H	Methyltransferase	Conservative	Tolerated
NS5	T336M	Polymerase	Non-conservative	Tolerated
NS5	T399I	Polymerase	Non-conservative	Tolerated
NS5	T631A	Polymerase	Conservative	Tolerated
NS5	I643V	Polymerase	Conservative	Tolerated
NS5	K749R	Polymerase	Conservative	Tolerated

The availability of genome sequence data has made possible the attempt to identify putative viral genetic factors that contributed to the continuing circulation of the Sumatran-Javan lineage and its epidemic potential. Two candidate residues were found in the E protein (Table 3.2) that may account for this difference however no conclusive evidence of positive selection was found. It is clear the genome sequencing and analysis strategy broadens the search for novel functional mutants beyond the scope of most previous studies which have mainly focused on using nucleotide sequences from the three dengue structural proteins.

The sequence data obtained in this study tells little about the role of population dynamics of the four serotypes in causing epidemics, but it clearly shows that the viruses that caused the 2004 epidemic in Indonesia are local strains that have been circulating in the region for a few decades. The identified Sumatran-Javan lineage of DENV-3 apparently is robust enough for sustained transmissions and has demonstrated strong epidemic potential spanning two decades.

3.6 Acknowledgements

We thank Lam Sai Kit and Amin Soebandario for encouraging the initiation of this collaboration, Klaus Ribbe of Novartis Indonesia for valuable assistance during site visits, and Edward C. Holmes for his advice on the phylogenetic analyses. We also wish to acknowledge the advice and discussions of Lim Siew Pheng from NITD and the technical skills of Khoo Chen Ai and Pauline Aw Poh Kim from the Infectious Disease Group together with the Cloning and Sequencing Group (headed by Yi Jun Ruan and Chia Lin Wei) at the Genome Institute of Singapore.

3.7 Author Contributions

SH Ong sequenced the genomes, assembled the sequences, performed the sequence analysis and wrote the manuscript. Contributions from co-authors are as listed: conceived and designed the experiments: WPM, MLH, SGV, MJS; provided the samples: AS; collected the samples: EL, BH, CGB; propagated the virus samples: WL, SH; sequenced the genomes: JTY, YLC; reviewed the manuscript: WPM, MLH, SGV and MJS.

3.8 References

16. Chungue, E., Cassar, O., Drouet, M.T., Guzman, M.G., Laille, M., Rosen, L., Deubel, V., 1995. Molecular epidemiology of dengue-1 and dengue-4 viruses. *J. Gen. Virol.* 76, 1877–1884.
17. Corwin, A.L., Larasati, R.P., Bangs, M.J., Wuryadi, S., Arjoso, S., Sukri, N., Listyaningsih, E., Hartati, S., Namursa, R., Anwar, Z., Chandra, S., Loho, B., Ahmad, H., Campbell, J.R., Porter, K.R., 2001. Epidemic dengue transmission in southern Sumatra, Indonesia. *Trans. R. Soc. Trop. Med. Hyg.* 95, 257–265.
18. Goncalvez, A.P., Escalante, A.A., Pujol, F.H., Ludert, J.E., Tovar, D., Salas, R.A., Liprandi, F., 2002. Diversity and evolution of the envelope gene of dengue virus type 1. *Virology* 303, 110–119.
19. Gubler, D.J., Clark, G.G., 1995. Dengue/dengue hemorrhagic fever: the emergence of a global health problem. *Emer. Infec. Dis.* 1, 55–57.
20. Gubler, D.J., 2002. Epidemic dengue/dengue haermorrhagic fever as a public health, social and economic problem in the 21st century. *Trends Microbiol.* 10, 100–103.
21. Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* 89, 10915–10919.
22. Lanciotti, R.S., Calisher, C.H., Gubler, D.J., Chang, G.J., Vorndam, A.V., 1992. Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. *J. Clin. Microbiol.* 30, 545–551.

23. Lanciotti, R.S., Lewis, J.G., Gubler, D.J., Trent, D.W., 1994. Molecular evolution and epidemiology of dengue-3 viruses. *J. Gen. Virol.* 75, 65–75.
24. Lanciotti, R.S., Gubler, D.J., Trent, D.W., 1997. Molecular evolution and phylogeny of dengue-4 viruses. *J. Gen. Virol.* 78, 2279–2286.
25. Leitmeyer, K.C., Vaughn, D.W., Watts, D.M., Salas, R., Villalobos, I., de Chacon, Ramos, C., Rico-Hesse, R., 1999. Dengue virus structural differences that correlate with pathogenesis. *J. Virol.* 73 (6), 4738–4747.
26. Li, K.B., 2003. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics* 19, 1585–1586.
27. Messer, W.B., Gubler, D.J., Harris, E., Sivananthan, K., de Silva, A., 2003. Emergence and global spread of a dengue serotype 3, subtype III virus. *Emer. Infec. Dis.* 9, 800–809.
28. Monath, T.P., 1994. Dengue: the risk to developed and developing countries. *Proc. Natl. Acad. Sci.* 91, 2395–2400.
29. Ng, P.C., Henikoff, S., 2001. Predicting deleterious amino acid substitution. *Genome Res.* 11, 863–874.
30. Pond, S.L., Frost, S.D., 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533.
31. Pond, S.L., Frost, S.D., 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222.
32. Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
33. Rico-Hesse, R., 1990. Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology* 174, 479–493.
34. Rico-Hesse, R., Harrison, L.M., Salas, R.A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M.T.R., Nogueira, R.M.R., da Roasa, A.T., 1997. Origins of dengue type-2 viruses associated with increased pathogenicity in the Americas. *Virology* 230, 244–251.
35. Roehrig, J.T., 2003. Antigenic structure of flavivirus proteins. *Adv. Virus Res.* 59, 141–175.

36. Seah, C.L., Chow, V.T., Tan, H.C., Can, Y.C., 1995. Rapid, single-step RT-PCR typing of dengue viruses using five NS3 gene primers. *J. Virol. Methods* 51, 193–200.
37. Schreiber, M.J., Ong, S.H., Holland, R.C.G., Hibberd, M.L., Vasudevan, S.G., Mitchell, W.P., Holmes, E.C., 2007. DengueInfo: a web portal to dengue information resources. *Infect. Genet. Evol.* 7, 540–541.
38. Sumarmo, 1987. Dengue haemorrhagic fever in Indonesia. *Southeast Asian J. Trop. Med. Public Health* 18, 269–274.
39. Suwandono, A., Kosasih, H., Nurhayati, Kusriastuti, R., Harun, S., Ma'roef, C., Wuryadi, S., Herianto, B., Yuwono, D., Porter, K.R., Beckett, C.G., Blair, P.J., 2006. Four dengue virus serotypes found circulating during an outbreak of dengue fever and dengue haemorrhagic fever in Jakarta, Indonesia, during 2004. *Trans. R. Soc. Trop. Med. Hyg.* 100, 855–862.
40. Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
41. Twiddy, S.S., Farrar, J.J., Vinh Chau, N., Wills, Gould, E.A., Gritsun,T., Lloyd, G., Holmes, E.C., 2002. Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology* 298, 63–72.
42. World Health Organization (WHO), 2002. Dengue and dengue haemorrhagic fever. Fact sheet no. 117.
43. World Health Organization (WHO), 2004. EPR: Dengue fever in Indonesia – update 4, May 11, 2004.
44. World Health Organization (WHO), 2005. Dengue haemorrhagic fever, Timor-Leste – update. *Wkly. Epidemiol. Rec.* 80, 85–86.

4. Genomic Epidemiology of a Dengue Virus Epidemic in Urban Singapore

Mark J. Schreiber[1], Edward C. Holmes[2,3], Swee Hoe Ong[1,4,5], Harold S.H. Soh[6], Wei Liu[1], Lukas Tanner[1], Pauline P.K. Aw[4], Hwee Cheng Tan[7], Lee Ching Ng[7], Yee Sin Leo[8], Jenny G H. Low[8], Adrian Ong[8], Eng Eong Ooi[9], Subhash G. Vasudevan[1†], Martin L. Hibberd[4]

1. Novartis Institute for Tropical Diseases, 10 Biopolis Road, Chromos #05-01, Singapore 138670.
2. Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA.
3. Fogarty International Center, National Institutes of Health, Bethesda, MD 20892. USA.
4. Genome Institute of Singapore, A*STAR, 60 Biopolis Street, Genome #02-01, Singapore 138672.
5. Swiss Tropical Institute, Socinstrasse 57, P.O. Box CH-4002 Basel, Switzerland.
6. Institute for High Performance Computing, A*STAR, 1 Science Park Road, #01-01 The Capricorn, Singapore Science Park II, Singapore 117528.
7. National Environment Agency, 40 Scotts Road, #13-00, Singapore 228231.
8. Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 308433.
9. DSO National Laboratories, 27 Medical Drive, #09-01, Singapore 117510.

† Current address: Emerging Infectious Diseases Program, Duke-NUS Graduate Medical School, 2 Jalan Bukit Merah, Singapore 169547.

This article has been published in:

Journal of Virology (2009) 83(9): 4163–4173

4.1 Abstract

Dengue is one of the most important emerging diseases of humans with no preventative vaccines or antiviral cures available at present. Although one-third of the world's population live at risk of infection, little is known about the pattern and dynamics of dengue virus (DENV) within outbreak situations. By exploiting genomic data from an intensively studied major outbreak we are able to describe the molecular epidemiology of DENV at a uniquely fine-scaled temporal and spatial resolution. Two DENV serotypes (DENV-1 and DENV-3), and multiple component genotypes, spread concurrently and with similar epidemiological and evolutionary profiles during the initial outbreak phase of a major dengue epidemic that took place in Singapore during 2005. Although DENV-1 and DENV-3 differed in viremia and clinical outcome, there was no evidence for adaptive evolution before, during, or after the outbreak, indicating that ecological or immunological rather than virological factors were the key determinants of epidemic dynamics.

4.2 Introduction

The phylogenetic analysis of gene sequence data is commonly used to determine the origins of disease outbreaks, particularly those caused by rapidly evolving RNA viruses (1, 2, 7, 9, 16, 26). Historically, such molecular epidemiological studies have usually utilized a small number of genes and have largely concentrated on determining the origins of disease outbreaks and retracing their pathways of spread. As such, these studies have rarely been able to shed light on the precise spatial and temporal dynamics of viral transmission. Recently, whole genome sequencing of viruses has been utilized to provide greater phylogenetic resolution on outbreak dynamics (12, 14, 15, 20) and is likely to become the benchmark in the near future.

One disease where genomic sequence may be especially informative is dengue fever (DF), an acute febrile disease caused by a mosquito-borne RNA virus (DENV; single-strand positive-sense, family Flaviviridae) and the most common vector-borne viral infection of humans; some 100 million dengue cases are reported on an annual basis, with epidemics especially

common in Southeast Asia (10, 11). The potential expansion of the viable geographic range for *Aedes aegypti* mosquitoes following global warming coupled with the current lack of an effective vaccine or anti-viral therapies make understanding the epidemic dynamics of this important emerging virus a key priority.

Dengue has the ability to cause major outbreaks in urban settings, often with high levels of morbidity. The Singapore dengue outbreak of 2005 was the largest of its kind in this locality, with a case rate of 335 per 100,000 population (17). The 2005 outbreak was also notable in that it was characterized by the co-circulation of two of the four viral serotypes – DENV-1 and DENV-3 – combined with a low level of DENV-2 transmission. The resurgence of dengue in Singapore is particularly striking given that an aggressive vector control program started in 1970 has resulted in a very low household index of mosquito breeding sites (21). Lessons learnt in Singapore, with its long history of commitment to dengue control, may therefore be vital to the overall global effort in dengue prevention.

By undertaking a detailed analysis of whole-genome data sampled from a major outbreak of dengue virus (DENV) in Singapore during 2005 – the first of its kind – we demonstrate how a synthesis of comparative genomics and fine-scale spatial and temporal epidemiological data provides unprecedented power to reveal the origins, causes and dynamics of this important emerging virus in a densely population urban environment.

4.3 Material and Methods

4.3.1 Collection of viral samples

Viral samples were collected as described in Low *et al.* (17). Briefly, blood samples were collected by a research nurse from consenting patients presenting with fever of $\geq$ 38°C for less than 72 hours. One ml of serum from samples confirmed as dengue positive by RT-PCR was inoculated onto the *Aedes albopictus* mosquito (C6/36) cell line (ATCC: CRL-1660). Cells were incubated at 37°C for up to 10 days or until 75% of the cell monolayer showed

cytopathic effects. Isolation of the virus was confirmed and serotyped by indirect immunofluorescence using DENV group-specific and DENV serotype-specific monoclonal antibodies.

4.3.2 Molecular Analysis

Viruses were propagated by two passages in C6/36 mosquito cell culture. Viral titer was measured using a plaque-assay. Viral titers of at least 1x106 were found to be required for optimal success in subsequent RT-PCR steps. Viral RNA was extracted from the culture supernatant using the QIAGEN QiaAmp kit and extraction protocol.

The extracted RNA was copied to cDNA using an RT reaction followed by PCR amplification. The virus was amplified as five overlapping fragments. Unless specified below the same conditions were used for all five fragments. RT of fragments 1, 3 and 4 was performed in a single tube. RT of fragments 2 and 5 was performed in separate tubes. Samples were kept on ice and pipetting was carried out using RNAase free filter tips. PCR primers used in the RT reaction are detailed in Supplementary Table 1 and reaction conditions are in Supplementary Table 2. The five RT fragments were subsequently amplified by PCR. The PCR primers that were used to amplify the RT fragments are detailed in Supplementary Table 3. PCR of the RT fragments was carried out in a thermal cycler using the program in Supplementary Table 4.

Gel electrophoresis was used for visualization of the PCR products as well as gel extraction and purification of products. The products were separated in 1% agarose TBE gels after visualization of ethidium bromide stained bands under UV light. DNA was extracted from bands excised from agarose gels using a Qiagen QiaQuick extraction kit.

Finally, gel-purified fragments were sequenced using an Applied Biosystems BigDye ddNTP capillary sequencer. Viral genome sequences generated in this study are deposited in GenBank with accession numbers EU081177-EU081281. All genome sequences, their GenBank accession

numbers and their standard strain names (27) used in the analyses in this paper are detailed in Supplementary Table 5.

4.3.3 Evolutionary analysis

To reveal the origins of the Singapore viruses we conducted a phylogenetic analysis of the complete coding region of the genome sequences of all those viruses sequenced here as well as those already available in GenBank. This resulted in data sets of the following size: DENV-1 = 145 taxa, 10,176 nt; DENV-2 = 116 taxa, 10,173 nt; DENV-3 = 122 taxa, 10,173 nt. To determine the best-fit model of nucleotide substitution we employed MODELTEST (24). In all cases the most general GTR+I+?4 model was favored. Maximum likelihood (ML) trees were then inferred under this model using PAUP* (30), employing TBR branch-swapping in each case. Finally, a neighbour-joining (NJ) bootstrap analysis (1000 replications), but employing the ML substitution model, was used to determine the robustness of key nodes on each phylogeny.

To determine the population dynamics of DENV-1 and DENV-3 during the 2005 dengue outbreak in Singapore we analyzed those isolates that clearly diversified during the course of the epidemic. For DENV-1, this meant our analysis was confined to 53 genome sequences from Singapore (genotype I), while 42 genomes (genotype III) were used in the case of DENV-3. There were insufficient sequences for an analysis of DENV-2. Demographic and evolutionary parameters for both serotypes were estimated using the Bayesian MCMC (Markov Chain Monte Carlo) approach implemented in the BEAST package (4). Because of the typically complex population dynamics we utilized the Bayesian skyline plot as a coalescent prior. This provides a piecewise graphical depiction of changes in the effective number of infections ($N_e\tau$), where N_e is the effective population size and τ the generation time. We also utilized both strict and relaxed (uncorrelated lognormal) molecular clocks. The GTR+Γ4 model of nucleotide substitution was used in all cases, with the invariant sites parameter (I) excluded as it tended to over-fit to the data. Uncertainty in the data for each estimated parameter is reflected in values of the 95% high probability density (HPD),

with all MCMC chains run for sufficient time (50 million steps, with 10% removed as burn-in) to ensure statistical convergence.

To determine the nature of selection pressures acting on each gene of DENV-1 and DENV-3 sampled from Singapore, we computed the mean ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions per site (d_N/d_S) using the Single Likelihood Ancestor Counting (SLAC) method available through the DATAMONKEY web interface (23) and assuming the GTR model of nucleotide substitution and an input NJ tree. This analysis also allowed us to compute the Tree Length (TL) in substitutions per site for each gene. In addition, we used the CODEML program within the PAML package (33) to estimate an overall d_N/d_S for the entire coding region of both serotypes (the 'one-ratio' model). This was compared to the case in which a separate d_N/d_S value was estimated for the external and internal branches of each data set (the 'two-ratio' model).

Finally, to determine the strength of spatial structure in both DENV-1 and DENV-3 we first obtained the physical address of each viral isolate and produced clusters according to their geographical proximity by K-means clustering. For DENV-1, the physical address was available for 48 isolates which were then placed into one of six different spatial groups (with a single letter character state code identifying each group). In the case of DENV-3, address information was available for 42 isolates and these were separated into four spatial groups. Although there are a variety of methods by which the extent of spatial structure in phylogenetic data can be determined, particularly utilizing parsimony character state mapping (28), we employed a Bayesian MCMC approach (22) thereby accounting for any error in the underlying phylogeny. This analysis was based on the trees output from the previous BEAST analysis (with a new BEAST analysis conducted on the 48 sequence DENV-1 data set), employing 1000 replications and with 10% of trees removed as burn-in. From these trees we computed the mean values, credible intervals, and significance of the parsimony score (PS) and association index (AI) statistics of the strength of geographical clustering (22).

4.4 Results and Discussion

Fortuitously, the 2005 outbreak coincided with the launch of a longitudinal "Early DENgue Infection and Outcome" (EDEN) study in the central Ang Mo Kio district of Singapore (17). From a sample of 133 RT-PCR dengue positive patients enrolled during the EDEN study, and collected between April and November 2005, serotyping determined 66 (48.9%) to be DENV-1, 62 (46.6%) to be DENV-3 and 5 to be DENV-2 (3.8%). No cases due to DENV-4 were observed (17), and one patient was found to be co-infected with serotypes 1 and 3. The detection of large numbers of DENV-3 in the Ang Mo Kio area was unusual as DENV-1 was the predominant serotype in the rest of Singapore. We were able to isolate 112 (84.2%) viruses, corresponding to 57 DENV-1, 50 DENV-3 and 5 DENV-2 isolates. Complete genome sequences were obtained for 54 DENV-1, 44 DENV-3 and 4 DENV-2 viruses. The remaining samples, although shown to be dengue positive by RT-PCR, did not yield sufficient viral RNA for genome sequencing.

To determine the origins of the viral isolates responsible for the 2005 dengue epidemic in Singapore we conducted a phylogenetic analysis of the complete genomes of the viruses sampled here combined with representative DENV isolates taken from GenBank. To assist in this comparison we also completed whole genome sequences of some historical DENV samples from Singapore. All but one of the DENV-1 genomes from this epidemic were classified as genotype I, which commonly circulates in Southeast Asia (Figure 4.1). The single outlier belongs to genotype III, which is predominantly found in Latin America and West Africa (8), although a genotype III virus was previously sampled in Singapore in 1993. The closest relatives of the 53 genotype I Singapore DENV-1 isolates are the Chinese isolates DENV-1/CN/Fj231/2004 and DENV-1/CN/ZH1067/XXX, suggesting that frequent transfer of DENV may occur between Singapore and China, and DENV-1/JP/20-Feb/XXX from Japan. Importantly, as three historical Singapore DENV-1 samples isolated in 2003 fell at the base of the 2005 cluster it is possible that this particular lineage of DENV-1 genotype I has been circulating continuously in Singapore since at least 2003.

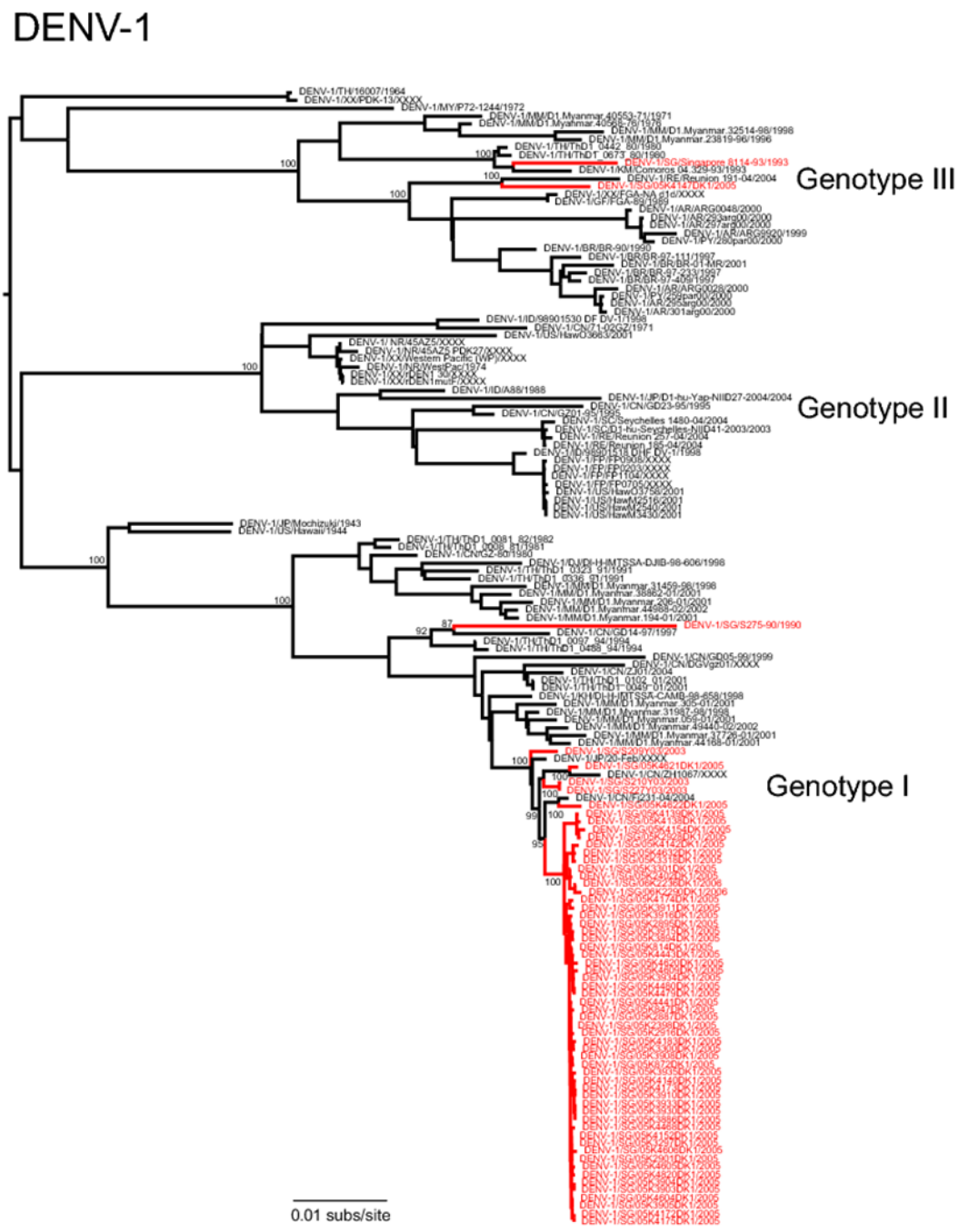


Figure 4.1. Phylogenetic relationships of 145 complete genomes DENV-1 sampled globally determined using a maximum likelihood method. Isolates sampled from Singapore are shown in red and individual genotypes are shown. Bootstrap values (>80%) are shown next to key nodes and all horizontal branch lengths are drawn to scale.

The four DENV-2 genomes form part of the “cosmopolitan” genotype (Figure 4.2), that has a wide distribution in tropical and subtropical localities (32). Close relatives of these strains include, DENV 2/ID/BA05i/2004 and DENV 2/ID/TB16i/2004, which were isolated during a dengue fever epidemic in Jakarta in 2004 (29) as well as three strains from Brunei Darussalam, one from China and an older isolate from Queensland, Australia (DENV-2/AU/TSV01/1993) (13), possibly introduced from the nearby Indonesian islands. As a number of these viruses were isolated between 2004 and 2006 it seems likely that this lineage was relatively common in this geographical area at the time of the outbreak.

The majority of DENV-3 genomes fell into genotype III (Figure 4.3). This genotype was originally associated with the Indian subcontinent until the mid-1990s when it was introduced into Latin America and the Caribbean (19). Of more importance from the perspective of this outbreak was that an isolate from genotype III was first detected in Singapore in 2003 (DENV-3/SG/S221/2003) and which fell basal to the 2005 outbreak viruses in our phylogenetic analysis. Such a phylogenetic pattern is compatible with the in situ evolution of this lineage in Singapore since at least 2003. Hence, as is also likely the case with DENV-1, the 2005 outbreak of DENV-3 may also be due to the amplification of a pre-existing viral lineage rather than the invasion of an ‘exotic’ DENV strain. To further test this hypothesis we obtained the additional genome sequence of a Singaporean DENV-3 genotype I isolate (DENV-3/SG/SS710/2004) sampled in 2004. As expected under the hypothesis of in situ evolution, this isolate occupies an intermediate position between the 2003 and 2005 strains. Finally, two of the 44 DENV-3 isolates from the 2005 outbreak in Singapore fall into genotype I which is endemic in the Malay archipelago. This observation provides an additional point of similarity between the DENV-1 and DENV-3 components of the 2005 DENV outbreak in Singapore: that individual epidemic serotypes can be composed of multiple viral genotypes.

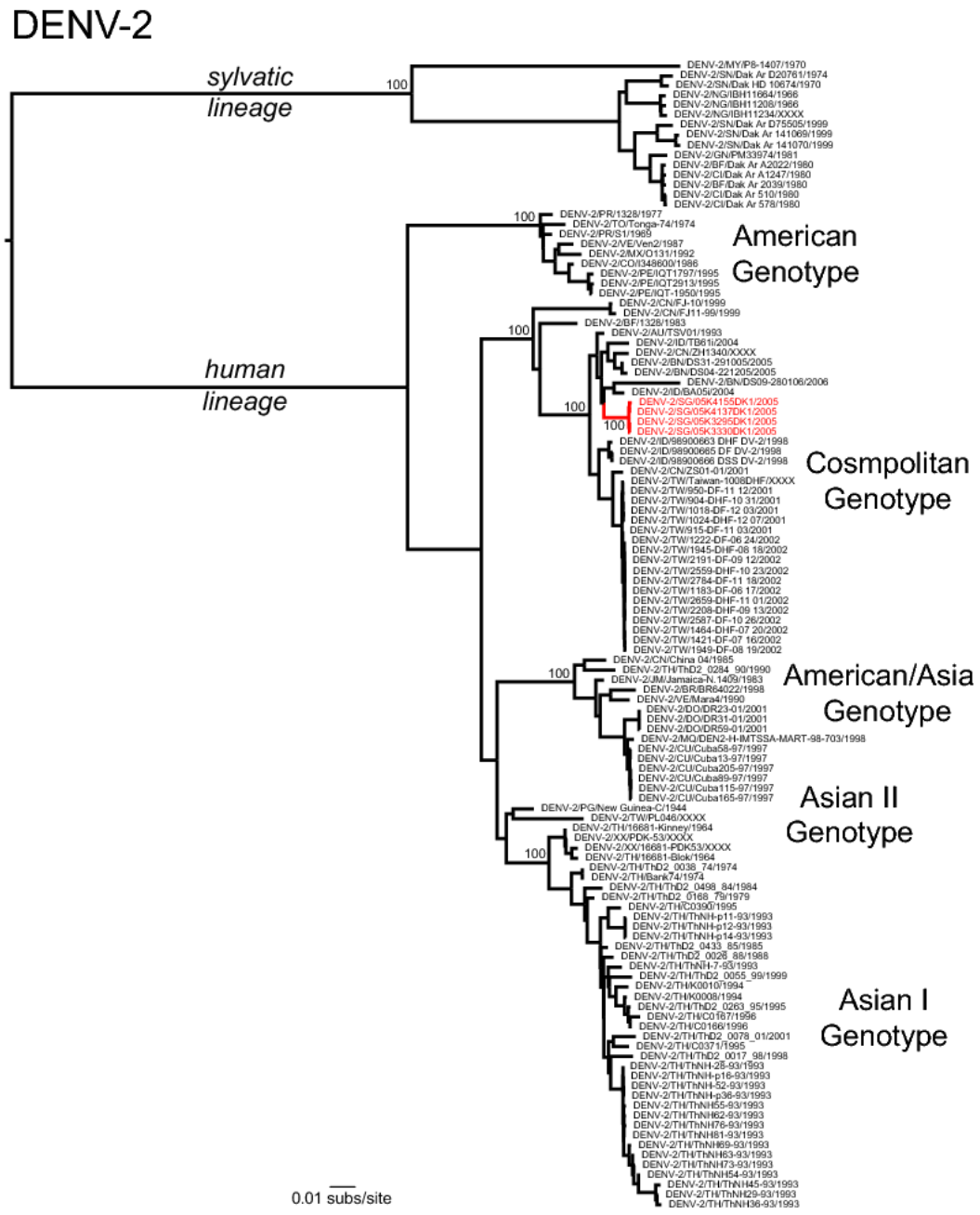


Figure 4.2. Phylogenetic relationships of 116 complete genomes DENV-2 sampled globally determined using a maximum likelihood method. Isolates sampled from Singapore are shown in red and individual genotypes are shown. Bootstrap values (>80%) are shown next to key nodes and all horizontal branch lengths are drawn to scale.

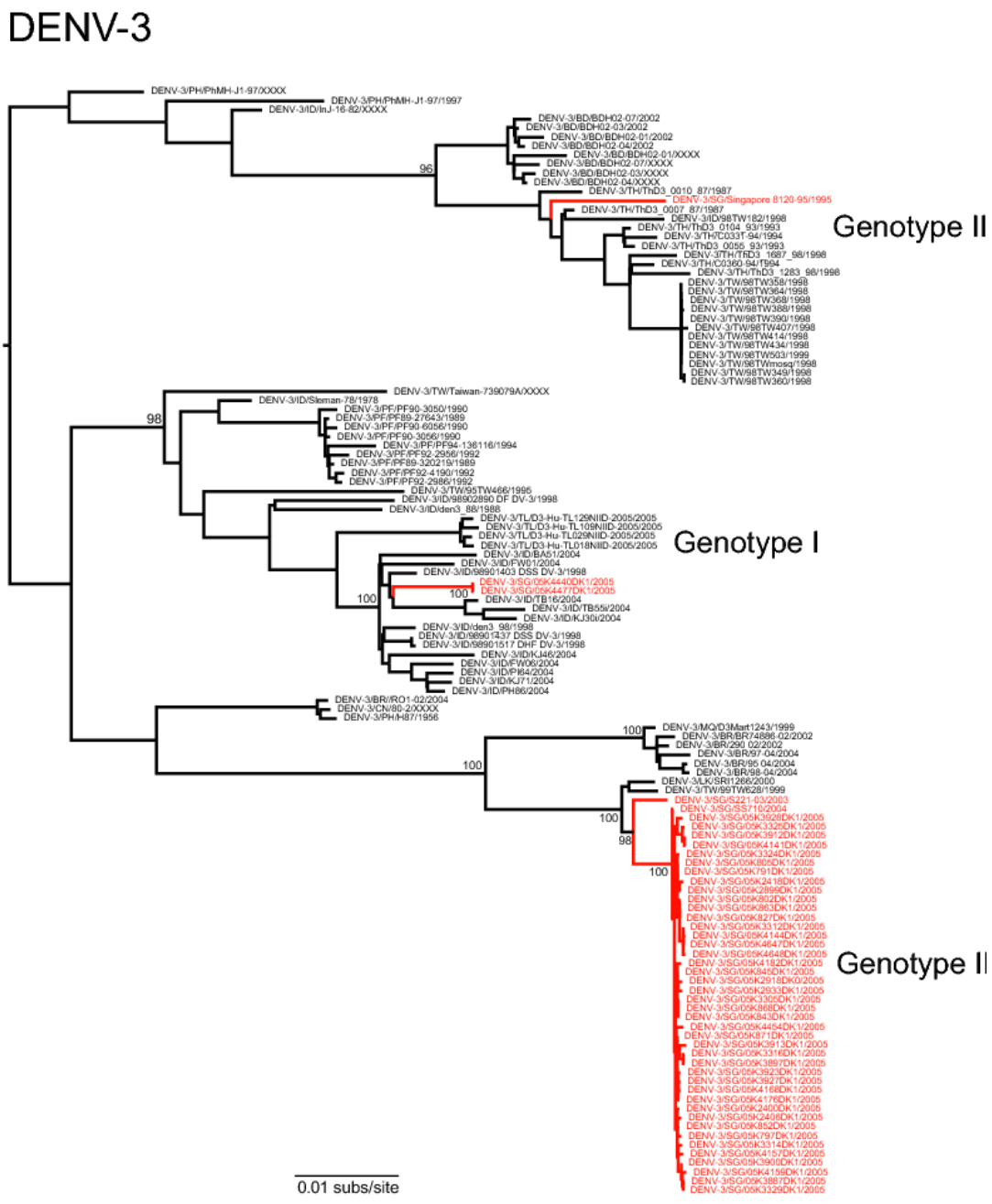


Figure 4.3. Phylogenetic relationships of 122 complete genomes DENV-3 sampled globally determined using a maximum likelihood method. Isolates sampled from Singapore are shown in red and individual genotypes are shown. Bootstrap values (>80%) are shown next to key nodes and all horizontal branch lengths are drawn to scale.

The majority of DENV-3 genomes fell into genotype III (Figure 4.3). This genotype was originally associated with the Indian subcontinent until the mid-1990s when it was introduced into Latin America and the Caribbean (19). Of more importance from the perspective of this outbreak was that an isolate from genotype III was first detected in Singapore in 2003 (DENV-3/SG/S221/2003) and which fell basal to the 2005 outbreak viruses in our phylogenetic analysis. Such a phylogenetic pattern is compatible with the in situ evolution of this lineage in Singapore since at least 2003. Hence, as is also likely the case with DENV-1, the 2005 outbreak of DENV-3 may also be due to the amplification of a pre-existing viral lineage rather than the invasion of an 'exotic' DENV strain. To further test this hypothesis we obtained the additional genome sequence of a Singaporean DENV-3 genotype I isolate (DENV-3/SG/SS710/2004) sampled in 2004. As expected under the hypothesis of in situ evolution, this isolate occupies an intermediate position between the 2003 and 2005 strains. Finally, two of the 44 DENV-3 isolates from the 2005 outbreak in Singapore fall into genotype I which is endemic in the Malay archipelago. This observation provides an additional point of similarity between the DENV-1 and DENV-3 components of the 2005 DENV outbreak in Singapore: that individual epidemic serotypes can be composed of multiple viral genotypes.

Figure 4.4a. Observed amino acid changes in DENV-1

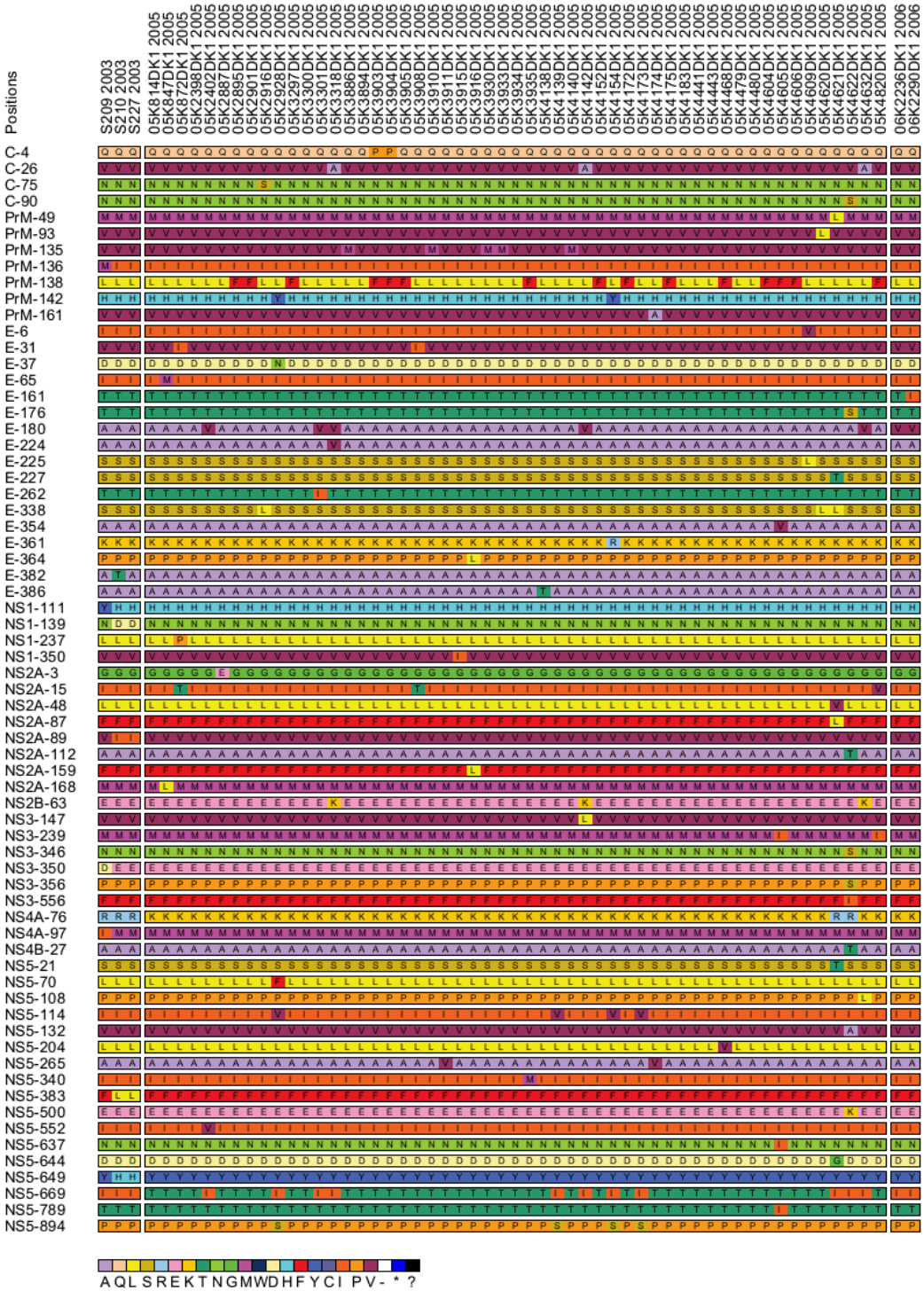


Figure 4.4b. Observed amino acid changes in DENV-3

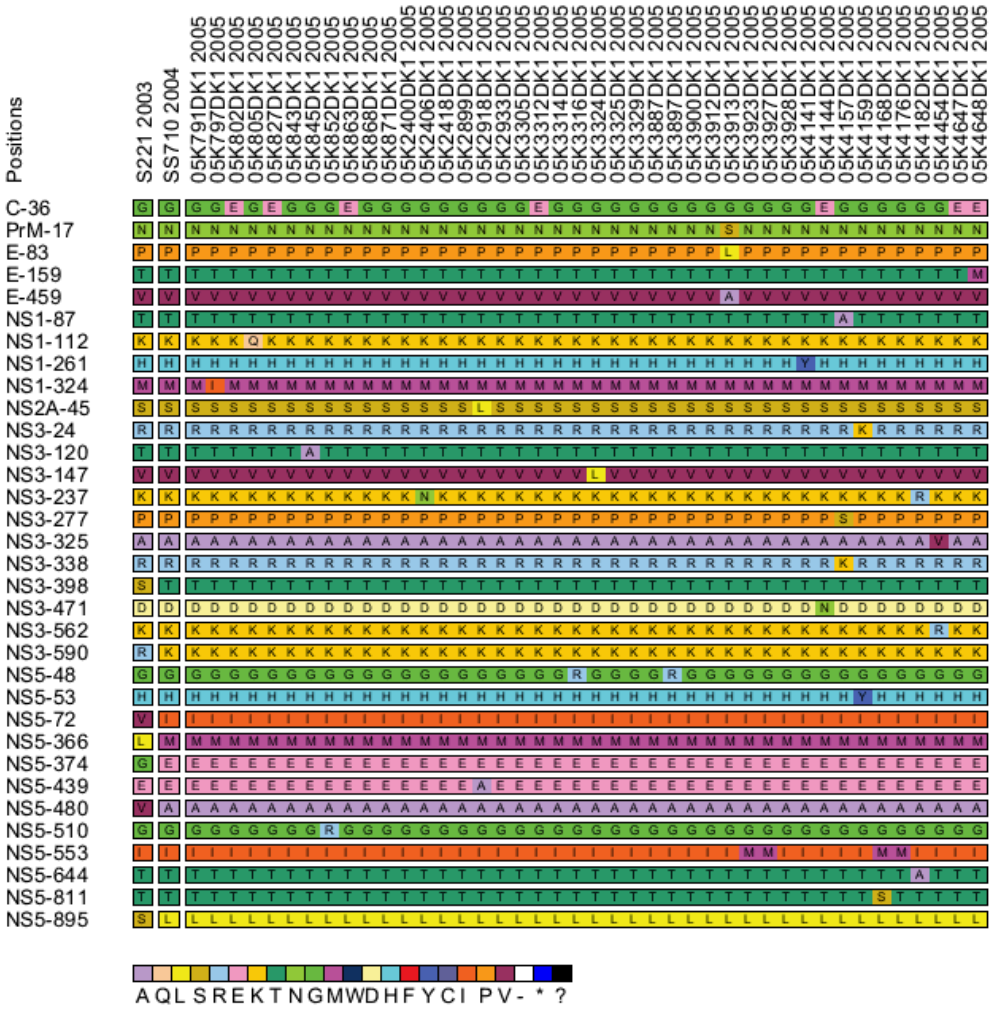


Figure 4.4. Observed amino acid changes in DENV-1 (a) and DENV-3 (b). Viral isolates are plotted on the x axis. The aligned polyproteins of each virus were compared to count the number and distribution of amino acid changes. Positions that were not completely conserved are shown with the individual protein name and the amino acid position within the protein on the Y axis. The color of each square indicates the type of amino acid residue found in isolate x at position y.

To determine the evolutionary processes that enabled the emergence of multiple DENV genotypes in a single outbreak we examined each of the amino acid changes associated with these viruses. Remarkably, no amino acid changes were completely fixed on the branches leading to the DENV-1 or DENV-3 isolates sampled in Singapore during 2005 (Figure 4.4a). In DENV-1 residue 76 of NS4A was observed to be K in all samples from 2003 and R in many 2005 samples. However, the penetration of the mutation was incomplete and is quite conservative in nature and therefore unlikely to be significant. Similarly, no amino acid variants in DENV-1 observed during 2005 were found in the two genomes isolated and sequenced from the 2006 non-epidemic year, indicating that no mutations that occurred in 2005 became fixed. Some substitutions were observed in DENV-3 between the 2003 and 2004 non-epidemic strains although all are conservative except for the change from serine to lysine at residue 895 in NS5. Position 895 is not conserved in the four serotypes and is usually S, P or E. In the recently solved structure of the DENV-3 RNA dependent RNA polymerase (34) position 895 is near the C terminal (position 900) and does not appear to functionally significant. More notably, there are no fixations between the 2004 isolate and the 2005 epidemic isolates. Although a number of non-conservative substitutions were observed within the 2005 isolates of DENV-3, that this serotype was not detected in 2006 indicates that none were capable of perpetuating the clade (Figure 4.4b). Finally, there was no evidence for positive selection in any gene of the Singapore viruses, with a relatively low ratio of nonsynonymous to synonymous substitutions per site (d_N/d_S) in all genes and no evidence for site-specific positive selection (Table 4.1; which also gives a variety of other gene-specific measures of genetic diversity).

Table 4.1. Phylogenetic and evolutionary patterns among the proteins of DENV-1 and DENV-3 sampled from Singapore during 2005.

Protein	Length (nt)	DENV-1			DENV-3		
		IS[a]	T_L[b]	d_N/d_S	IS	T_L	d_N/d_S
Capsid	342	5	0.026	0.359	3	0.023	0.064
Membrane	498	9	0.034	0.332	1	0.008	0.175
Envelope	1479-1485[c]	11	0.031	0.206	9	0.013	0.095
NS1	1056	7	0.025	0.042	4	0.011	0.204
NS2A	654	6	0.029	0.318	3	0.009	0.067
NS2B	390	8	0.043	0.029	1	0.010	0
NS3	1857	16	0.024	0.059	4	0.014	0.265
NS4A	450	6	0.029	0.037	2	0.009	0
NS4B	744-747[c]	6	0.031	0.019	3	0.009	0
NS5	2679-2700[c]	19	0.022	0.157	11	0.011	0.131
Genome (coding)	*10,158 - 10,170*	*96*	*NA*[d]	*NA*	*41*	*NA*	*NA*

[a] Number of parsimony informative sites, [b] Tree length in substitutions per site, [c] Length for DENV-1 and DENV-3 respectively, [d] Not applicable. No significant evidence for positive selection was observed in any gene in either DENV-1 or DENV-3.

To infer the epidemiological dynamics of DENV-1 and DENV-3 during the Singapore outbreak we used a Bayesian coalescent approach (3, 5) incorporating data on the exact day of viral sampling. For both serotypes we estimated the changing patterns of relative genetic diversity through time as reflected in the effective number of infections ($N_e\tau$) using a Bayesian skyline plot and assuming a relaxed (uncorrelated lognormal) molecular clock (although very similar results were observed under a strict molecular clock; results available from the authors on request). Similar epidemic profiles were observed in both viruses, comprising a rapid growth phase followed by a constant population size, although the mean age of the common ancestor was significantly greater in DENV-1 (1740 days; 95% HPD = 741 – 3222 days) compared to DENV-3 (298 days; 95% HPD = 225 – 387 days), indicating that already diverse lineages of DENV-1 were present in Singapore at the outset of the 2005 outbreak (Figure 4.5). Similarly, mean estimates of peak $N_e\tau$ were smaller in DENV-3 (1632; 95% HPD = 56 – 4334) than DENV-1 (7211; 95% HPD = 637 – 23130), although with overlapping HPD values. Interestingly DENV-3 was not widely reported in Singapore during the outbreak and appeared to be mainly contained to the sampling area which supports this result. The substitution dynamics of both serotypes were also similar, with mean evolutionary rates of 1×10^{-3} subs/site/year (95% HPD, 0.4×10^{-3} – 1.6×10^{-3} subs/site/year) and 1.3×10^{-3} subs/site/year (95% HPD, 8.7×10^{-4} – 1.8×10^{-3} subs/site/year) for DENV-1 and DENV-3, respectively, and equivalent to those estimated previously for DENV (32).

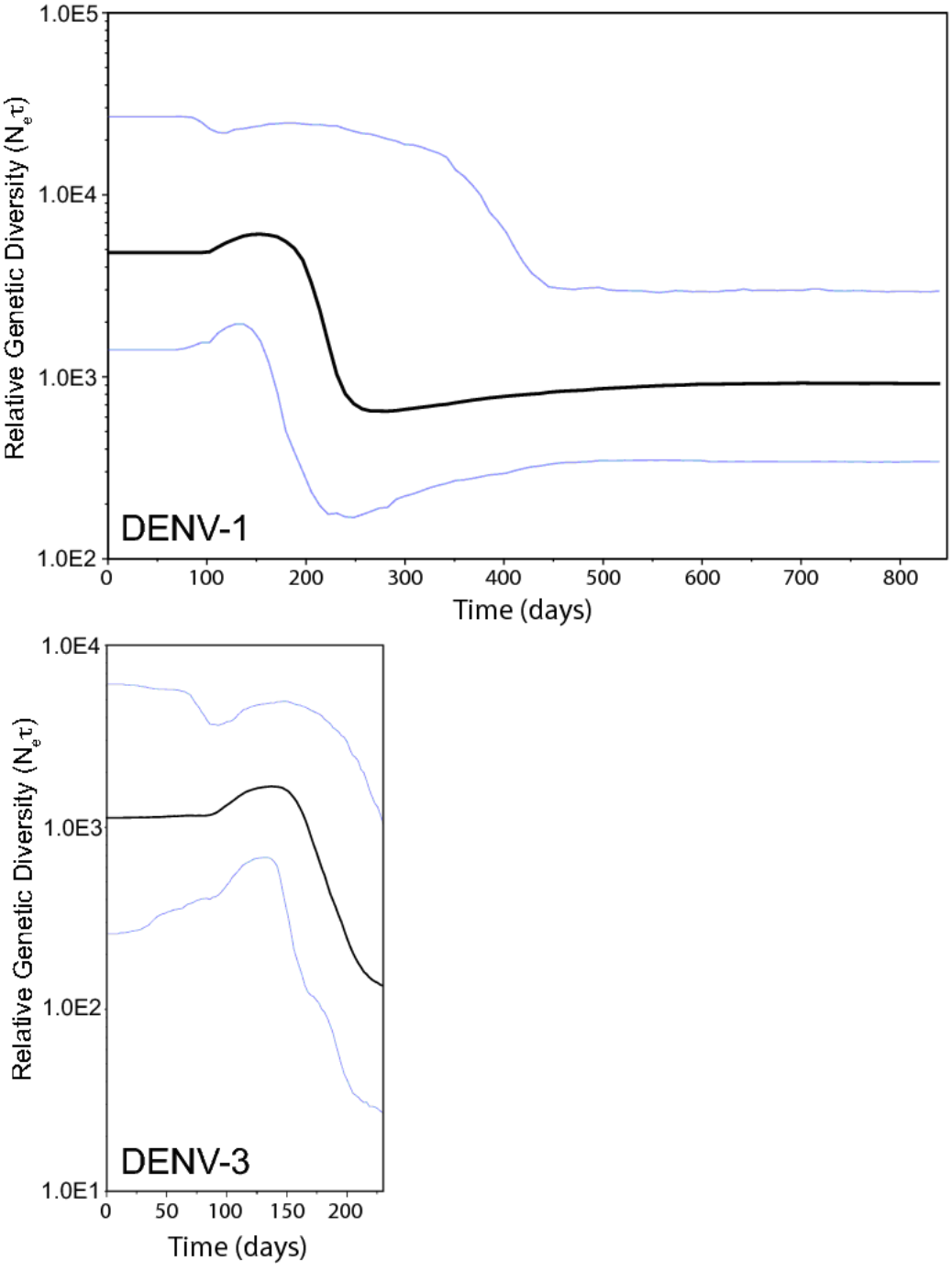


Figure 4.5. Population dynamics of DENV-1 and DENV-3 in Singapore during 2005 depicted using Bayesian skyline plots. The plots show changes in relative genetic diversity, depicted as the effective number of infections ($N_e\tau$), through time. The black line represents the mean estimate of $N_e\tau$, while the 95% HPD intervals are shown in blue. Time is shown as the number of days from the most recent sample. To aid interpretation, DENV-1 and DENV-3 have been shown on the same time axis.

As an additional analysis of evolutionary dynamics we determined the genetic distance (under the ML substitution model) for each pair of DENV-1 and DENV-3 sequences and compared these values to time intervals of sampling (based on day of fever onset) (Figure 4.6). Interestingly, it appears that genetic distances are often higher between isolates separated by shorter periods of time but then decline between pairs sampled over a longer time period (although this analysis does not take into account phylogenetic structure). This may in part be due to the presence of transient deleterious mutations in samples that are only separated by short time-periods (such that genetic distances strongly reflect the background mutation rate), which are later purged by purifying selection, so that longer term genetic distances are more indicative of the population substitution rate (12). This is supported by the observation that d_N/d_S is higher on external (0.130 and 0.157) than internal (0.081 and 0.030) branches of the complete coding region phylogenies for both DENV-1 and DENV-3, respectively, as expected if most nonsynonymous polymorphisms are deleterious (25).

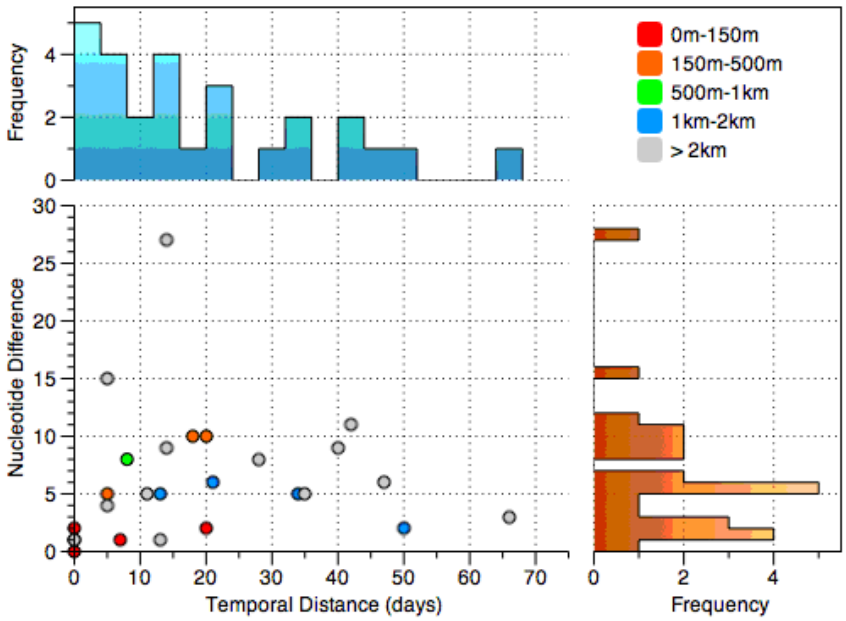


Figure 4.6. Distribution of pairwise nucleotide, temporal and geographic distances among complete coding sequences of DENV-1 and DENV-3 combined. Histograms show the frequency of temporal and nucleotide distances on the X and Y axes respectively. Each pair is represented as a point and the colour of the point indicates the physical distance between the home address of the patients from whom the viruses were isolated.

Interestingly, our analysis of DENV spatial dynamics in Singapore revealed significant population substructure (i.e. the existence of distinct spatial clusters) in both DENV-1 and DENV-3. This spatial dynamic was especially strong in the case of DENV-3 ($p < 0.001$ for both the PS and AI statistics) compared to DENV-1 ($p = 0.008$ and 0.037 for the PA and AI statistics, respectively). Hence, although these viruses were sampled from a relatively restricted region within Singapore, the movement of hosts and/or vectors is sufficiently limited that spatial structure is present in the data.

Finally, the viremia from each patient's sera was estimated from a cross-over threshold (Ct) calculated using quantitative RT-PCR (Q-PCR) at 1 to 3 days and 4 to 7 days post-fever onset. Low C_t value of the RT-PCR, indicating high viremia levels, in the first sampling has been previously shown to be predictive of severe thrombocytopenia in our cohort (31). The C_t value for DENV-1 was significantly lower ($p = 0.002$) than for DENV-3 at both the first (17.07 vs. 19.76) and second (26.87 vs. 29.57) serum samplings, indicating a higher viremia level for DENV-1. DENV-1 also resulted in a significantly ($p = 0.021$) higher ratio of hospitalizations among the sampled population compared to DENV-3 (0.74 vs. 0.50) and may also be reflected in the higher values of $N_e\tau$ for DENV-1 than DENV-3.

Overall, these results suggest that ecological and/or immunological factors, rather than aspects of viral evolution, were central in shaping the dynamics of this dengue outbreak. Most notably, two different serotypes, and multiple co-circulating genotypes, emerged simultaneously, experienced similar epidemiological dynamics, and seemingly spread without the aid of positive selection, accumulating no amino acid fixations. Similarly, it is clear that viral evolution did not succeed in extending the epidemic; despite numerous mutations (Table 4.1), the number of dengue cases declined rapidly in 2006, suggesting that few, if any, of these mutations provided any selective advantage in the face of rising immunity. However, both DENV-1 and DENV-3 appear to possess an inherent robustness that allows them to persist at a low level of infection at times when ecological and immunological conditions do not favour an outbreak. This is supported by the observation

that the DENV-1 and DENV-3 lineages that characterize this outbreak were found in Singapore as early as 2003 but did not result in an outbreak until 2005. However, a wider sampling of viral isolates from neighbouring geographic areas is needed to fully test this analysis. Furthermore, dengue epidemics in Singapore follow a regular 6-7 year periodic cycle, which is difficult to explain by patterns of viral evolution or by vector population density and might be more attributable to changing levels of herd immunity (6). Additionally, in Colombia it has been observed that epidemic years are correlated with an increase in the infection rate of mosquitoes and not the total number of mosquitoes per household (18) indicating surveillance of mosquito populations may also be important in understanding epidemics.

Together our results have important implications for the future study and control of DENV epidemics. In particular, the epidemic surveillance of viral genome sequences in this case would not have been sufficient to predict the 2005 outbreak. Hence, incumbent strains with apparently inherent epidemic potential are required but apparently not sufficient to spark an outbreak. Concurrent surveillance of viral isolates, mosquito vector (including proportion of mosquitoes infected with the dengue virus) and periodic surveys of seroprevalence rates of the population may therefore provide the additional required predictive information. The chance discovery of the DENV-3 outbreak also highlights the value of comprehensive city-wide fever surveys in detecting rare events.

4.5 Acknowledgements

We wish to thank the staff involved in undertaking the EDEN study together with Lai Yee Ling from the Environmental Health Institute, and Fu Xiuju from the Institute for High Performance Computing and Leng Marcel from the National University of Singapore. We thank professors Edison Liu and Duane Gubler for their useful comments on the manuscript.

4.6 Author Contributions

SH Ong assembled the sequences, performed the sequence analysis and reviewed the manuscript. Contributions from co-authors are as listed: conceived and designed the experiments: EEO, SGV, MLH, AO; contributed samples and reagents: JGHL; analyzed patient clinical profiles: YSL; propagated the virus samples: HCT, WL; sequenced the genomes: LT, PPKA; performed the population dynamics analysis: HS, ECH; wrote the manuscript: MJS, ECH; reviewed the manuscript: LCN, EEO, SGV, MLH.

4.7 References

1. Bennett, S. N., E. C. Holmes, M. Chirivella, D. M. Rodriguez, M. Beltran, V. Vorndam, D. J. Gubler, and W. O. McMillan. 2006. Molecular evolution of dengue 2 virus in Puerto Rico: positive selection in the viral envelope accompanies clade reintroduction. *J Gen Virol* 87:885-93.
2. de Oliveira, T., O. G. Pybus, A. Rambaut, M. Salemi, S. Cassol, M. Ciccozzi, G. Rezza, G. C. Gattinara, R. D'Arrigo, M. Amicosante, L. Perrin, V. Colizzi, and C. F. Perno. 2006. Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. *Nature* 444:836-7.
3. Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88.
4. Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
5. Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185-92.
6. Egger, J. R., E. E. Ooi, D. W. Kelly, M. E. Woolhouse, C. R. Davies, and P. G. Coleman. 2008. Reconstructing historical changes in the force of infection of dengue fever in Singapore: implications for surveillance and control. *Bull World Health Organ* 86:187-96.
7. Gilbert, M. T., A. Rambaut, G. Wlasiuk, T. J. Spira, A. E. Pitchenik, and M. Worobey. 2007. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A* 104:18566-70.

8. Goncalvez, A. P., A. A. Escalante, F. H. Pujol, J. E. Ludert, D. Tovar, R. A. Salas, and F. Liprandi. 2002. Diversity and evolution of the envelope gene of dengue virus type 1. *Virology* 303:110-9.
9. Guan, Y., B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li, L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K. Y. Yuen, J. S. Peiris, and L. L. Poon. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302:276-8.
10. Gubler, D. J. 2002. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends Microbiol* 10:100-3.
11. Halstead, S. B. 2007. Dengue. *Lancet* 370:1644-52.
12. Jarman, R. G., E. C. Holmes, P. Rodpradit, C. Klungthong, R. V. Gibbons, A. Nisalak, A. L. Rothman, D. H. Libraty, F. A. Ennis, M. P. Mammen, Jr., and T. P. Endy. 2008. Microevolution of Dengue viruses circulating among primary school children in Kamphaeng Phet, Thailand. *J Virol* 82:5494-500.
13. Johansson, M., A. J. Brooks, D. A. Jans, and S. G. Vasudevan. 2001. A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-beta and the viral helicase, NS3. *J Gen Virol* 82:735-45.
14. King, C. C., D. Y. Chao, L. J. Chien, G. J. Chang, T. H. Lin, Y. C. Wu, and J. H. Huang. 2008. Comparative analysis of full genomic sequences among different genotypes of dengue virus type 3. *Virol J* 5:63.
15. Klungthong, C., C. Zhang, M. P. Mammen, Jr., S. Ubol, and E. C. Holmes. 2004. The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. *Virology* 329:168-79.
16. Lanciotti, R. S., J. T. Roehrig, V. Deubel, J. Smith, M. Parker, K. Steele, B. Crise, K. E. Volpe, M. B. Crabtree, J. H. Scherret, R. A. Hall, J. S. MacKenzie, C. B. Cropp, B. Panigrahy, E. Ostlund, B. Schmitt, M. Malkinson, C. Banet, J. Weissman, N. Komar, H. M. Savage, W. Stone, T. McNamara, and D. J. Gubler. 1999. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* 286:2333-7.

17. Low, J. G., E. E. Ooi, T. Tolfvenstam, Y. S. Leo, M. L. Hibberd, L. C. Ng, Y. L. Lai, G. S. Yap, C. S. Li, S. G. Vasudevan, and A. Ong. 2006. Early Dengue infection and outcome study (EDEN) - study design and preliminary findings. *Ann Acad Med Singapore* 35:783-9.
18. Mendez, F., M. Barreto, J. F. Arias, G. Rengifo, J. Munoz, M. E. Burbano, and B. Parra. 2006. Human and mosquito infections by dengue viruses during and after epidemics in a dengue-endemic region of Colombia. *Am J Trop Med Hyg* 74:678-83.
19. Messer, W. B., D. J. Gubler, E. Harris, K. Sivananthan, and A. M. de Silva. 2003. Emergence and global spread of a dengue serotype 3, subtype III virus. *Emerg Infect Dis* 9:800-9.
20. Ong, S. H., J. T. Yip, Y. L. Chen, W. Liu, S. Harun, E. Lystiyaningsih, B. Heriyanto, C. G. Beckett, W. P. Mitchell, M. L. Hibberd, A. Suwandono, S. G. Vasudevan, and M. J. Schreiber. 2008. Periodic re-emergence of endemic strains with strong epidemic potential-a proposed explanation for the 2004 Indonesian dengue epidemic. *Infect Genet Evol* 8:191-204.
21. Ooi, E. E., K. T. Goh, and D. J. Gubler. 2006. Dengue prevention and 35 years of vector control in Singapore. *Emerg Infect Dis* 12:887-93.
22. Parker, J., A. Rambaut, and O. G. Pybus. 2008. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* 8:239-46.
23. Pond, S. L., and S. D. Frost. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531-3.
24. Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817-8.
25. Pybus, O. G., A. Rambaut, R. Belshaw, R. P. Freckleton, A. J. Drummond, and E. C. Holmes. 2007. Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol* 24:845-52.
26. Rico-Hesse, R., L. M. Harrison, A. Nisalak, D. W. Vaughn, S. Kalayanarooj, S. Green, A. L. Rothman, and F. A. Ennis. 1998. Molecular evolution of dengue type 2 virus in Thailand. *Am J Trop Med Hyg* 58:96-101.

27. Schreiber, M. J., S. H. Ong, R. C. Holland, M. L. Hibberd, S. G. Vasudevan, W. P. Mitchell, and E. C. Holmes. 2007. DengueInfo: A web portal to dengue information resources. *Infect Genet Evol* 7:540-1.
28. Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123:603-13.
29. Suwandono, A., H. Kosasih, Nurhayati, R. Kusriastuti, S. Harun, C. Ma'roef, S. Wuryadi, B. Herianto, D. Yuwono, K. R. Porter, C. G. Beckett, and P. J. Blair. 2006. Four dengue virus serotypes found circulating during an outbreak of dengue fever and dengue haemorrhagic fever in Jakarta, Indonesia, during 2004. *Trans R Soc Trop Med Hyg* 100:855-62.
30. Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. . Sinauer Associates. Sunderland, Mass.
31. Tanner, L., M. Schreiber, J. G. Low, A. Ong, T. Tolfvenstam, Y. L. Lai, L. C. Ng, Y. S. Leo, L. Thi Puong, S. G. Vasudevan, C. P. Simmons, M. L. Hibberd, and E. E. Ooi. 2008. Decision tree algorithms predict the diagnosis and outcome of dengue Fever in the early phase of illness. *PLoS Negl Trop Dis* 2:e196.
32. Twiddy, S. S., E. C. Holmes, and A. Rambaut. 2003. Inferring the rate and time-scale of dengue virus evolution. *Mol Biol Evol* 20:122-9.
33. Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-91.
34. Yap, T. L., T. Xu, Y. L. Chen, H. Malet, M. P. Egloff, B. Canard, S. G. Vasudevan, and J. Lescar. 2007. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J Virol* 81:4753-65.

5. Inferring the History of Dengue Virus Diversity in Malaysia from Sequence Data

Swee Hoe Ong[1,2,3], Pauline P.K. Aw[3], Jin Teen Yip[1], Jasmin Schmid[1], Wei Liu[1], Siew Gyan Khoo[5], Subhash G. Vasudevan[1†], Pei Yong Shi[1], Martin L. Hibberd[3], Shamala Devi[4], Mark J. Schreiber[1]

1. Novartis Institute for Tropical Diseases, 10 Biopolis Road, Chromos #05-01, Singapore 138670.
2. Swiss Tropical Institute, Socinstrasse 57, P.O. Box CH-4002 Basel, Switzerland.
3. Genome Institute of Singapore, A*STAR, 60 Biopolis Street, Genome #02-01, Singapore 138672.
4. Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Malaysia.
5. School of Computer Engineering, Block N4, Nanyang Technological University, Nanyang Avenue, Singapore 639798.

† Current address: Emerging Infectious Diseases Program, Duke-NUS Graduate Medical School, 2 Jalan Bukit Merah, Singapore 169547.

Manuscript in preparation.

5.1 Abstract

To understand the molecular epidemiology of dengue viruses in Malaysia, the complete genomes of 46 clinical isolates sampled over an 11 year period between 1995 and 2005 were sequenced. High sequence diversity was found with viruses from all four dengue serotypes and various subtypes detected. In addition, all publicly-available sub-genomic length sequences from Malaysian isolates were used to construct a snapshot of dengue virus diversity in the country since 1968. This study suggests that Malaysia, situated at the center of Southeast Asia which is often thought to be the source of dengue virus diversity, has experienced several episodes of importation in addition to maintaining several lineages of viruses that have been in sustained transmission for several decades.

5.2 Introduction

Dengue has become the most important arboviral disease in the world, affecting an estimated 50-100 million people each year (WHO, 2009). It is transmitted by the mosquito vectors *Aedes aegypti* and *Aedes albopictus* which are widely found in most tropical and subtropical countries. The health burden of dengue is particularly heavy in Southeast Asian countries where epidemic DHF was first recognised in the 1950s (Quinlos *et al*., 1954; Hammond *et al*., 1960). Since that time the countries have had a long and sustained history of dengue infections (Gubler & Clark, 1995). A lack of antiviral drugs or vaccines means there is little that can be done, apart from vector control measures, to stem the upward trend in the incidence rate of dengue.

Dengue disease is caused by any one of the four closely-related but antigenically-distinct dengue viruses (DENV) denoted as DENV-1, DENV-2, DENV-3 and DENV-4. The genome of the dengue virus is a single-stranded, positive-sense RNA molecule approximately 10,700 nucleotides in length that encodes three structural proteins (C, capsid; prM, precursor to membrane; E, envelope) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). Accumulated molecular data indicate there is great diversity

of dengue viruses. Each of the four serotypes can be further divided into distinct subtypes (also called genotypes) based on their nucleotide sequence. Phylogenetic analysis of the viral nucleotide sequences often allows the origin of virus isolates to be identified and their transmission across time and space to be tracked.

Malaysia is a tropical country situated in the middle of Southeast Asia, bordered by Thailand to the north and by Singapore and Indonesia to the south. Classical dengue fever cases have been reported in Malaysia since 1901-1902 (Skae, 1902) and laboratory-confirmed DHF was first reported in November 1962 in its northern state of Penang (Rudnick *et al*., 1965). Dengue was made a notifiable disease in 1973 and the number of reported cases has been on the increase ever since (Lam, 1994). Short sequences from Malaysian DENV isolates have been routinely included in phylogeny studies (Rico-Hesse *et al*., 1990; Lewis *et al*., 1993; Lanciotti *et al*., 1994, Wang *et al*., 2000). However, despite Malaysia's long association with dengue, relatively little is known about the diversity and molecular epidemiology of the etiological agent circulating in the country.

The few previous studies of dengue molecular epidemiology in Malaysia have multifariously involved sequencing either the 240-nucleotide region spanning the E/NS1 junction (Fong *et al*., 1998; Chee and AbuBakar, 2003), a 1050-nucleotide region spanning the complete C, prM and part of E (Kobayashi *et al*., 1999) or the complete E gene sequence (AbuBakar *et al*., 2002). Three complete genomes of sylvatic DENV from the 1970s have been made available recently (Vasilakis *et al*., 2007) but no complete genome from post-1970s urban dengue outbreaks have been reported thus far.

In order to fill the gap in knowledge and to gain a better understanding of DENV molecular epidemiology in Malaysia, we sequenced the complete genomes of virus samples collected in the Malaysian capital city of Kuala Lumpur over an 11-year period. Using the complete genome sequences as signposts we devised a routine that employs all available partial sequences of Malaysian isolates for phylogeny inference. The results were then assembled to produce a snapshot of DENV diversity in Malaysia since 1968.

5.3 Material and Methods

5.3.1 Sample inclusion/collection criteria

Blood samples were obtained from clinically diagnosed dengue patients admitted to the University Malaya Medical Centre (UMMC), Kuala Lumpur from 1995-2005. These were classified as dengue fever (DF) or dengue haemorrhagic fever (DHF) based on the WHO dengue classification scheme and case definition (WHO, 1997). The serum samples collected were subjected to an in-house IgM capture ELISA for IgM antibodies (Lam *et al*., 1987). Samples that were IgM negative were then subjected to virus isolation and a second sample was requested. All serum samples were stored at -80°C until use.

5.3.2 Virus isolation and serotyping

Dengue viruses were cultured in the C6/36 *Ae. albopictus* cell line (Igarashi *et al*., 1982). The C6/36 cell layer was grown in a 25 cm^3 angle-neck tissue culture flask inoculated with 100 µl of patient serum and kept at 37°C for 1 hour to allow for virus adsorption. Flasks were then topped up with 5 ml of maintenance media (RPMI containing 2% FCS) and incubated at 28°C for 10 days after which the infected cells were harvested and fixed onto Teflon-coated slides. The presence of virus was determined by indirect immunofluorescence antibody test (IFAT) using dengue-specific monoclonal antibodies (DENV-1: MAB D2-1F1-3; DENV-2: MAB 3H2-1-21; DENV-3: MAB D6-8A1-12; and DENV 4: MAB 1410-6-7, kindly provided by the Centers for Disease Control and Prevention, Fort Collins, CO, USA). Viral antigen in fixed infected cells was visualised using an UV microscope (Olympus).

5.3.3 Complete genome sequencing

Complete genome sequencing was performed as previously described (Ong *et al*., 2008). Briefly, viral RNA was extracted using the QIAamp Viral RNA Mini Kit (QIAGEN) and cDNA templates were generated from viral RNA by reverse transcription. Five serotype-specific primer pairs were then used to produce five slightly-overlapping PCR fragments from the cDNA. The PCR

fragments were sequenced by capillary sequencing using a 3730xl DNA Analyzer (Applied Biosystems). The obtained sequences were assembled using SeqScape version 2.5 (Applied Biosystems) and manually edited for quality before being aligned to other DENV complete genomes.

5.3.4 Phylogeny inference of sequenced complete genomes

Complete DENV genome sequences annotated with country and year of isolation were obtained from the DengueInfo database (Schreiber *et al*., 2007). Distance-based neighbour-joining phylogenetic trees were generated from all available complete genome sequences for each of the four DENV serotypes using the Minimum Evolution (ME) tree inference method (Rzhetsky and Nei, 1993) as implemented in MEGA version 4.0 (Tamura *et al*., 2007). The nucleotide substitution model used was the Tamura-Nei model (Tamura and Nei, 1993) and the obtained tree topology was tested using 1000 bootstrap replications. Based on the generated ME trees, highly-similar complete genome sequences in overly dense branches were removed and the pruned data set used to produce the final set of trees using both Minimum Evolution and Maximum Likelihood (ML) methods. Pruning was not required for DENV-4 due to the limited number of available genome sequences.

The ML phylogenetic trees were constructed using the maximum likelihood (ML) method implemented in PAUP* (Swofford, 2002), with the best-fit model of nucleotide substitution selected by Akaike Information Criterion (AIC) as implemented in ModelTest (Posada and Crandall, 1998). Branch topology was verified by generating 1000 neighbour-joining bootstraps using the ML substitution model and the scores on tree nodes represent the number of bootstrap replicates (presented in percentage) supporting each node. The length of the tree branches is proportional to the number of nucleotide changes.

The DENV subtype classification scheme follows that of Goncalvez *et al*. (2002) for DENV-1, Twiddy *et al*. (2002) for DENV-2, Lanciotti *et al*. (1994) for DENV-3 and Lanciotti *et al*. (1997) for DENV-4.

5.3.5 Search for sub-genomic length sequences

DENV sub-genomic length sequences were retrieved on April 15, 2009 from the National Center for Biotechnology Information (NCBI) GenBank database (http://www.ncbi.nlm.nih.gov/) using the query string *txid11052[orgn] AND 1[SLEN]:9999[SLEN]*. The number 11052 is the taxonomy ID for the Dengue Virus Group as defined by the NCBI Taxonomy Browser; the balance of the query string specified that sequences with length between 1 and 9,999 nucleotides be retrieved. Sequences originating from virus samples collected in Malaysia and their isolation year were identified by parsing the respective GenBank records and/or from literature search.

5.3.6 Subtype classification for sub-genomic length sequences

A multiple sequence alignment of the Malaysian sub-genomic length sequences was generated to identify the most sequenced regions. A procedure was then devised based on the sequence coverage in the multiple sequence alignment to rationally apportion all sub-genomic length sequences for subtype classification. Sequences were divided into discrete sets starting with those that have the highest coverage, i.e. the complete E gene and followed by the 240-nucleotide E/NS1 junction. Lastly, the remaining analogous sequences of identical length were divided into yet more discrete sets. This process was repeated for each serotype. Analogous sequences extracted from complete genomes were then added to the resulting sets of sub-genomic length sequences (Table 5.2).

To avoid discrepancies introduced through the use of different tree construction methods, phylogenetic trees were constructed from each of the sets using the Minimum Evolution method as described in section 5.3.4. The subtype of the sub-genomic length sequences was determined based on their respective clustering location in the trees, and the consistency in subtype classification of the completely sequenced isolates was also scrutinised. Finally, the subtype and year of isolation information of all DENV sequences from Malaysia were combined to produce a snapshot of the history of virus diversity in the country.

5.4 Results

5.4.1 Complete genome sequencing of clinical isolates from Malaysia

A total of 46 DENV complete genomes were obtained with all four dengue serotypes represented, producing a data set with 10 DENV-1, 5 DENV-2, 23 DENV-3 and 8 DENV-4 sequences (Table 5.1). Multiple subtypes were observed for DENV-1 and DENV-3 whereas only single subtypes were observed for DENV-2 and DENV-4, although this may be due to the limited number of samples. The obtained sequences have been deposited in GenBank with the accession numbers FN429881–FN429926.

Table 5.1. The GenBank accession number, isolate name, serotype, year of sample collection, subtype classification and some patient-related information of the 46 dengue virus genomes sequenced in this study.

Accession	Isolate Name	Serotype	Year	Subtype	Severity	Patient Age	Patient Sex
FN429881	D1MY95-3891	DENV-1	1995	IV	DF	6	M
FN429882	D1MY95-3928	DENV-1	1995	IV	DF	40	M
FN429883	D1MY96-7814	DENV-1	1996	IV	DF	21	F
FN429884	D1MY96-7891	DENV-1	1996	IV	DF	21	F
FN429885	D1MY96-7905	DENV-1	1996	IV	DF	22	F
FN429886	D1MY96-8064	DENV-1	1996	IV	DF	-	F
FN429887	D1MY96-8080	DENV-1	1996	IV	DF	27	F
FN429888	D1MY96-8081	DENV-1	1996	IV	DF	19	F
FN429889	D1MY97-10245	DENV-1	1997	IV	DF	27	M
FN429890	D1MY05-33915	DENV-1	2005	I	DF	5	F
FN429891	D2MY97-10340	DENV-2	1997	Cosmopolitan	DF	27	F
FN429892	D2MY00-22563	DENV-2	2000	Cosmopolitan	DF	37	F
FN429893	D2MY04-32618	DENV-2	2004	Cosmopolitan	DF	22	F
FN429894	D2MY04-32883	DENV-2	2004	Cosmopolitan	DF	21	M
FN429895	D2MY04-33054	DENV-2	2004	Cosmopolitan	DF	24	M
FN429896	D3MY95-2471	DENV-3	1995	II	DF	10	F
FN429897	D3MY95-3952	DENV-3	1995	II	DF	22	F
FN429898	D3MY96-4269	DENV-3	1996	II	DF	24	M
FN429899	D3MY97-12440	DENV-3	1997	I	DF	1	M
FN429900	D3MY99-21531	DENV-3	1999	I	DF	5	F

FN429901	D3MY00-22447	DENV-3	2000	I	DHF	-	M
FN429902	D3MY00-22460	DENV-3	2000	I	DF	7	M
FN429903	D3MY00-22550	DENV-3	2000	I	DF	32	M
FN429904	D3MY00-22583	DENV-3	2000	I	DF	30	F
FN429905	D3MY00-22366	DENV-3	2000	II	DHF	7	M
FN429906	D3MY01-22939	DENV-3	2001	I	DF	35	M
FN429907	D3MY01-24056	DENV-3	2001	III	DF	-	M
FN429908	D3MY02-25811	DENV-3	2002	II	DF	20	M
FN429909	D3MY02-25850	DENV-3	2002	II	DF	21	F
FN429910	D3MY03-27834	DENV-3	2003	I	DF	16	M
FN429911	D3MY03-28526	DENV-3	2003	I	DF	28	F
FN429912	D3MY04-32645	DENV-3	2004	I	DF	25	F
FN429913	D3MY04-33077	DENV-3	2004	II	DF	57	F
FN429914	D3MY05-33464	DENV-3	2005	I	DF	11	F
FN429915	D3MY05-33506	DENV-3	2005	I	DF	6	M
FN429916	D3MY05-33610	DENV-3	2005	I	DF	32	M
FN429917	D3MY05-33927	DENV-3	2005	I	DF	12	F
FN429918	D3MY05-34640	DENV-3	2005	I	DF	60	M
FN429919	D4MY95-328	DENV-4	1995	II	DF	20	F
FN429920	D4MY01-22713	DENV-4	2001	II	DF	42	M
FN429921	D4MY01-23476	DENV-4	2001	II	DF	26	M
FN429922	D4MY02-26658	DENV-4	2002	II	DF	49	F
FN429923	D4MY03-27498	DENV-4	2003	II	DF	1	M
FN429924	D4MY03-27949	DENV-4	2003	II	DF	40	M
FN429925	D4MY04-31586	DENV-4	2004	II	DF	3	M
FN429926	D4MY04-32729	DENV-4	2004	II	DF	15	F

Abbreviations: DF: dengue fever and DHF: dengue haemorrhagic fever according to WHO guidelines.

5.4.2 Phylogeny of DENV complete genomes between 1995-2005

Nine of the ten sequenced DENV-1 genomes, collected between 1995 and 1997, form a distinct clade within subtype IV (Figure 5.1). This suggests that an unbroken chain of transmission by this viral lineage was partly responsible for the dengue incidences in Malaysia during that 3-year period. The tenth DENV-1 sample was collected in 2005 and its sequence clusters near a subtype I virus collected at the tail end of the 2005 epidemic in Singapore described by Schreiber *et al*. (2009). This could either be evidence of an importation event from Singapore into Malaysia, or the same subtype of DENV-1 could have been circulating in both countries in 2005. However, since its isolation date is not known the exact role of this Malaysian isolate in the Singapore epidemic cannot be ascertained. Until more DENV-1 isolates from Kuala Lumpur during and immediately before 2005 become available, it will not be possible to determine the relatedness of viruses in Kuala Lumpur and Singapore in 2005. Nevertheless, it can be concluded that subtype I and IV DENV-1 form part of the local reservoir of dengue viruses in the region due to the fact that the Malaysian isolates in both clades are closely related to isolates from nearby countries like Singapore, Indonesia, and Brunei and the more distant region of southern China.

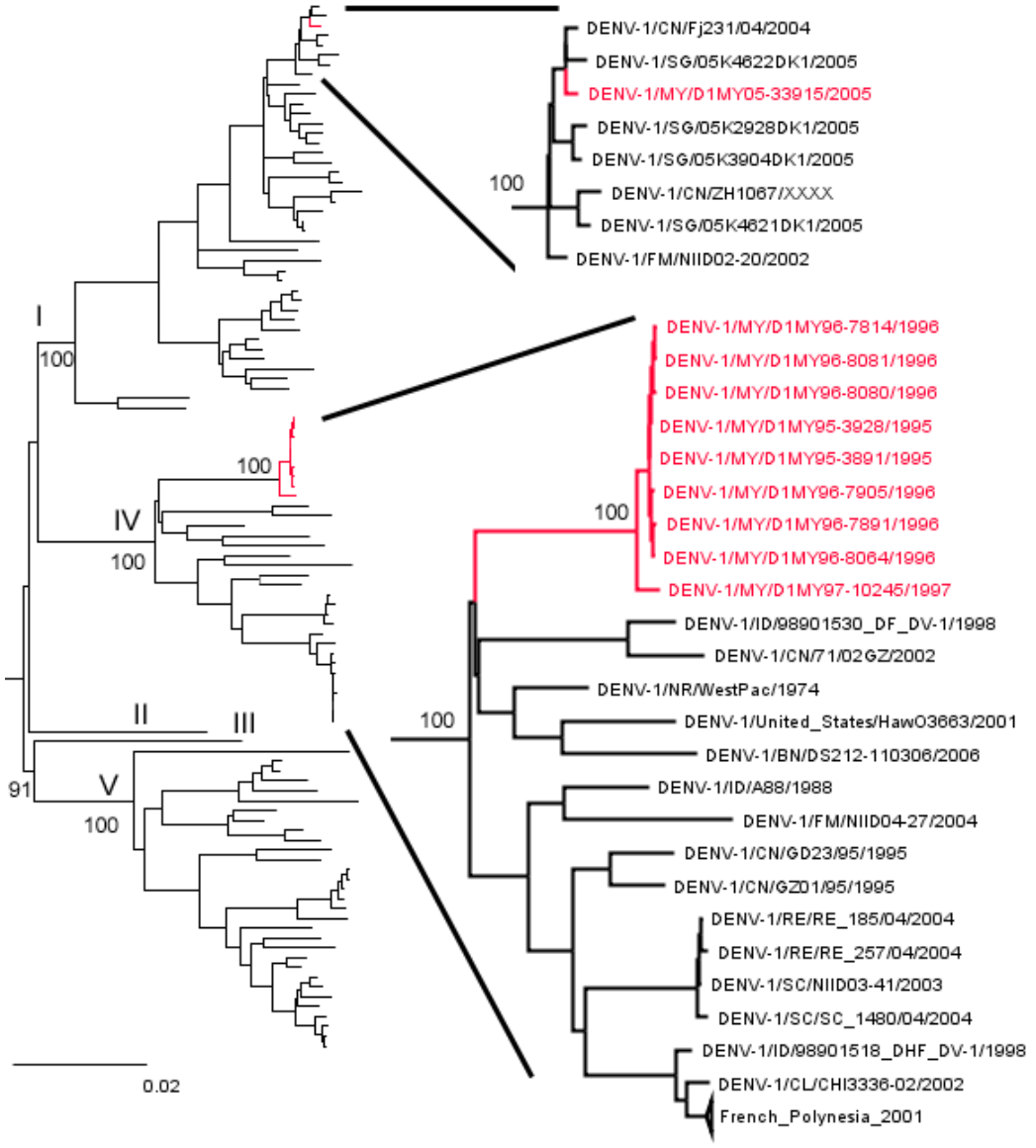


Figure 5.1. Minimum evolution (ME) phylogenetic tree built from 107 full-length DENV-1 genome sequences. The DENV-1 subtypes are indicated on the main branches and the 10 Malaysian clinical isolates sequenced in this study are shown in red. The tree is mid-point root for clarity only, all horizontal branch lengths are drawn to a scale of nucleotide substitutions per site, and bootstrap values are shown for key nodes.

All 5 DENV-2 genome sequences obtained belong to the Cosmopolitan subtype and they also cluster with recent isolates from neighbouring countries such as Indonesia, Singapore and Brunei (Figure 5.2). The distribution of these isolates within the Cosmopolitan subtype tree branch strongly suggests that this subtype is prevalent in those Southeast Asian countries and points to frequent transmissions across national borders.

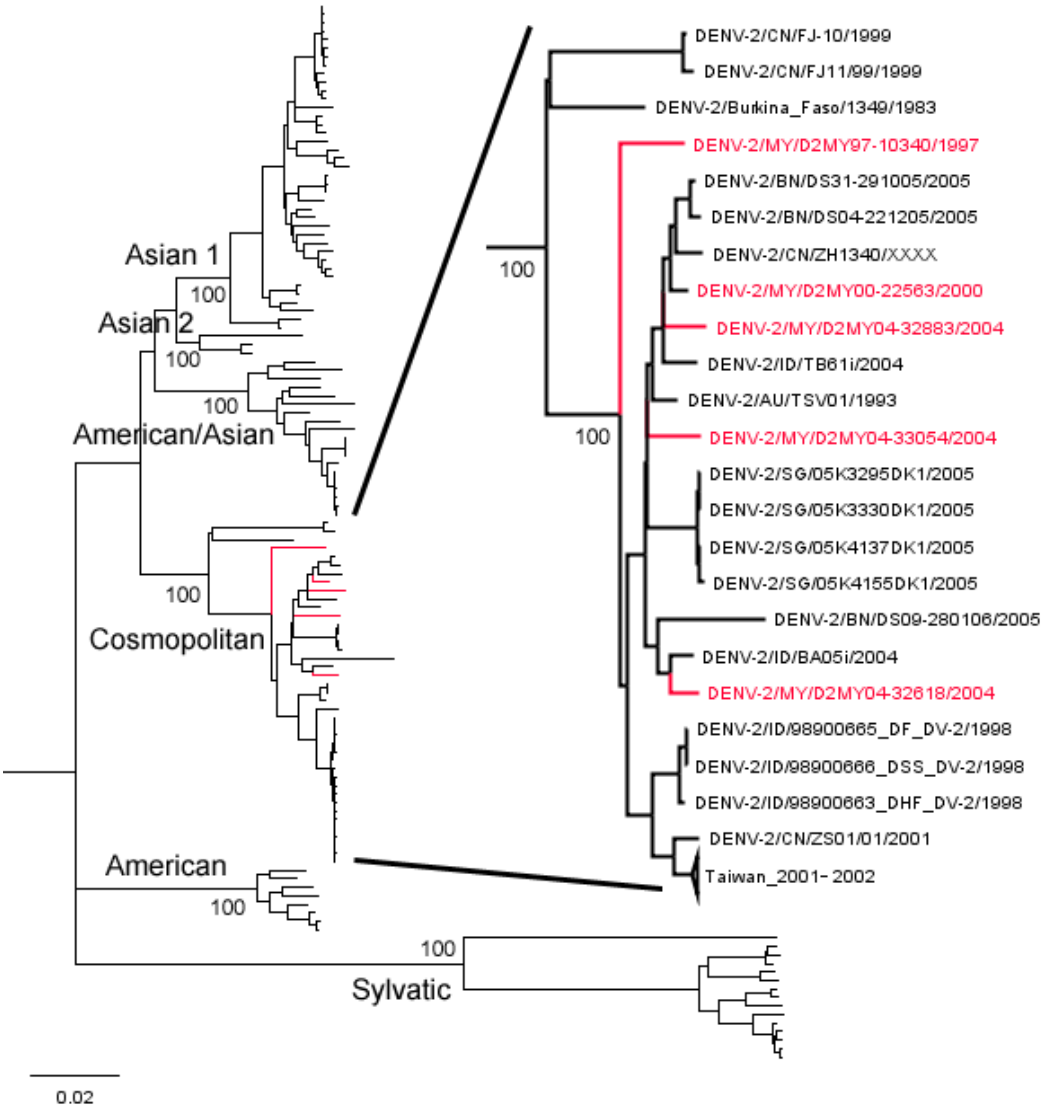
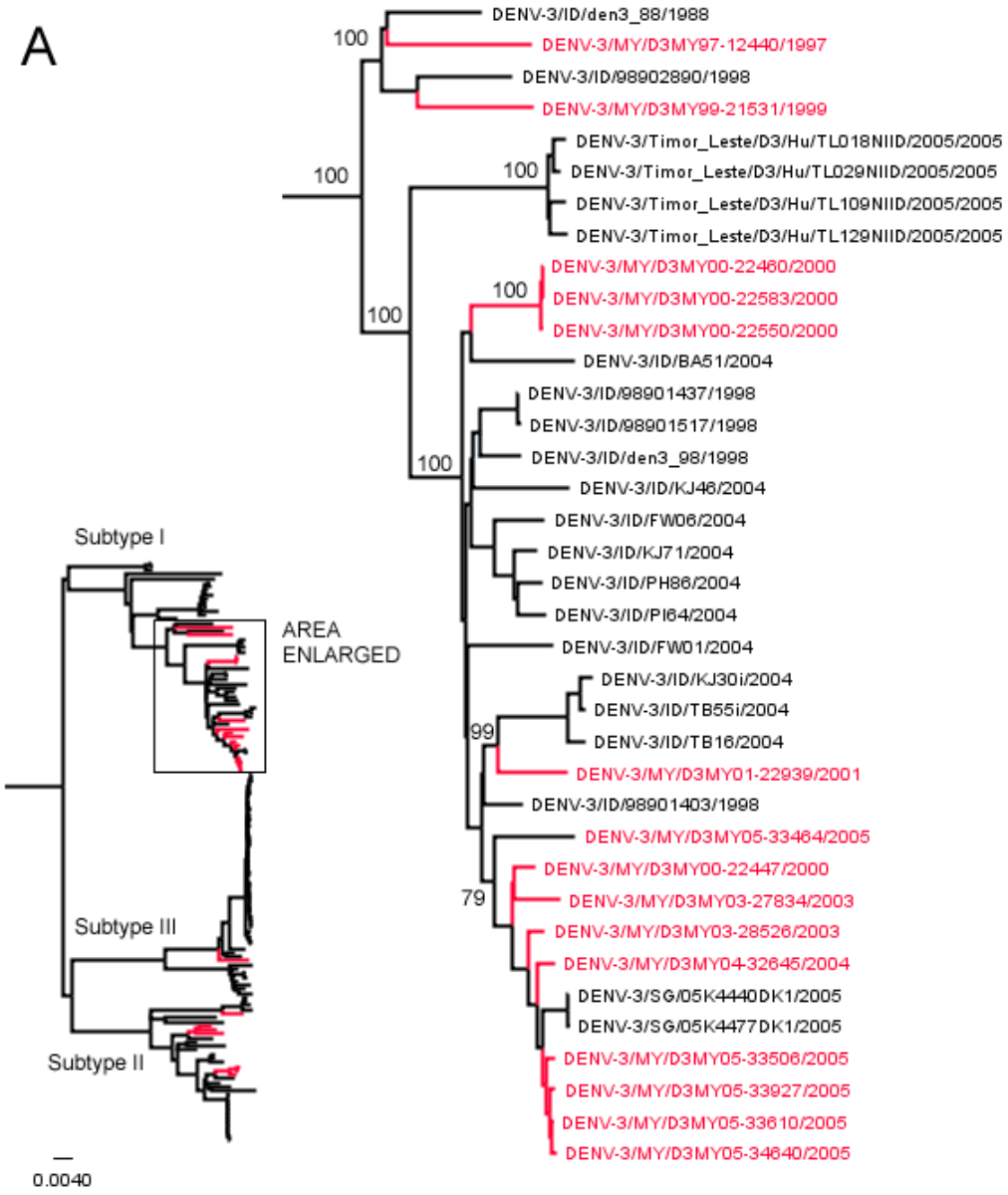
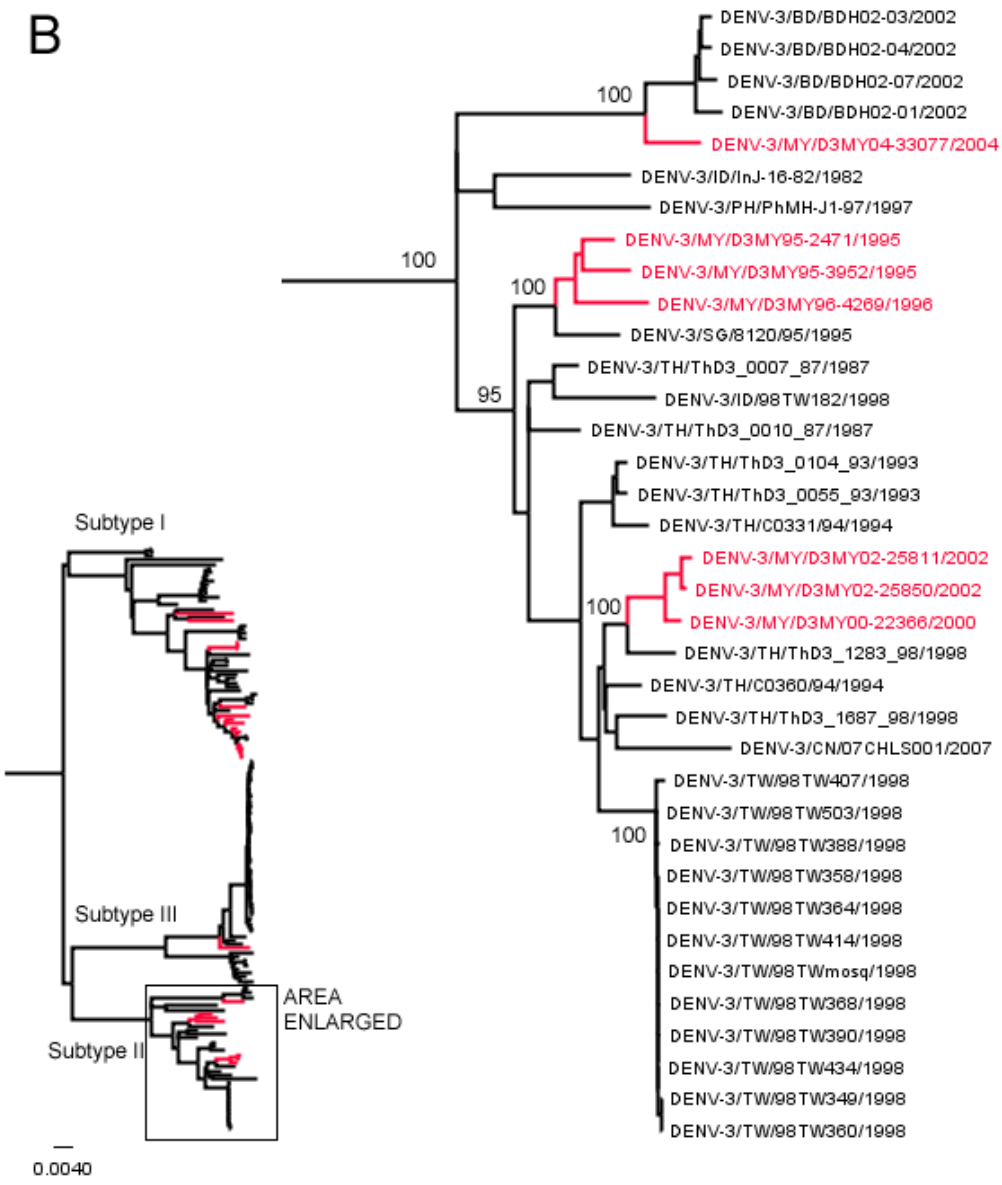


Figure 5.2. Minimum evolution (ME) phylogenetic tree built from 125 full-length DENV-2 genome sequences. The DENV-2 subtypes are indicated on the main branches and the 5 Malaysian clinical isolates sequenced in this study are shown in red. The tree is mid-point root for clarity only, all horizontal branch lengths are drawn to a scale of nucleotide substitutions per site, and bootstrap values are shown for key nodes.

DENV-3 forms the bulk of the genomes sequenced in this study. With 23 isolates belonging to three different subtypes DENV-3 is also the serotype exhibiting the greatest genetic diversity. The distribution of the subtype I isolates suggest great degree of intermixing with isolates from Indonesia and Singapore (Figure 5.3A). The subtype II isolates showed the same pattern, clustering closely with isolates from neighbouring Singapore, Indonesia and Thailand (Figure 5.3B).





The solitary subtype III isolate clustered to the base of the clade that led to the DENV-3 isolates from the 2005 epidemic in Singapore (Figure 5.3C). From the branching topology it is most likely that the Malaysian isolate is not a direct ancestor of the Singapore outbreak. Two other isolates, from Taiwan and Sri Lanka respectively, are genetically closer to the Singapore isolates. If all isolates in this clade were from autochthonous dengue infections, then it is most likely subtype III DENV-3 has established a strong foothold in SEA and all the way to southern China and Taiwan since its high epidemic potential was first reported in Sri Lanka (Messer *et al*., 2003).

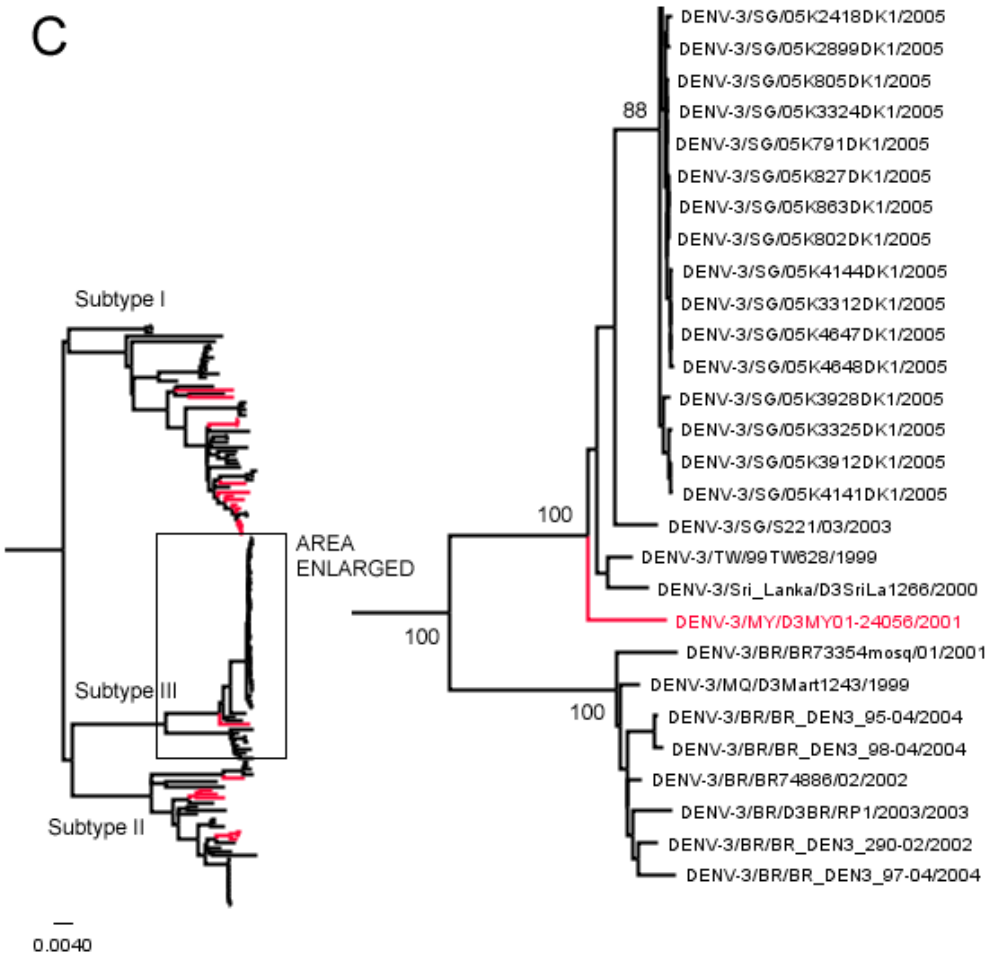


Figure 5.3. Minimum evolution (ME) phylogenetic tree built from 143 full-length DENV-3 genome sequences. The DENV-3 subtypes are indicated on the main branches and the 23 Malaysian clinical isolates sequenced in this study are shown in red. The tree is mid-point root for clarity only, all horizontal branch lengths are drawn to a scale of nucleotide substitutions per site, and bootstrap values are shown for key nodes.

The eight DENV-4 genomes obtained in this study added considerable new information to the relatively small pool of publicly-available DENV-4 complete genome sequences. The eight isolates all belong to subtype II (Figure 5.4) with the older 1995 isolate clustering with an isolate from Singapore of the same year at the base of a clade of viruses from the Americas. The other seven isolates fell into the Southeast Asian clade of subtype II viruses where post-2000 isolates from Malaysia, Singapore,

Indonesia and Thailand all clustered together. This clade likely represents a lineage that is currently endemic in Southeast Asia.

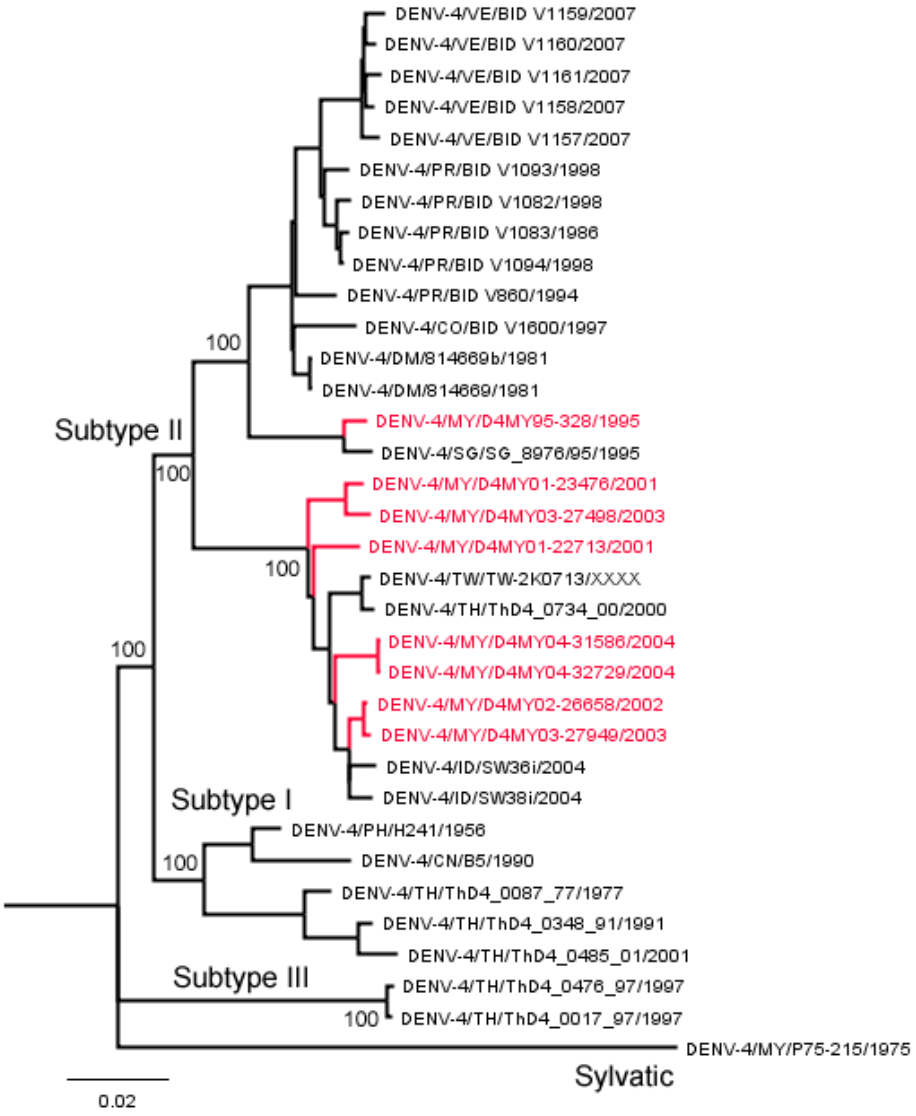


Figure 5.4. Minimum evolution (ME) phylogenetic tree built from 34 full-length DENV-4 genome sequences. The DENV-4 subtypes are indicated on the main branches and the 8 Malaysian clinical isolates sequenced in this study are shown in red. The tree is mid-point root for clarity only, all horizontal branch lengths are drawn to a scale of nucleotide substitutions per site, and bootstrap values are shown for key nodes.

Apart from differences in branch lengths, no incongruence was detected between the minimum evolution tree (Figure 5.4) and the maximum likelihood (ML) tree generated from the DENV-4 complete genome sequences (Figure 5.5). The same observation was seen for the trees of the other three dengue serotypes. This suggests that the computationally less intensive ME method is a viable alternative for the ML method for phylogenetic inference when using the complete genome sequences of DENV.

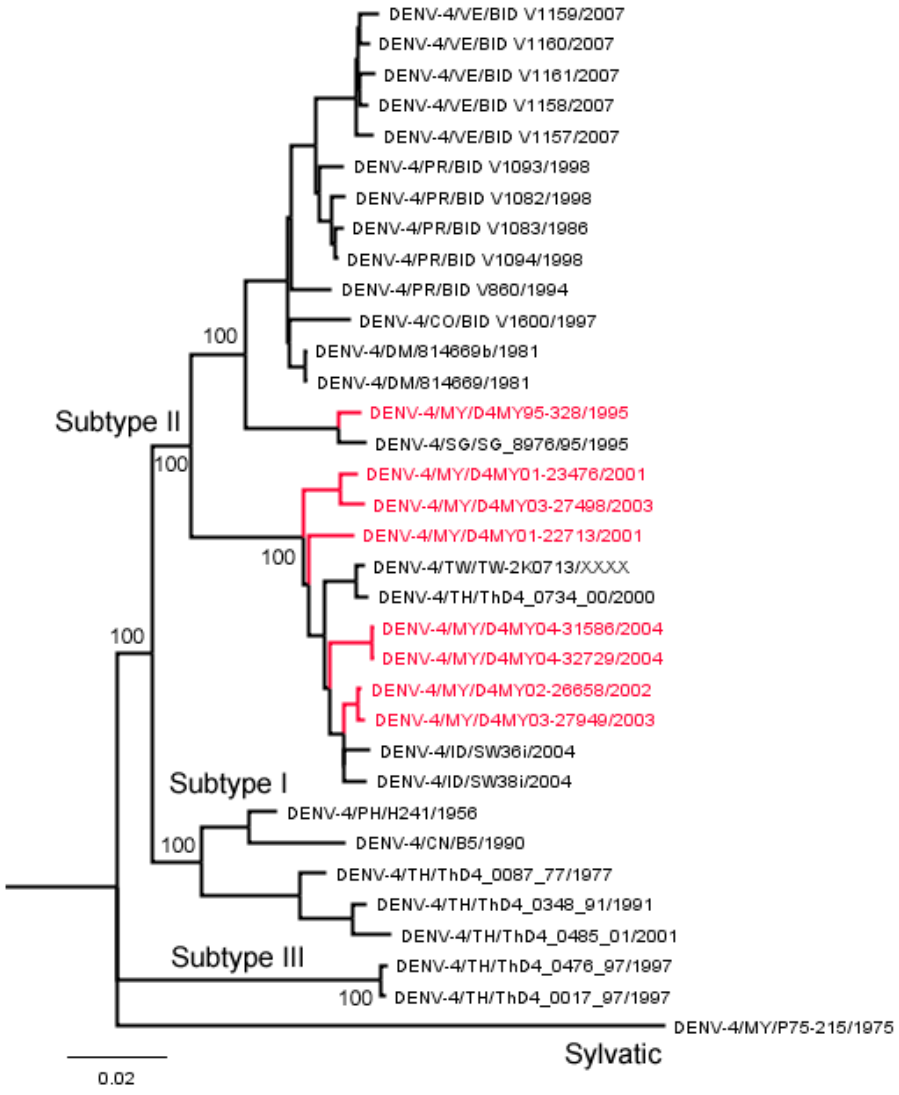


Figure 5.5. Maximum likelihood (ML) phylogenetic tree built from 34 full-length DENV-4 genome sequences. The DENV-4 subtypes are indicated on the main branches and the 8 Malaysian clinical isolates sequenced in this study are shown in red. The tree is mid-point root for clarity only, all horizontal branch lengths are drawn to a scale of nucleotide substitutions per site, and bootstrap values are shown for key nodes.

5.4.3 Alignment of sub-genomic length sequences

From a total of 5165 DENV sequences of sub-genomic length obtained from GenBank, 88 were identified as originating from samples collected in Malaysia from either the accompanying meta data or from literature search: 10 DENV-1, 34 DENV-2, 34 DENV-3 and 10 DENV-4 sequences. In addition, from the available annotation 12 of the DENV-2 sequences were found to originate from the same set of 3 isolates (X15434, X17338, X51708 and X51711 from isolate M1; X15433, X17339, X51709 and X51712 from isolate M2; X15214, X17340, X51710 and X51713 from isolate M3) and were therefore combined into 3 contiguous sequences. DENV-1 sequences with accession numbers AF425622 and AF426121 were also from the same isolate (P72-1244) but these were not merged because they are not contiguous.

An additional 24 DENV-2 E/NS1 junction sequences (Fong *et al*., 1998) and 5 DENV-1 complete E sequences (A-Nuegoonpipat *et al*., 2004) from Malaysian isolates were found described in the literature but these sequences were not submitted to GenBank and thus have no Accession number. These short sequences were obtained from the respective labs and included in this analysis.

From the alignment of the resulting 108 sequences to a complete dengue genome scaffold it was evident the great majority of the sub-genomic length sequences were from either the three structural genes or the E/NS1 junction (Figure 5.6). The complete E gene and the E/NS1 junction were found to be the two most sequenced regions of the DENV genome. The rare exceptions were four DENV-3 NS3 sequences and one DENV-1 NS2A-NS2B sequence. As often the case, DENV-2 is the most studied dengue serotype with the most number of partial sequences available.

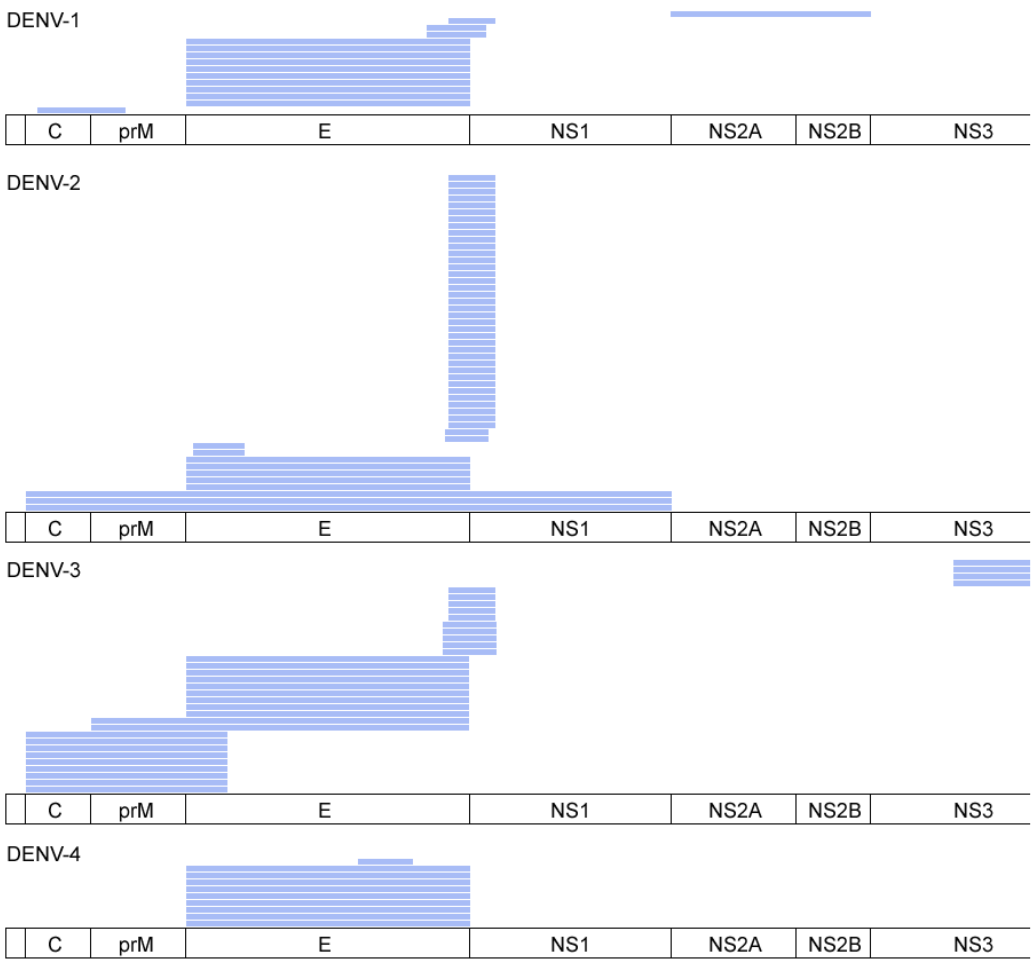


Figure 5.6. Alignment of 108 dengue virus sub-genomic length sequences originating from samples collected in Malaysia.

5.4.4 Phylogeny inference of sub-genomic length sequences

The 108 non-redundant DENV sub-genomic length sequences from Malaysian isolates were apportioned into 14 discrete sets (Table 5.2) according to the protocol described in 5.3.6. After combining the sub-genomic length sequences with the respective analogous sub-sequences from the complete genomes, 14 phylogenetic trees were constructed from the resulting sequence data sets.

Table 5.2. Apportionment of sub-genomic length DENV sequences from identified Malaysian isolates into discrete sets, sorted by serotype and genomic region. Analogous sub-sequences from complete genomes were then added to these sequences for phylogeny inference.

Serotype	Gene Region	Length	Total	Accession Numbers
DENV-1	Complete E	1485	11	AB111067, AF425622, EY448395, EY448400 & EY448410 & set of 6[1]
DENV-1	E/NS1 junction	240	1	M32898
DENV-1	C/partial prM	454	1	EU005249
DENV-1	E/NS1	306	2	DQ016630 & DQ016642
DENV-2	Complete E	1485	8	AF231715, AF231716, AF231717, X51708, X51709, X51710, Q518635 & DQ518636
DENV-2	E/NS1 junction	240	37	AJ556803, AJ556804, AJ556805, AJ556806, AJ556807, AJ556808, AJ556809, AJ556810, AJ556811, AJ556812, AJ556813, AJ556814, AF400004 & set of 24[2]
DENV-2	Partial E	266	2	AF435487 & AF435488
DENV-2	E/NS1 junction	225	2	AF436114 & AF436115
DENV-3	Complete E	1479	11	L11427, L11429, AF147457, AF147458, AF147459, AF147460, AY338492, AY338493, AY338494, EU448438 & EU448439
DENV-3	E/NS1 junction	240	10	AF029794, AF029795, AF029796, AF029797, AF029798, AF400003, AF400021, AF400022, AF400023 & AF400024

DENV-3	C/prM/partial E	1050	9	AB010982, AB010983, AB010984, AB010985, AB010986, AB010987, AB010988, AB010989 & AB010990
DENV-3	Partial NS3	396	4	U11669, U11670, U11671 & U11672
DENV-4	Complete E	1485	9	AF231722, AF231723, AF231724, AF231725, AJ428556, AJ428557, AJ428558, AJ428559 & AJ428560
DENV-4	Partial NS1	282	1	AF051110

Reference: [1] A-Nuegoonpipat *et al*., 2004 and [2] Fong *et al*., 1998.

The 14 phylogenetic trees were characterised by low bootstrap values at most branches. In general, the number of well-supported branches increased proportionately with the length of the sequences used for tree building. As an example, among the four DENV-2 trees listed in Table 5.2 the tree constructed from the 1485-nucleotide complete E gene sequences had the highest number of well-supported branches (Supplementary Figure 1). This underscores the importance of using longer sequences when utilising distance-based methods in phylogeny inference.

The relative clustering positions of the five DENV-2 isolates sequenced in this study vary greatly in the four DENV-2 sub-genomic length sequence trees (Supplementary Figure 1). Using these five as signposts, it is clear that there is substantial incongruence between the relatively-popular complete E gene tree and the E/NS1 junction tree. This implies that any interpretation from these trees, particularly of leaves near terminal branches, must be used with caution. However, apart from two DENV-3 sequences (AF400003 and AF400022) all other isolates were consistently classified at the subtype level in the 14 trees, which means the phylogenetic trees constructed from these sub-genomic length sequences can be reliably used to catalogue the presence of the various DENV subtypes for a geographical area.

5.4.5 Inferring the history of DENV diversity in Malaysia

The 46 complete DENV genomes sequenced in this study were from virus samples circulating in Malaysia during the 11-year period between 1995 and 2004. This data set contributed a total of 24 data points (boxes coloured in blue) to Figure 5.7 and sequences of all four dengue serotypes were represented. Only one subtype each of DENV-4 and DENV-2 was detected in the period, i.e. subtype II and the Cosmopolitan subtype respectively. Two subtypes of DENV-1, subtype I and IV, were also found to be present. DENV-3 showed the greatest diversity with three subtypes of DENV-3 detected. The fourth known DENV-3 subtype, subtype IV, has not been reported since the 1970s and no complete genome of that subtype is available.

Evidence of sylvatic DENV-1, DENV-2 and DENV-4 was only available from samples collected in the 1960s and 1970s. This is not surprising because Malaysia has undergone rapid development and urbanisation since the beginning of the 1980s and sylvatic transmission would be expected to decline.

Understandably, very little sequence data were available from earlier decades, which is especially true for DENV-1 and DENV-4 (Figure 5.7). DENV-1 was the first dengue serotype isolated in Malaysia in 1954 (Smith, 1956) and DENV-1 has also been reported to be the predominant serotype in Malaysia for many years since the 1970s (Lam, 1994). In spite of this, there are few historical DENV-1 sequences available. Even so, it can be surmised that subtypes I and IV have probably been the endemic DENV-1 subtypes in Malaysia for at least the past 20 years.

The Cosmopolitan subtype is almost certainly the prevalent DENV-2 subtype as it has been detected persistently since 1968. There appears to be several distinct genetic lineages within the Cosmopolitan subtype rather than a single lineage that has been in unbroken transmission since the 1960s (Supplementary Figure 1). The Asian 1 subtype, which is common in Thailand and Vietnam, was detected in Malaysia in the 1980s and 1990s but its continued transmission in the new millennium is in doubt due to a lack of

molecular evidence. The presence in 1987 of the American/Asian subtype, traditionally associated with Thailand, Vietnam and southern China and later the Latin Americas, could represent an imported case although the country of origin cannot be inferred reliably from the phylogenetic tree (data not shown).

The snapshot (Figure 5.7) also indicates that subtypes I and II are the predominant DENV-3 subtypes in Malaysia. The subtype I clinical isolates clustered to the same branch as viruses from neighbouring Indonesia, Singapore and Timor Leste, indicating it is a widespread lineage that is found in a large geographical area in SEA (Figure 5.3). The sub-genomic length sequences point to the presence of subtype I in Malaysia since at least 1974. This lineage of viruses has been circulating in Indonesia since the early 1970s and has been implicated in causing at least four epidemics in the past two decades (Ong *et al*., 2008). It is highly probable that this subtype accounts for most of the DENV-1 incidences reported in Malaysia since the 1950s (Smith, 1956).

Based on sequence data from 1993 and 1994 Kobayashi *et al*. (1999) had previously suggested that the epidemic from 1992-1994 in Malaysia was caused by subtype II viruses introduced from Thailand. Our phylogenetic trees showed a short sequence from 1992 that clustered closely with the data set from Kobayashi *et al*. (data not shown) and so did the 1995 and 1996 DENV-3 samples sequenced in this study. This shows that the duration of the epidemic resulting from the introduced Thai strain lasted from at least 1992 to 1996. A recent study has shown that it could take a few years before an introduced virus causes an outbreak (Schreiber *et al*., 2009). However, the lack of sequence data from the 1989-1991 period prevented a better estimation of the time of the alleged introduction of subtype II DENV-3 from Thailand to Malaysia.

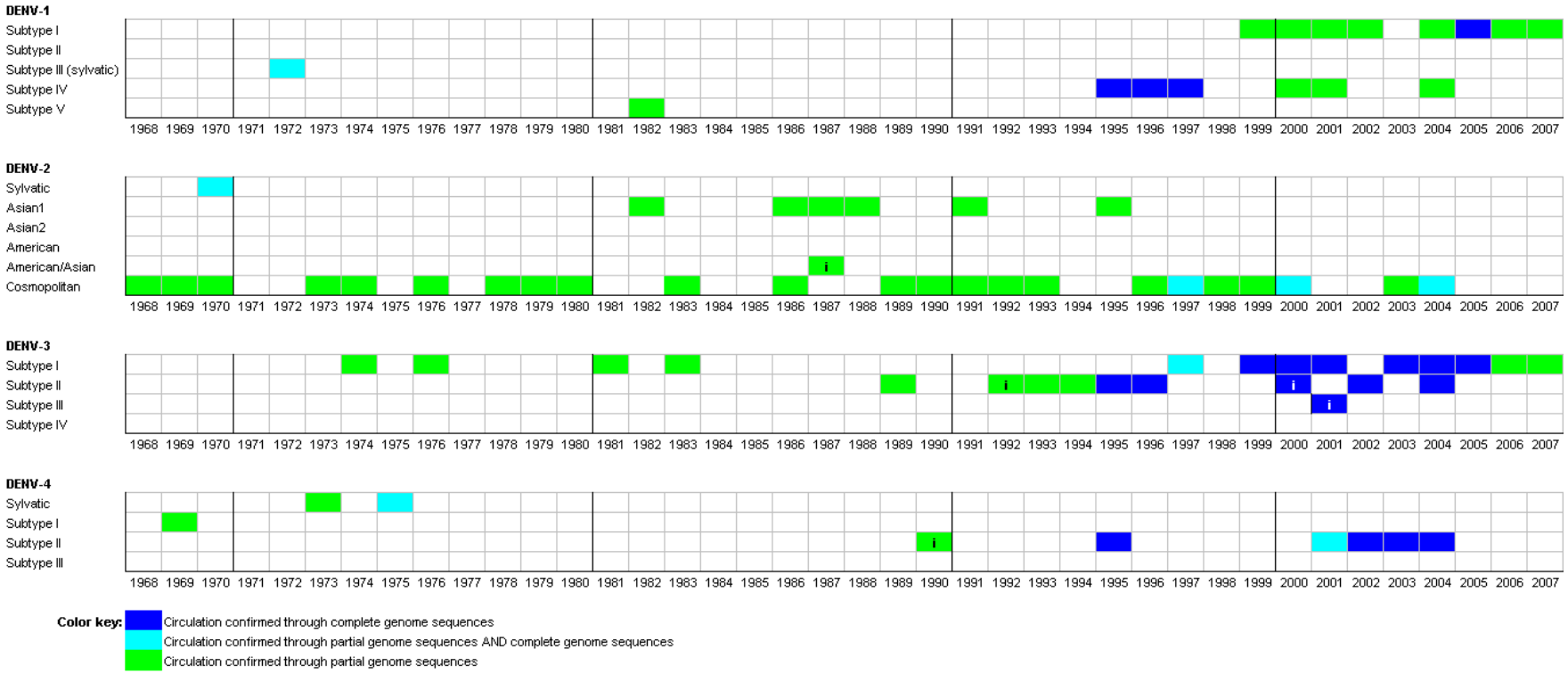
An earlier subtype II DENV-3 sequence from 1989 was found to be clustered with another lineage from Thailand and Vietnam. Although it is unlikely to be related to the above-mentioned epidemic, this could be evidence of another importation event but one which did not lead to an epidemic. Subtype II genome sequences from 2000 onwards clustered

together with a separate lineage of Thai isolates, indicating a third possible DENV-3 importation event from north of the Malaysian border.

Subtype III of DENV-3, which had never been reported in Malaysia, made a brief appearance in 2001. This subtype can be traced to Sri Lanka and it has been implicated in causing sustained epidemics in Central and South America since its introduction to Nicaragua and Panama in 1994 (Usuku *et al*., 2001; Miagostovich *et al*., 2002; Messer *et al*., 2003; Usme-Ciro *et al*., 2008). The detection of the subtype III isolate D3MY01-24056 points to an importation event from an unknown foreign source. The reason it did not develop into a sustained transmission could be due to the prevalence of subtype I and II DENV-3 which may have made it more difficult for another highly-similar virus strain to get established in the local vector population.

DENV-4 was reported to be the predominant serotype isolated from DF patients in Malaysia during the period from 1967 to 1969 (Rudnick, 1986). Our phylogenetic trees indicate the lone 1969 DENV-4 sequence as belonging to subtype I, whereas the more recent samples belong to subtype II. AbuBakar *et al*. (2002) had previously reported the presence of subtype II DENV-4 in 2001 and speculated it was the result of recombination events. All except one of our DENV-4 sequences were isolated from 2001 onwards and they all clustered near to the AbuBakar data set. The lone 1995 sequence was located on another branch of subtype II meaning it is not related to the 2001-2004 sequences. Again, due to a lack of sequence information from the period just prior to 2001 it is currently not possible to refute or corroborate the hypothesis put forward by AbuBakar *et al*. regarding the genesis of DENV-4 in Malaysia. A short 282-nucleotide sequence from 1990 (accession number AF051110, isolate LN-72992) clustered with sequences from Central America and is most likely the result of another importation from foreign sources.

Figure 5.7. History of dengue virus diversity in Malaysia from 1968 to 2007 as inferred through available sequence data. Coloured boxes with an "i" indicate likely importation events.



5.5 Discussion

Southeast Asia is generally considered as a source for dengue virus diversity (Kyle and Harris, 2008). Some of these virus strains are strongly associated with DHF/DSS and may be inherently more virulent than others as exemplified by their worldwide spread and their reported displacement of endogenous DENV strains in the Americas (Rosen, 1977; Rico-Hesse *et al*., 1997; Leitmeyer *et al*., 1999).

The 46 complete genome sequences obtained in this study reaffirmed the fact that dengue is hyper-endemic in Malaysia, with all four serotypes and multiple subtypes of DENV recovered during the sampling period. The effect of this hyper-endemicity and great viral genetic diversity on the epidemiology of the disease in SEA is unclear. However, a consistent theme encountered when analysing the phylogeny of the DENV serotypes is that the SEA countries, in particular Malaysia, Singapore and Indonesia, tend to share the same circulating subtypes of viruses. This not only hints strongly of frequent transmission and free exchange of virus strains across national borders, but also predicates that any future dengue control programs (e.g. vector eradication, vaccination et cetera) need to be implemented in a concerted manner between all regional governments and international bodies to have any real chance of success.

Malaysia consists of two parts: West Malaysia which borders Thailand to the north and Singapore to the south, and East Malaysia on the island of Borneo which shares borders with Brunei and Indonesia. The clinical isolates sequenced in this study were collected in the capital city of Kuala Lumpur which occupies a central location in West Malaysia. With a few exceptions, the uncovered sub-genomic length sequences were also from West Malaysia. Therefore, the results arising from this study may only be considered as representative of the circulating viruses in cities in West Malaysia but not of those in East Malaysia.

From the minimum evolution (ME) and maximum likelihood (ML) trees generated from 34 DENV-4 complete genome sequences (Figure 5.4 and Figure 5.5) we have shown that there is little to differentiate the two except for branch lengths. As the number of publicly available DENV complete genome sequences has been increasing at a tremendous pace in recent years, the popular but computationally-demanding character-based ML method of phylogeny inference could soon prove to be too resource-straining for regular application in the not-too-distant future. Provided the whole length of the DENV genome sequences are used, the computationally less intensive distance-based ME method should be considered as a viable alternative to the ML method for practical reasons. We have also shown that isolate classification based on sub-genomic length sequences such as the complete E gene and E-NS1 junction is consistent down to the subtype level. We could therefore take advantage of this fact to infer the history of DENV subtype prevalence in Malaysia using all the available sequence data.

The comprehensiveness of the Malaysian sub-genomic length sequences included in this study can only be described as a *best effort*. Due to the less-than-perfect way metadata of sequences are recorded in public sequence depositories such as GenBank, it is highly possible there are other usable sequences from Malaysian samples that have been missed. There is also the possibility that there exist more sequences which were only described in literature but were not submitted to public sequence depositories – as exemplified by the 24 E/NS1 DENV-2 sequences from Fong *et al*. (1998) and 6 complete E DENV-1 sequences from A-Nuegoonpipat *et al*. (2004) that added considerable amount of valuable information to this study. All in all, it is not unreasonable to conclude that the total amount of DENV sequence from Malaysia is also a very small sample of the total diversity that might be (or have been) circulating in the country.

Although the snapshot of dengue virus diversity was constructed from all available sequence data, great care should be taken when interpreting it due to the relative scarcity of sequences from the earlier decades. For example it has been reported that DENV-1 was the serotype most commonly

encountered between 1973 and 1982 (Lam, 1994), yet a dearth of DENV-1 sequence data from that period may give the wrong impression that DENV-1 was not found in the country until subtype V was detected in 1982. In fact, DENV-1 was detected in Malaysia in as early as 1954 (Smith, 1956). A more accurate picture can only emerge if historical virus samples were subjected to sequencing and the newly-derived sequence information used to enhance the completeness of the snapshot. Looking forward to the future, it is strongly recommended that sampling and complete genome sequencing of DENV be conducted on a regular basis, not just during epidemic years but also during inter-epidemic periods.

Since the two criteria for inclusion in this study are the availability of sequence data and isolation year, the constructed snapshot (Figure 5.7) gives no information about the predominant serotype or subtype. Information about the former can only be gleaned from extensive serological testing of clinical samples, whereas the predominant subtype currently can only be achieved through large-scale sequencing which is unlikely to be carried out in an outbreak situation. Consequently, it should be remembered that the *raison d'setre* of such snapshots is to provide qualitative and not quantitative information. This gap in quantitative information underlies the difficulty in linking dengue outbreaks to the introduction of new strains of viruses that led to a change in the predominant DENV subtype in circulation. Decades of dengue epidemiological research has so far unearthed only a few examples of introduced virus lineages with strong correlation to epidemics of increased severity (Rico-Hesse *et al*, 1997; Messer *et al*., 2003).

This study has demonstrated that data from complete genome and sub-genomic length sequences can be used synergistically to construct a picture of the history of the dengue virus for a particular geographical area. The value of having complete genomes has been heightened by using them as signposts that allowed sub-genomic length sequences to be reliably incorporated into phylogenetic analyses. This simple approach could open doors for new molecular epidemiological studies of the dengue virus –

especially for localities where currently only fragmentary sequence data are available and where there is little possibility of sequencing historical isolates.

In conclusion, the sequencing of the 46 complete genomes of DENV isolated between 1995 and 2005 has contributed to a better understanding of the molecular epidemiology of dengue in Malaysia. If more historical isolates from the country could be obtained for sequencing, we are certain a more complete picture could be constructed. We look forward to concurrently using complete genome and fragmentary sequence data originating from neighbouring countries such as Thailand, Singapore and Indonesia to draw a clearer picture of dengue virus molecular epidemiology in Southeast Asia.

5.6 Acknowledgements

We thank Lam Sai Kit and Fong Mun Yik from the University of Malaya for providing the 24 DENV-2 E/NS1 sequences and John Aaskov from the Queensland University of Technology for providing the 6 DENV-1 complete E sequences which are not found in GenBank. Inclusion of these data sets contributed to the completeness of this study.

5.7 Author Contributions

SH Ong sequenced the genomes, assembled the sequences, designed and performed the sequence analyses, and wrote the manuscript. Contributions from co-authors are as listed. Conceived the study: SGV, MLH, MJS; collected and isolated the samples: SD; propagated the virus samples: WL; sequenced the genomes: JS, PPKA, JTY, PYS; performed the sequence analysis: KSG; reviewed the manuscript: SD, SGV, MLH, MJS.

5.8 References

1. A-Nuegoonpipat A, Berlioz-Arthaud A, Chow V, Endy T, Lowry K, Mai le Q, Ninh TU, Pyke A, Reid M, Reynes JM, Su Yun ST, Thu HM, Wong SS, Holmes EC and Aaskov J (2004) Sustained transmission of dengue virus type 1 in the Pacific due to repeated introductions of different Asian strains. *Virology*. 329:505–512.
2. AbuBakar S, Wong PF and Chan YF (2002) Emergence of dengue virus type 4 genotype IIA in Malaysia. *J Gen Virol*. 83(10):2437–2442.
3. Chee HY and AbuBakar S (2003) Phylogenetic investigation of dengue virus type 2 isolated in Malaysia. *Dengue Bulletin*. 27:100–107.
4. Fong MY, Koh CL and Lam SK (1998) Molecular epidemiology of Malaysian dengue 2 viruses isolated over twenty-five years (1968-1993). *Res Virol*. 149(6):457–464.
5. Goncalvez AP, Escalante AA, Pujol FH, Ludert JE, Tovar D, Salas RA and Liprandi F (2002) Diversity and evolution of the envelope gene of dengue virus type 1. *Virology*. 303:110–119.
6. Gubler DJ and Clark GG (1995) Dengue/Dengue Hemorrhagic Fever: the emergence of a global health problem. *Emer Infec Dis.* 1(2):55–57.
7. Hammon WM, Rudnick A and Sather GE (1960) Viruses associated with epidemic hemorrhagic fevers of the Philippines and Thailand. *Science*. 131:1102–1103.
8. Igarashi A, Fujita N and Okuno Y (1982) Isolation of dengue viruses from patients with dengue haemorrhagic fever (DHF) and those with fever of unknown origin (FUO) in Jarkata, Indonesia, in the years of 1981 and 1982. ICMR Annals Kobe University 2:7–17.
9. Kobayashi N, Thayan R, Sugimoto C, Oda K, Saat Z, Vijayamalar B, Sinniah M and Igarashi A (1999) Type-3 dengue viruses responsible for the dengue epidemic in Malaysia during 1993-1994. *Am J Trop Med Hyg*. 60(6):904–909.
10. Kyle JL and Harris E (2008) Global spread and persistence of dengue. *Annu Rev Microbiol*. 62:71–92.
11. Lam SK (1994) Two Decades of Dengue in Malaysia. *Tropical Biomedicine*. 35(4):195–200.

12. Lam SK, Devi S and Pang T (1987). Detection of specific IgM in dengue infections. *Southeast Asian J. Trop. Med. Publ. Hlth*. 18:532–538.
13. Lanciotti RS, Lewis JG, Gubler DJ and Trent DW (1994) Molecular evolution and epidemiology of dengue-3 viruses. *J Gen Virol*. 75:65–75.
14. Lanciotti RS, Gubler DJ and Trent DW (1997) Molecular evolution and phylogeny of dengue-4 viruses. *J Gen Virol*. 78:2279–2286.
15. Leitmeyer KC, Vaughn DW, Watts DM, Salas R, de Chacon IR, Ramos C and Rico-Hesse R (1999) Dengue virus structural differences that correlate with pathogenesis. *J. Virol*. 73(6):4738–4747.
16. Lewis JA, Chang GJ, Lanciotti R, Kinney RM, Mayer LW and Trent DW (1993) Phylogenetic relationships of dengue-2 viruses. *Virology*. 197:216–224.
17. Messer WB, Gubler DJ, Harris E, Sivananthan K and de Silva A (2003) Emergence and global spread of a dengue serotype 3, subtype III virus. *Emer Infec Dis*. 9:800–809.
18. Miagostovich MP, dos Santos FB, de Simone TS, Costa EV, Filippis AMB, Schatzmayr HG and Nogueira RMR (2002) Genetic characterization of dengue virus type 3 isolates in the State of Rio de Janeiro, 2001. *Braz J Med Biol Res.* 35: 869–872.
19. Ong SH, Yip JT, Chen YL, Liu W, Harun S, Lystiyaningsih E, Heriyanto B, Beckett CG, Mitchell WP, Hibberd ML, Suwandono A, Vasudevan SG, and Schreiber MJ (2008) Periodic re-emergence of endemic strains with strong epidemic potential-a proposed explanation for the 2004 Indonesian dengue epidemic. *Infect Genet Evol*. 8:191–204.
20. Posada D and Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 14, 817–818.
21. Quinlos FN, Lim LE, Juliano A, Reyes A and Lacson P (1954) Haemorrhagic fever observed among children in the Philippines. *Philip J Pediatrica*. 3:1–19.
22. Rzhetsky A and Nei M (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol*. 10(5):1073–1095.
23. Rico-Hesse (1990) Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology*. 174:479–493.

24. Rico-Hesse R, Harrison LM, Salas RA, Tovar D, Nisalak A, Ramos C, Boshell J, de Mesa MT, Nogueira RM and da Rosa AT (1997) Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology*. 230(2):244–251.
25. Rosen L (1977) The emperor's new clothes revisited, or reflections on the pathogenesis of dengue hemorrhagic fever. *Am J Trop Med Hyg*. 26(3):337–343.
26. Rudnick A, Tan EE, Lucas JK and Omar MB (1965) Mosquito-borne haemorrhagic fever in Malaya. *Br Med J.* 1(5445):1269–1272.
27. Rudnick A (1986) Dengue fever epidemiology in Malaysia 1901-1980. In Rudnick A & Lim TW (ed.), Dengue Fever Studies in Malaysia, Bulletin No. 23, pp. 9–38. Kuala Lumpur: Institute of Medical Research.
28. Schreiber MJ, Ong SH, Holland RCG, Hibberd ML, Vasudevan SG, Mitchell WP and Holmes EC (2007) DengueInfo: A web portal to dengue information resources. *Infect Genet Evol*. 7(4):540–541.
29. Schreiber MJ, Holmes EC, Ong SH, Soh HS, Liu W, Tanner L, Aw PP, Tan HC, Ng LC, Leo YS, Low JG, Ong A, Ooi EE, Vasudevan SG and Hibberd ML (2009) Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J Virol*. 83(9):4163–4173.
30. Skae PMT (1902) Dengue fever in Penang. *Brit. Med. J*. 2:1581–1582.
31. Smith CEG (1956) Isolation of three strains of type 1 dengue virus from a local outbreak of the disease in Malaya. *J Hyg (Lond)*. 54(4):569–580.
32. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
33. Tamura K & Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 10(3):512–526.
34. Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol*. 24(8):1596–1599.
35. Twiddy SS, Farrar JJ, Nguyen VC, Wills B, Gould, EA, Gritsun T, Lloyd G and Holmes EC (2002) Phylogenetic relationships and differential

selection pressures among genotypes of dengue-2 virus. *Virology*. 298:63–72.

36. Usme-Ciro JA, Mendez JA, Tenorio A, Rey GJ, Domingo C, Gallego-Gomez JC (2008) Simultaneous circulation of genotypes I and III of dengue virus 3 in Colombia. *Virol J*. 5:101.
37. Usuku S, Castillo L, Sugimoto C, Noguchi Y, Yogo Y and Kobayashi N (2001) Phylogenetic analysis of dengue-3 viruses prevalent in Guatemala during 1996–1998. *Arch Virol*. 146:1381–1390.
38. Vasilakis N, Holmes EC, Fokam EB, Faye O, Diallo M, Sall AA and Weaver SC (2007) Evolutionary processes among sylvatic dengue type 2 viruses. *J Virol*. 81(17):9591–9595.
39. Wang E, Ni H, Xu R, Barrett AD, Watowich SJ, Gubler DJ and Weaver SC (2000) Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J Virol*. 74(7):3227–3234.
40. World Health Organization (1997) Dengue haemorrhagic fever: diagnosis, treatment, prevention and control. 2nd ed. Geneva: World Health Organization.
41. World Health Organization (2009) Dengue and dengue haemorrhagic fever. (http://www.who.int/mediacentre/factsheets/fs117/en/, accessed 1 May 2009).

6. Discussion and Conclusions

In the last two decades of the twentieth century the world has witnessed the emergence or re-emergence of a number of viral pathogens such as human immunodeficiency virus (HIV), West Nile virus, severe acute respiratory syndrome coronavirus (SARS-CoV) and a number of influenza A subtypes. The majority of these pathogens crossed the species barrier from their natural reservoirs to human populations due to human encroachments to the natural habitat of the reservoir and vector species. In a relatively short period of time, these pathogens were disseminated globally by jet travels or migratory birds.

Another commonality of the viruses cited above is that they are all RNA viruses. RNA viruses tend to have high mutation rates and consequently they can evolve rapidly to resist treatments and to evade host immune responses. Disease emergence can be influenced by several complex interacting factors such as environmental, ecological, socio-economic and behavioral influences (Morens *et al*., 2004). Because of this complexity, studying the genetic blueprint of the viral pathogen is a relatively simple approach to unraveling their virulence or pathogenesis.

6.1 Argument for viral genome sequencing

RNA viruses lack proof-reading mechanisms in their RNA polymerases and undergo constant mutations. Most of these mutations tend to be either neutral or deleterious - in which case the deleterious mutations are purged from the viral population via natural selection. However, on rare occasions a mutation can endow the virus with the ability to infect a different species, to trigger diseases that are more serious, or to develop resistance to treatment.

The only way to study such molecular events is through the comparative analysis of the complete genetic blueprint of the viruses. Fortunately, in most cases viral genome sequences are miniscule in size compared to microbial and eukaryote genomes. Due to their small size, viral genomes are amenable to large-scale sequencing efforts involving sufficient

number of isolates with sufficient sampling frequency. This is something that is currently challenging for the genomes of protozoan or metazoan organisms.

Another strong reason to advocate the sequencing of viral genomes is due to the fact that viruses are *ephemeral* (the exception being retroviruses). They leave little trace of their genetic makeup once an outbreak is over and they are cleared from the affected host population. The value of the genetic blueprint in helping to understand the pathogenesis of a viral infectious agent was underscored by the well-documented effort to sequence the 1918 Spanish influenza virus strains from archival autopsy tissues (Taubenberger *et al*. 1997; Reid *et al*., 2003).

The main body of work described in this thesis covers: (1) the development of a reliable laboratory protocol to sequence and assemble the complete (+)ssRNA genome of the dengue viruses, (2) the efforts to sequence the complete genomes of dengue virus clinical samples which were collected in three dengue-endemic Southeast Asian countries, Indonesia, Malaysia and Singapore, (3) the development of an informatics infrastructure to store, organize, manipulate and analyse the sequence data, and (4) the derivation of knowledge from the annotated sequence data to answer pertinent molecular epidemiological questions.

Before analyses and knowledge discovery can be performed efficient informatics architectures for the storage and easy retrieval of carefully annotated viral genome sequence data are required. The knowledge discovery step relies mainly on using comparative genomics and molecular evolutionary approaches to predict the origin and spread of dengue viruses, and to understand the causes of epidemics and the genetic basis of virulence. The dividends to be gained from such comparative sequence analysis are tremendous: it may make prediction of future dengue epidemics a possibility and may also point to new routes for the development of antiviral drugs or vaccines.

6.2 Complete genome versus partial sequencing

In the early 1990s partial sequencing of the dengue genome was first used to uncover the origin of viral strains during outbreak situations (Rico-Hesse, 1990). Since that time many applications including the discovery of virulence factors, tracking the spread of dengue virus strains, study of evolutionary selection pressure and vaccine development have relied on partial genome sequencing. This preference for limited sequencing of course has in it elements of the technological limitations of the time, affordability and expediency. Nevertheless, recent advances in sequencing technology and the availability of established protocols for complete genome sequencing have made sequencing the relatively small genomes of dengue viruses within reach of most workers.

Frequently investigations to link phenotypic traits or epidemic potential to viral genetics were inconclusive because only small fragments of the viral genome, in particular the E gene, were sequenced and studied (Fong *et al*., 2004; A-Nuegoonpipat *et al*., 2004). Mutations in genes other than E (Twiddy *et al*., 2002) or in the untranslated regions of the DENV genome (Leitmeyer *et al*. 1999; Durbin *et al*., 2001; Lodeiro *et al*., 2008) have previously been identified as the genetic determinants for traits such as neurovirulence and attenuation.

Similarly, due to the availability of only partial sequences molecular epidemiology studies often reported inconclusive results when attempting to undercover the genetic basis of disease severity. This has been the case from the earliest studies (Chungue *et al*, 1993) to the more recent ones (Figueiredo *et al*., 2008). In addition, analyzing only partial sequences could potentially lead to erroneous conclusions. Only a thorough analysis of the complete genetic blueprint of the DENV – obtained through sequencing of the complete genome – could lead to unambiguous conclusions.

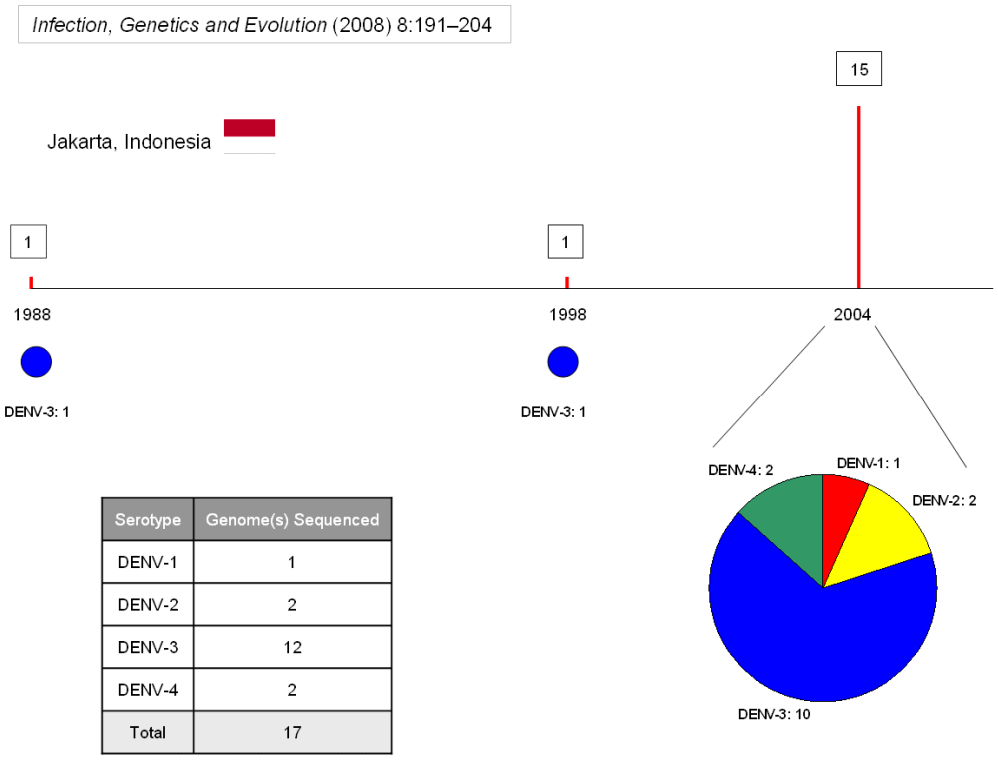
All sequencing in this study was done with the aim of obtaining the complete genome sequences of the dengue virus samples. This allowed unambiguous interpretation of the sequence analysis results. Additionally, regions extracted from the complete genome sequences obtained could also be used, in a concurrent manner, with existing partial sequences to construct a probable history of dengue virus diversity. These same regions could also be extracted and included in previous analyses using only existing partial sequences to enhance the breadth if not the accuracy of those studies.

6.3 Sampling strategies and findings

Each of the three DENV genome sequencing projects described in Chapters 3, 4 and 5 adopted a different approach in terms of the sampling strategy. As a result, the DENV genome sequences thus obtained allowed different kinds of questions to be asked. Likewise, there are dissimilar shortcomings in each of the studies due to the differences in the sampling strategy.

6.3.1 The 2004 epidemic in Jakarta, Indonesia

For this study, a medium quantity of virus samples (n=69) were obtained from hospitals within the city of Jakarta during a dengue epidemic in 2004. Only 15 virus samples were successfully amplified and sequenced. As DENV-3 was found to be the predominant serotype, historical DENV-3 samples from the city were obtained for sequencing (Figure 6.1). The collected genome data allowed two main questions to be answered. The first was whether the same genotype of DENV-3 had been circulating locally since at least 1988 (the oldest available historical sample from Jakarta), or whether the 2004 genotype was a strain imported from a foreign source. Secondly, assuming the former hypothesis was true, the sequence data would allow the examination of amino acid changes over the 16-year period to look for possible signs of virus adaptive evolution that contributed to the epidemic in 2004.



Serotype	Genome(s) Sequenced
DENV-1	1
DENV-2	2
DENV-3	12
DENV-4	2
Total	17

Figure 6.1. The strategy employed for virus sampling in the Jakarta 2004 DENV genome sequencing study. Sample size by year and serotype is indicated as boxed figures and coloured pie chart respectively, and the total serotype breakdown is indicated in the inset table.

The two historical DENV-3 samples, from 1998 and 1988 respectively, were found to be from the same genotype as the 2004 viruses. Phylogenetic inference indicates they are from the same lineage (known as subtype I) that could be traced to viruses isolated in the same country as far back as 1973. This effectively ruled out recent importation and/or introduction of a new virulent DENV-3 strain as the cause of the 2004 epidemic in Jakarta.

An additional four subtype I DENV-3 complete genome sequences dating from 1998 were found to be available publicly. These samples were collected in the city of Palembang on the island of Sumatra, some 200 km to the north of Jakarta during an epidemic in that year. The sixteen subtype I DENV-3 complete genomes (12 from this study) were therefore grouped together for a comparative analysis of their deduced amino acid sequences in order to answer the second question.

A total of 98 positions out of the 3390 amino acid residues in the subtype I DENV-3 polyprotein were found to have non-synonymous mutations (Figure 6.2). Viruses from 1998 and 2004 were almost indistinguishable at the amino acid level and none of the observed amino acid changes could be suggested as the trigger for the epidemic in 2004. However, there are a number of notable differences between the older 1988 sequence and the other fifteen samples, in particular at E-301, NS1-93, NS1-98, NS1-127, NS2B-60 and NS3-60. E-301 is just next to the putative receptor-binding site on domain III of the dengue envelope protein. Sadly, no follow-up experiments could be pursued to examine the effects of the aforementioned mutations on virus replication efficiency and infectivity due to restrictions in the material transfer agreement governing their use.

One major deficiency of this study is the total lack of serological information of the affected population in Jakarta. If the inhabitants were exposed to the same virus genotype barely 6 years ago, there ought to be a certain percentage that has acquired immunity against it. And, assuming further that the subtype I viruses were absent in the intervening years between 1998 and 2004, the age of the affected patients would reflect this. There were, however, no relevant data to verify or refute this hypothesis.

The second limitation of this study is the fact that only one sample from 1988 was available. There is no compelling reason to believe this sample is representative of the DENV-3 viruses circulating in Jakarta in 1988. The third shortcoming is that there were no virus samples from the intervening years between the three epidemics in 1988, 1998 and 2004. Indonesia experiences dengue perennially and there should be no lack of samples if active collection of virus sample were in practice. Regular sample collection would have given investigators the means to understand better the role of virus adaptive evolution in the outbreak of epidemics.

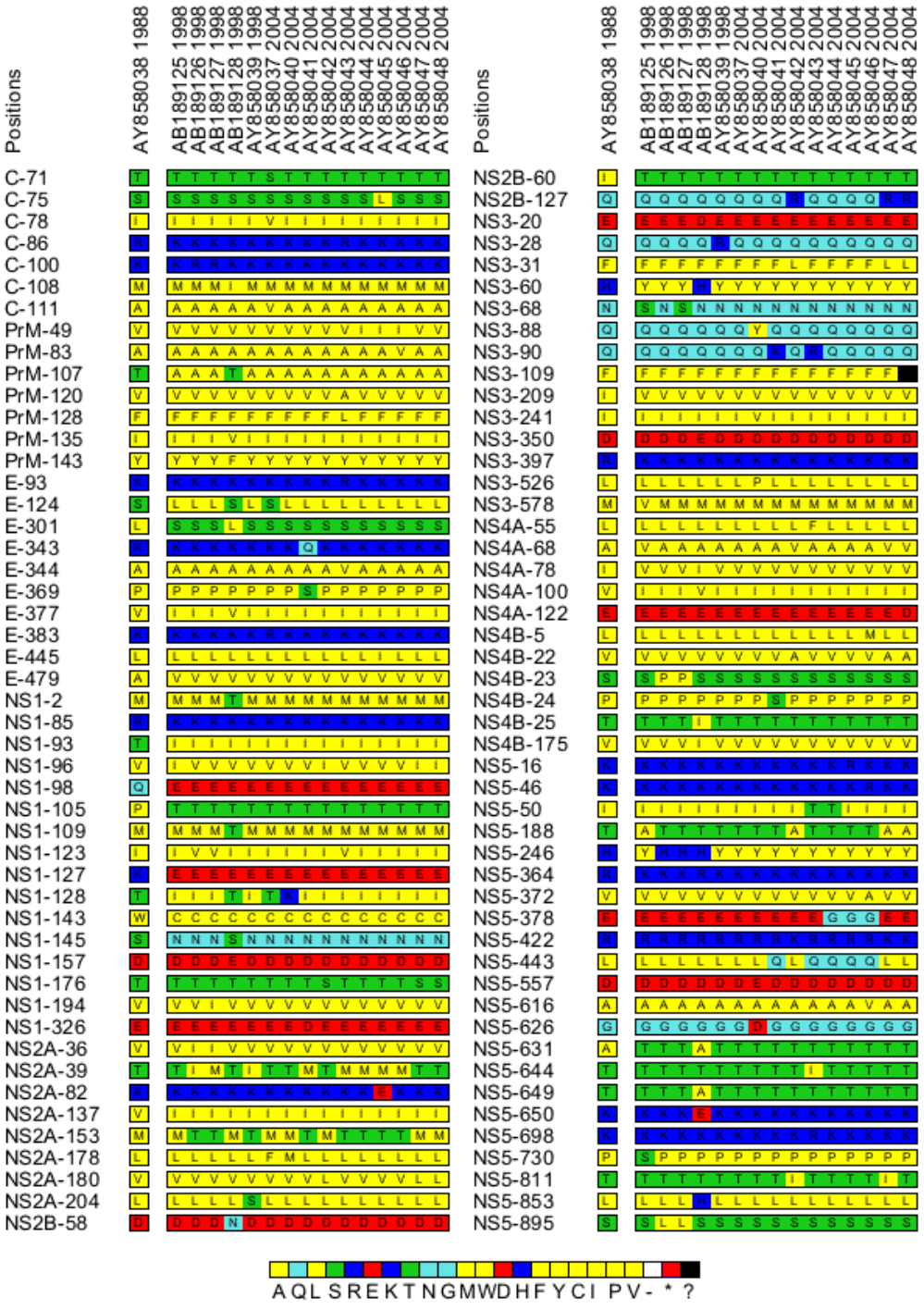


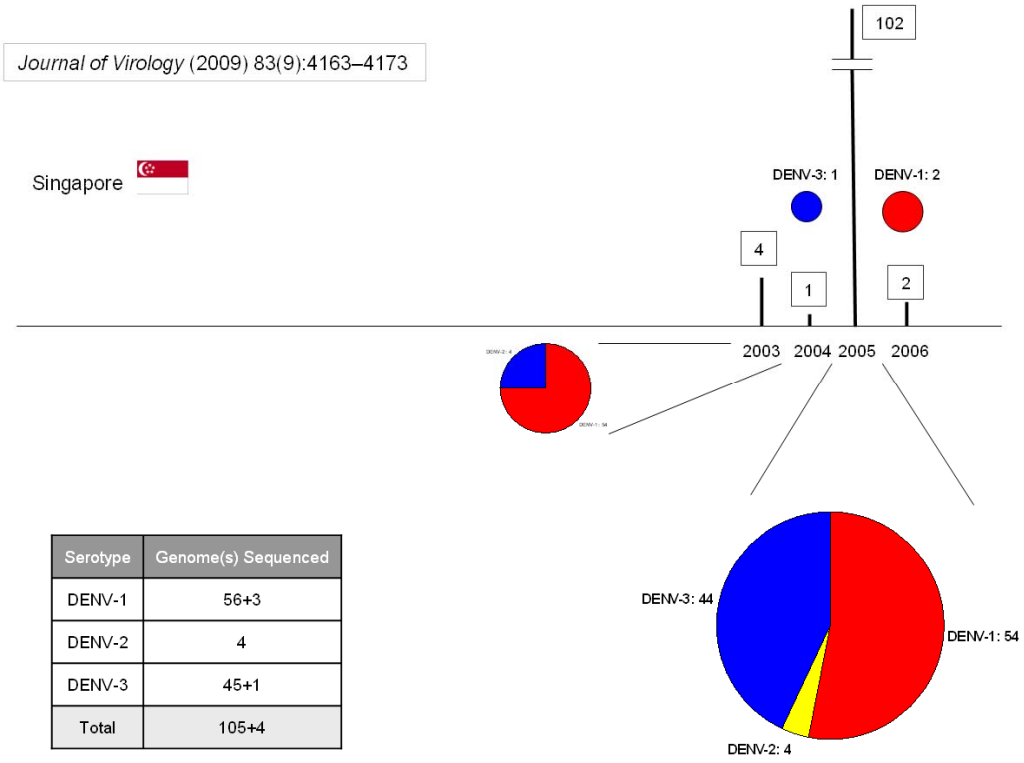
Figure 6.2. Amino acid differences among the Indonesian subtype I DENV-3 strains isolated in 1988, 1998 and 2004. The sequences are sorted in chronological order followed by accession number, starting from the left. Only 98 positions out of 3390 of the DENV-3 polyprotein showed any mutations, and none could be suggested as the trigger for the epidemic in 2004. Amino acids are color-coded by their chemical properties as shown in the key beneath the main diagram.

Although DENV-1, DENV-2 and DENV-4 were also detected in the 2004 epidemic, the presence of all four serotypes during an outbreak is not unusual and the number of sequences obtained from these serotypes was too small to allow any in-depth analysis. Phylogenetic analysis from the genome sequences did reveal that, as in the case for DENV-3, they all belong to lineages that have been endemic in the Indonesian archipelago for several decades.

In conclusion, virus sample collection during non-epidemic years on a regular basis is recommended for countries experiencing perennial dengue infections. For a country as large as Indonesia it would be impractical to get a large enough sample size that can reveal the entire diversity of DENV in the whole country. Sample archiving efforts should therefore be concentrated in densely populated urban centers like Jakarta and Palembang where the probability of dengue epidemics occurring is highest. Being important political, commercial and cultural centers on the islands of Java and Sumatra respectively, it is also highly likely that these cities could serve as the starting points of dengue epidemics that then propagate to the surrounding regions. Spatial-temporal patterns of DHF incidences in Thailand have pointed to the city of Bangkok as playing such a role (Cummings *et al*., 2004).

6.3.2 The 2005 epidemic in Singapore

The strategy adopted during the epidemic in Singapore in 2005 was to acquire very high-resolution sequence data by doing high-frequency sampling throughout the duration of the epidemic (Low *et al*., 2006). This approach allowed the study of DENV molecular epidemiology at an unprecedented temporal and spatial resolution. The 105 complete genome sequences generated from this study was the first high-resolution DENV genome data set ever reported (Schreiber *et al*., 2009). Four other pre-epidemic samples were subsequently obtained and sequenced in order to examine the role of adaptive evolution in the triggering of the 2005 epidemic (Figure 6.3).



Serotype	Genome(s) Sequenced
DENV-1	56+3
DENV-2	4
DENV-3	45+1
Total	105+4

Figure 6.3. The strategy employed for virus sampling in the Singapore 2005 DENV genome sequencing study. Sample size by year and serotype is indicated as boxed figures and coloured pie chart respectively, and the total serotype breakdown is indicated in the inset table.

Serological tests of the collected samples revealed the presence of DENV-1, DENV-2 and DENV-3 during the epidemic. DENV-1 and DENV-3 accounted for the majority of the cases while DENV-2 registered only a nominal presence. The number of genome sequences obtained also reflected this pattern. Phylogenetic analyses revealed the presence of more than one virus subtype for DENV-1 and DENV-3. In both cases, there was one dominant subtype while the second subtype caused only isolated cases. Although multiple subtypes of DENV-2 have previously been found circulating in a single epidemic in Thailand (Pandey and Igarashi, 2000), the complexity of virus subtypes observed in this study was much higher.

The four DENV-2 isolates were found to belong to the Cosmopolitan subtype, a virus lineage that has long been known to be at large in the Southeast Asian region. Unsurprisingly subtype I, which is another virus lineage long associated with the region, was the predominant DENV-1 in the epidemic. The solitary outlier belongs to subtype V which is predominantly found in Latin America and West Africa, although its close clustering with a 2004 sequence from Réunion suggests it could be an imported case by travellers from the Indian Ocean island.

The oddity observed in the DENV-3 cases is that the endemic subtype I was not the predominant subtype. Instead, subtype III DENV-3, which was originally associated with the Indian subcontinent, was responsible for the majority of the DENV-3 cases. The subtype III viruses have been implicated in causing epidemics in Latin America and are considered as possessing high epidemic potential. Phylogenetic analysis of the nucleotide sequences and comparative studies of the deduced amino acid sequences of two DENV-3 samples from 2003 and 2004 both point to the occurrence of *in situ* evolution of this lineage since its introduction to Singapore. Although the actual time of the alleged introduction cannot be ascertained, it is likely to have occurred in 2003 or even earlier. This indicates that the introduced lineage, despite the strong correlation between its introduction and epidemics in Latin America, did not cause an immediate epidemic in Singapore.

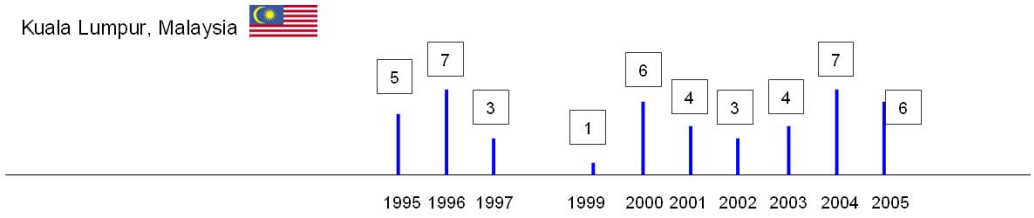
This high-resolution data set allowed the population dynamics of the DENV-1 and DENV-3 to be measured and compared in an epidemic situation, which is something that was never before possible. The Bayesian coalescent analysis used in the study revealed that the two group of viruses exhibited similar epidemic profiles, despite the difference in their probable origin as an endemic lineage (DENV-1) and a recently introduced lineage (DENV-3). The conclusion that can be drawn is that the evolution of the virus is unlikely to be the central factor in shaping the dynamics of this dengue outbreak.

Despite the huge amount of data collected in this study, the relative difficulty in getting pre-epidemic samples for comparative analysis underscores the fact that regular sequencing of clinical isolates during non-epidemic years is essential in order to have a better understanding of the role of viral genetics in causing epidemics. One final observation is that epidemic surveillance of viral genome sequences alone would not have been sufficient to predict the 2005 outbreak. Concurrent surveillance of virus diversity, mosquito vectors (including the proportion of mosquitoes infected with DENV), and periodic survey of population sero-prevalence rates may provide the additional required predictive information.

6.3.3 The 1995-2005 collection from Kuala Lumpur, Malaysia

The clinical samples sequenced in this study differed radically from the previous two in that they did not originate from a single epidemic but were collected longitudinally over a period of 11 years from a single locality (Figure 6.4). As such, the kind of questions that could be asked also differs from the two previous studies to a great extent.

Malaysia is subjected to perennial dengue outbreaks and the data collected may be considered as coming from an 11-year-long epidemic. The number of sequences obtained (46) can be considered as large when compared to other sequencing efforts. However, except probably for DENV-3, the number of sequences available per serotype per year is still insufficient for any meaningful analysis apart from allowing the identification of virus subtypes for the years in which samples were available (Figure 6.4). This reveals that the number of samples collected per year in this study was not high enough to adequately investigate the diversity of circulating dengue viruses during the study period.



Serotype	Genome(s) Sequenced
DENV-1	10
DENV-2	5
DENV-3	23
DENV-4	8
Total	46

Figure 6.4. The distribution by year of the dengue clinical isolates collected between 1995 and 2005 in Kuala Lumpur, Malaysia. Sample size by year and serotype is indicated as boxed figures and coloured pie chart respectively, and the total serotype breakdown is indicated in the inset table.

To extend the usefulness of the obtained sequence data, a simple but laborious method was devised that allowed all existing partial sequences from Malaysian DENV samples to be used concurrently to construct a qualitative snapshot of the diversity of dengue viruses in the country. Short sequences are known to be unreliable for interpreting detailed phylogenetic relationships but this study showed that such sequences produced consistent classification of DENV at the subtype level. As a result, the subtype classification information was used to build the first (and admittedly incomplete) picture of the history of dengue virus diversity in Malaysia.

The results showed that Malaysia has experienced several episodes of importation in addition to maintaining several lineages of viruses that have been in sustained transmission for several decades. Evidently, only one of these importation events (of subtype II DENV-3 from Thailand circa 1992) led to a change in the predominant circulating virus subtype and a pronounced epidemic that lasted several years (Kobayashi *et al*., 1999). Eight years later this lineage of virus was inexplicably no longer found in the sample. The subtype II viruses from 2000 and beyond probably shared a common ancestor with the 1992 strains but clearly belonged to a different lineage - perhaps the result of another importation event.

There is no straightforward explanation for the observed clade extinction event of the subtype II DENV-3. Because of Malaysia's fairly constant tropical climate there are no identifiable climatic factors that could account for the disappearance of dengue virus clades as has been reported elsewhere (Zhang *et al.*, 2005). With a population size of 25.7 million (July 2009 estimate), herd immunity, another possible explanation for viral clade extinction, is also unlikely to be the factor. Although the actual causes of the clade replacement is unclear, the generalisation that can be made is that imported viruses do, occasionally, possess the ability to cause widespread epidemics for several years. Monitoring changes in the prevalent DENV subtypes (not just serotypes) may assist in predicting future dengue outbreaks.

The various phylogenetic analyses showed that the three Southeast Asian countries of Indonesia, Malaysia and Singapore to a large extent share the same subtypes of dengue viruses. That the distribution and transmission of DENV strains transcends national borders should come as no surprise. One important implication arising from this observation is that dengue control programs such as vector eradication (and vaccination, when a vaccine is available) must be implemented in a concerted manner between all regional governments and international bodies to increase the likelihood of success.

6.4 Importance of sustained sequence monitoring

Molecular epidemiology approaches provide a viable way to understand the role of viral genetics in disease transmission and clinical severity. Of the three components involved in the transmission cycle of dengue (virus, vector and host), the viral component is the most amenable to genetic investigations. Most of the groundwork required for investigating the role of dengue virus genetics has been demonstrated in the course of this project. What is needed next is continued sampling and sequencing of clinical isolates and sustained monitoring of the accumulated sequence data.

Among the most pertinent questions that need answering is how the dengue virus would react to the introduced selective pressure when vaccines and antivirals for dengue eventually become available. Being an RNA virus, it is a good bet that it is just a matter of time before mutations that allow it to escape the effect of the interventions would appear. In general, detection of virus adaptive evolution tends to become increasingly difficult as time goes on. Advantageous mutations are likely to be selected and quickly become one of the numerous “conserved” residues when sequence analysis is done on samples collected long after the mutations have achieved fixation. Two recent examples of viral adaptation that were successfully detected, the WNV NS3-T249P (Brault *et al*., 2007) and CHIKV E1-A226V (Schuffenecker *et al*., 2006) mutations, were found relatively soon after the mutational events were believed to have taken place. Data from currently available dengue genome sequences shows there are 1315 residues in the DENV polyprotein (nearly

39%) that are 100% conserved across all four dengue serotypes. This suggests that many advantageous mutations have already reached fixation before the start of the global spread of dengue disease in the 1950s. Regular monitoring of DENV sequence evolution would enable the detection of such adaptive events as they occur.

It is a certainty that the dengue virus will continue to evolve, and probably at an accelerated rate after a drug or vaccine becomes available. This means it will be important to keep monitoring the changes in the virus genome so that any future adaptive events can be detected quickly and its impact on the virulence and drug/vaccine efficacy be assessed.

6.5 Recommendations for future studies

It is evident from the experience gained from the three separate sequencing projects that both the *number* and *resolution* of sequence data are of paramount importance in determining the kind of research questions that can be adequately addressed. Before the start of sample collection and sequencing the desired scientific question(s) should be clearly defined and followed by careful experimental design. Putting the cart before the horse would be a poor way of conducting DENV genome sequencing projects.

The second lesson learned is that every little thing matters. During the data acquisition phase, curation skills and standard nomenclature are required to maintain the accuracy of sample information starting from the stage of sample collection, storage and subsequent lab manipulation. The assembly of sequencing chromatograms was probably the most mundane and yet most crucial task in the entire study. A great deal of effort was invested in manual editing to remove artifacts from primer sequences and to deal with less-than-perfect sequence quality. For data management and interpretation, manipulation of sequence data of such massive scale could not be accomplished reliably without the help of customised informatics solutions, which again require vigorous verification and validation in their development process. To avoid the classic problem of garbage in, garbage out, all these steps were routinely monitored for quality.

The third point to highlight is the importance of doing experimentation to verify the insights gained from sequence-based analyses. Although the lack of suitable experimentation models for the dengue disease is a serious deficiency that impedes the identification of viral genetic virulence factors, there are still a number of things that can be investigated. An example would be to examine the effect of amino acid changes on virus infectivity. The effect of amino acid mutations found in the studies of the 2004 epidemic in Jakarta, Indonesia and the 2005 epidemic in Singapore could not be performed for a number of reasons, primarily due to a lack of resources, expertise and material transfer agreements.

Frustratingly, unlike most branches of scientific research, many of the hypotheses generated in the field of molecular epidemiology are not easily amenable to controlled experimentation. Frequently, a conclusion can not be reached years or decades after an initial observation was made. The fourth lesson is that workers in the field have to be patient and be on constant lookout for natural experiments. A good example of a natural experiment, though unrelated to the topic under discussion, occurred in 2008. By imposing strict pollution control over the city of Beijing during the 2008 Olympics, the Chinese government inadvertently created a huge natural laboratory for understanding how pollution affects climate. The direct downside of such convenience sampling is that the hypothesis often has to be built around the nature of data collected.

The fifth and final point is that it is important to consider the big picture at all times. Being a vector-borne disease, dengue transmission is a complex phenomenon that involves the human host, the mosquito vector and the dengue viruses. The high degree of heterogeneity in the molecular data and the distribution of virus subtypes observed in the three studies underscores the plurality of the last component. Although not within the scope of this project, a multitude of factors affect the human host and mosquito vector and contribute to the transmission and maintenance of the dengue disease. In future studies it would be prudent to consider molecular data concurrently with

clinical, demographic and epidemiological data, both at the planning and data interpretation phase.

In spite of the limitations highlighted, a lot about the history and evolutionary dynamics of DENV in the three Southeast Asian countries has been learned from this study. The three sequencing studies all differed in important aspects which allowed different kinds of questions to be asked, and many pertinent lessons have been learned that would enable future studies to be better designed for understanding the molecular epidemiology of the dengue virus.

6.6 References

1. A-Nuegoonpipat A, Berlioz-Arthaud A, Chow V, Endy T, Lowry K, Mai le Q, Ninh TU, Pyke A, Reid M, Reynes JM, Su Yun ST, Thu HM, Wong SS, Holmes EC and Aaskov J (2004) Sustained transmission of dengue virus type 1 in the Pacific due to repeated introductions of different Asian strains. *Virology*. 329:505–512.
2. Brault AC, Huang CY, Langevin SA, Kinney RM, Bowen RA, Ramey WN, Panella NA, Holmes EC, Powers AM and Miller BR (2007) A single positively selected West Nile viral mutation confers increased virogenesis in American crows. *Nat Genet*. 39(9):1162–1166.
3. Chungue E, Deubel V, Cassar O, Laille M and Martin PMV (1993) Molecular epidemiology of dengue 3 viruses and genetic relatednes among dengue 3 strains isolated from patients with mild or severe form of dengue fever in French Polynesia. *J Gen Virol*. 74:2765–2770.
4. Cummings DAT, Irizarry RA, Huang NE, Endy TP, Nisalak A, Ungchusak K and Burke DS (2004) Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature*. 427:344–347.
5. Durbin AP, Karron RA, Sun W, Vaughn DW, Reynolds MJ, Perreault JR, Thumar B, Men R, Lai CJ, Elkins WR, Chanock RM, Murphy BR and Whitehead SS (2001) Attenuation and immunogenicity in humans of a live dengue virus type-4 vaccine candidate with a 30 nucleotide deletion in its 3'-untranslated region. *Am J Trop Med Hyg*. 65(5):405–413.

6. Figueiredo LB, Cecílio AB, Ferreira GP, Drumond BP, de Oliveira JG, Bonjardim CA, Ferreira PCP and Kroon EG (2008) Dengue virus 3 genotype 1 associated with dengue fever and dengue hemorrhagic fever, Brazil. *Emerg Infect Dis*. 14(2):314-316.
7. Fong MY, Yusup R, Yusof R and Lam SK (2004) Neurovirulence of four encephalitogenic dengue 3 virus strains isolated in Malaysia (1992-1994) is not attributed to their envelope protein. *Trans R Soc Trop Med Hyg.* 98(6):379–381.
8. Kobayashi N, Thayan R, Sugimoto C, Oda K, Saat Z, Vijayamalar B, Sinniah M and Igarashi A (1999) Type-3 dengue viruses responsible for the dengue epidemic in Malaysia during 1993-1994. *Am J Trop Med Hyg*. 60(6):904–909.
9. Leitmeyer KC, Vaughn DW, Watts DM, Salas R, de Chacon IR, Ramos C and Rico-Hesse R (1999) Dengue virus structural differences that correlate with pathogenesis. *J. Virol*. 73(6):4738–4747.
10. Lodeiro MF, Filomatori CV and Gamarnik AV (2008). Structural and functional studies of the promoter element for dengue virus RNA replication. *J Virol*. 83(2):993–1008.
11. Low JG, Ooi EE, Tolfvenstam T, Leo YS, Hibberd ML, Ng LC, Lai YL, Yap GS, Li CS, Vasudevan SG and Ong A (2006) Early Dengue infection and outcome study (EDEN) - study design and preliminary findings. *Ann Acad Med Singapore*. 35(11):783–789.
12. Morens DM, Folkers GK and Fauci AS (2004) The challenge of emerging and re-emerging infectious diseases. *Nature*. 430:242–249.
13. Pandey BD and Igarashi A (2004) Severity-related molecular differences among nineteen strains of dengue type 2 viruses. *Microbiol Immunol*. 44(3):179–88.
14. Reid AH, Janczewski TA, Lourens RM, Elliot AJ, Daniels RS, Berry CL, Oxford JS and Taubenberger JK (2003) 1918 influenza pandemic caused by highly conserved viruses with two receptor-binding variants. *Emerg. Infect. Dis*. 9(10):1249–1253.
15. Rico-Hesse (1990) Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology*. 174:479–493.

16. Schreiber MJ, Holmes EC, Ong SH, Soh HS, Liu W, Tanner L, Aw PP, Tan HC, Ng LC, Leo YS, Low JG, Ong A, Ooi EE, Vasudevan SG and Hibberd ML (2009) Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J Virol*. 83(9):4163–4173.
17. Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, Lavenir R, Pardigon N, Reynes JM, Pettinelli F, Biscornet L, Diancourt L, Michel S, Duquerroy S, Guigon G, Frenkiel MP, Bréhin AC, Cubito N, Desprès P, Kunst F, Rey FA, Zeller H and Brisse S (2006) Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med*. 3(7):e263.
18. Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE and Fanning TG (1997) Initial genetic characterization of the 1918 “Spanish” influenza virus. *Science*. 275:1793–1796.
19. Twiddy SS, Woelk CH and Holmes EC (2002) Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *J Gen Virol*. 83(7):1679–1689.
20. Zhang C, Mammen MP Jr, Chinnawirotpisan P, Klungthong C, Rodpradit P, Monkongdee P, Nimmannitya S, Kalayanarooj S and Holmes EC (2005) Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. *J Virol*. 79(24):15123–15130.

Appendix A

The SQL file for the structure of the dengue genome database (without the data) generated by the m*ysqldump* tool.

```
-- MySQL dump 10.11

--

-- Host: localhost    Database: dbflavivirus

-- ------------------------------------------------------

-- Server version           5.0.67-community-nt


/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;

/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;

/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;

/*!40101 SET NAMES utf8 */;

/*!40103 SET @OLD_TIME_ZONE=@@TIME_ZONE */;

/*!40103 SET TIME_ZONE='+00:00' */;

/*!40014 SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0 */;

/*!40014 SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS,
FOREIGN_KEY_CHECKS=0 */;

/*!40101 SET @OLD_SQL_MODE=@@SQL_MODE,
SQL_MODE='NO_AUTO_VALUE_ON_ZERO' */;

/*!40111 SET @OLD_SQL_NOTES=@@SQL_NOTES, SQL_NOTES=0 */;


--

-- Table structure for table `tblannotation`

--

DROP TABLE IF EXISTS `tblannotation`;

SET @saved_cs_client     = @@character_set_client;

SET character_set_client = utf8;
```

```
CREATE TABLE `tblannotation` (
  `annotation_accession` varchar(12) NOT NULL default '',
  `annotation_species` tinyint(2) NOT NULL default '0',
  `annotation_strain` varchar(40) NOT NULL default 'n/a',
  `annotation_nature` varchar(20) NOT NULL default 'n/a',
  `annotation_severity` varchar(10) default 'n/a',
  `annotation_country` varchar(45) NOT NULL default 'n/a',
  `annotation_year` year(4) default NULL,
  `annotation_collection` varchar(25) default 'n/a',
  `annotation_subtype` varchar(20) default 'n/a',
  `annotation_date` date default NULL,
  `annotation_remarks` mediumtext,
  PRIMARY KEY  (`annotation_accession`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
SET character_set_client = @saved_cs_client;


--
-- Table structure for table `tblcountry`
--
DROP TABLE IF EXISTS ` tblcountry`;
SET @saved_cs_client     = @@character_set_client;
SET character_set_client = utf8;
CREATE TABLE `tblcountry` (
  `country_iso` char(2) NOT NULL,
  `country_name` varchar(80) NOT NULL,
  `country_printable_name` varchar(80) NOT NULL,
  `country_iso2` char(2) default NULL,
```

```
  `country_iso3` char(3) default NULL,
  `country_numcode` smallint(6) default NULL,
  PRIMARY KEY  (`country_iso`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
SET character_set_client = @saved_cs_client;

--
-- Table structure for table `tblnucleotide`
--
DROP TABLE IF EXISTS `tblnucleotide`;
SET @saved_cs_client     = @@character_set_client;
SET character_set_client = utf8;
CREATE TABLE `tblnucleotide` (
  `nucleotide_accession` varchar(12) NOT NULL default '',
  `nucleotide_bases` mediumtext NOT NULL,
  PRIMARY KEY  (`nucleotide_accession`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
SET character_set_client = @saved_cs_client;

--
-- Table structure for table `tblspecies`
--
DROP TABLE IF EXISTS `tblspecies`;
SET @saved_cs_client     = @@character_set_client;
SET character_set_client = utf8;
CREATE TABLE `tblspecies` (
  `species_id` int(11) NOT NULL auto_increment,
```

```
  `species_name` varchar(30) default NULL,
  `species_abbrev` varchar(8) default NULL,
  `species_txid` varchar(8) default NULL,
  PRIMARY KEY  (`species_id`)
) ENGINE=InnoDB AUTO_INCREMENT=10 DEFAULT CHARSET=latin1;
SET character_set_client = @saved_cs_client;

/*!40103 SET TIME_ZONE=@OLD_TIME_ZONE */;
/*!40101 SET SQL_MODE=@OLD_SQL_MODE */;
/*!40014 SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS */;
/*!40014 SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS */;
/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;
/*!40111 SET SQL_NOTES=@OLD_SQL_NOTES */;
```

Appendix B

Binding positions of RT-PCR primers used in Chapter 3: Periodic re-emergence of endemic strains with strong epidemic potential—A proposed explanation for the 2004 Indonesian dengue epidemic.

Primers for DENV-1	Binding Positions (sense strand)
d1a5	8558-8577
d1a9	6551-6573
d1a13	4544-4561
d1a17	2540-2559
d1a23	10716-10735
d1s6	2201-2223
d1s10	4213-4231
d1s14	6216-6235
d1s18	8211-8232
d1s22	1 to 20

Primers for DENV-2	Binding Positions (sense strand)
d2a6	8468-8488
d2a10	6477-6497
d2a14	4461-4484
d2a18	2455-2474
d2a23	10704-10723
d2s5	2182-2201
d2s9	4175-4197
d2s13	6193-6213
d2s16	7669-7692
d2s23	1 to 20

Primers for DENV-3	Binding Positions (sense strand)
d3a6	8342 – 8361
d3a10	6339-6360
d3a14	4334-4356
d3a18	2361-2380
d3a23	10688-10707
d3s5	2035-2053
d3s8	3532-3553
d3s13	6032-6053
d3s17	8025-8046

d3s23	1 to 20

Primers for DENV-3	Binding Positions (sense strand)
AF1	109-131
AF5	1927-1949
BF1	3629-3652
CF1	7111-1734
AR1	2708-2727
BR1	4133-4153
CR1	7664-7687
DR1	10606-10637

Appendix C

Sequencing and amplification primers used in Chapter 3: Periodic re-emergence of endemic strains with strong epidemic potential—A proposed explanation for the 2004 Indonesian dengue epidemic.

Primer name	Sequence	Serotype
d1a1	5'- ACAGCTTCCCCTGGTGTTGG-3'	DENV-1
d1a2	5'- DTCTTCCCAACTGGAYACATG-3'	DENV-1
d1a3	5'- YACRCARTCATCTCCRCTGAT-3'	DENV-1
d1a4	5'- CACTCCACTGAGTGAATTCTCTCT-3'	DENV-1
d1a5	5'- GGRATRACATCCCATGGTTT-3'	DENV-1
d1a6	5'- AGRACACGTAACGTTCTWCCTTC-3'	DENV-1
d1a7	5'- CCTACCTCCTCCTARAGATTTCA-3'	DENV-1
d1a8	5'- CAAGTCCCATCAATATAGCTGC-3'	DENV-1
d1a9	5'- CCAGTYARCACAGCTATCAAAGC-3'	DENV-1
d1a10	5'- TCTCTCYGGCTCAAAGAGGG-3'	DENV-1
d1a11	5'- CRTAGCCTGARTTCCATGATCT-3'	DENV-1
d1a12	5'- CCTCGTCCTCAATCTCTGGTAG-3'	DENV-1
d1a13	5'- TTCCACTTCYGGAGGGCT-3'	DENV-1
d1a14	5'- CCGGAAGCCATGTTGTTTT-3'	DENV-1
d1a15	5'- GCATYTTTCTRCTCCATCTGGATC-3'	DENV-1
d1a16	5'- CARCTTCCARGTYTCGTTCTT-3'	DENV-1
d1a17	5'- CCAATGGCYGCTGAYAGTCT-3'	DENV-1
d1a18	5'- AAAGGTGGYTCYGYYTCAAT-3'	DENV-1
d1a19	5'- GTTTGTGGACRAGCCATGATT-3'	DENV-1
d1a20	5'- CGTCTTCAAGAGTTCAATGTCC-3'	DENV-1
d1a21	5'- CATYGCAATRAGRGTGCACAT-3'	DENV-1
d1a22	5'- AGCTTCCGATTCGAAACTGT-3'	DENV-1
d1a23	5'- AGAACCTGTTGATTCAACAG-3'	DENV-1
d1s1	5'- TRGCTCCATCGTGGGGAT-3'	DENV-1
d1s2	5'- TTGCTYTCAGGCCAAGGACC-3'	DENV-1
d1s3	5'- AAACGTTCCGTSGCACTGGC-3'	DENV-1
d1s4	5'- TGTGTGTCGMCGAACGTT-3'	DENV-1
d1s5	5'- GCAATGCACACYGCGTTG-3'	DENV-1
d1s6	5'- GGYTCTATAGGAGGRGTGTTCAC-3'	DENV-1
d1s7	5'- GGCCCAAGGRAARAAAATG-3'	DENV-1
d1s8	5'- ACAAACAGCAGGGCCRTGGCA-3'	DENV-1
d1s9	5'- CCTAGCYYTGATGGCYACTTT-3'	DENV-1
d1s10	5'- RGCYGGSCCACTAATAGCT-3'	DENV-1
d1s11	5'- AAGAGRCTGGAACCRAGYTGGGC-3'	DENV-1
d1s12	5'- AAATGGCAGAGGCGCTCAAGGG-3'	DENV-1
d1s13	5'- ACAAAAAAYAAYGACTGGGACTAT-3'	DENV-1
d1s14	5'- ATGGRGAAAGGAACAACCAG-3'	DENV-1
d1s15	5'- GGATAGCGGCCTCYATCATACT-3'	DENV-1
d1s16	5'- GCAAARGCYACTAGAGAAGCTCAA	DENV-1
d1s17	5'- GAAACRACYAAACAYGCAGTG-3'	DENV-1
d1s18	5'- CCACYCATGAAATGTAYTGGGT-3'	DENV-1
d1s19	5'- GCCARGTGGTTATGGGGTTT-3'	DENV-1

d1s20	5'- GGATGATCTTCAGAATGAGGC-3’	DENV-1
d1s21	5'- TYATGAAGGATGGGAGGGA-3’	DENV-1
d1s22	5’- AGTTGTTAGTCTACGTGGAC-3’	DENV-1
d2a1	5'- AGGAAACGAAGGAACGCC-3’	DENV-2
d2a2	5'- ACGCCATGCGTACAGCTT-3’	DENV-2
d2a3	5'- CCGTYGTCATCCATTCATG-3’	DENV-2
d2a4	5'- TTTCTTCTGTGRCTGTCAGGTG-3’	DENV-2
d2a5	5'- TCTGCTGCCTTTTGCCTT-3’	DENV-2
d2a6	5'- CATGGTAWGCCCAYGTTTTGT-3’	DENV-2
d2a7	5'- TTCTGGCGGRRTGAAGAA-3’	DENV-2
d2a8	5'- TGACACYGCAATGGTAGTGTT-3’	DENV-2
d2a9	5'- CAATGCTATGTCTCARCATTGGTGT-3’	DENV-2
d2a10	5'- TACGCCCTTCCRCCTGCTTCA-3’	DENV-2
d2a11	5'- CCAGTGTGCACAGTCTTCATCAT-3’	DENV-2
d2a12	5'- ATGGRTCTCTRCTTCCCGG-3’	DENV-2
d2a13	5'- CACCATTACCATAAAGACCCAC-3’	DENV-2
d2a14	5'- GCCGTGATTGGTATTGATACAGGA-3’	DENV-2
d2a15	5'- GTGCAACTCACTTTCCATGC-3’	DENV-2
d2a16	5'- CGGCTGTGACCAAGGAGTT-3’	DENV-2
d2a17	5'- CCGCTGACATGAGTTTTGAGTC-3’	DENV-2
d2a18	5'- CCACTGCCACATTTCAGTTC-3’	DENV-2
d2a19	5'- GGCGRCCTAAGACATRTCTTTT-3’	DENV-2
d2a20	5'- GCCATARCCTGTCARTTCTGC-3’	DENV-2
d2a21	5'- CTGAAACCCCTTCTACAAAGTCTC-3’	DENV-2
d2a22	5'- TGTGGTTCTCCGTTACGTGT-3’	DENV-2
d2a23	5’- AGAACCTGTTGATTCAACAG-3’	DENV-2
d2s1	5'- GCAACAGCTGACAAAGAGATTCTC-3’	DENV-2
d2s2	5'- CACCACRGGAGAACAYAGAAGA-3’	DENV-2
d2s3	5'- CAGCCTAAAWGAAGAGCAGGA-3’	DENV-2
d2s4	5'- GCGAAGAAACAGGATGTTGTTG-3’	DENV-2
d2s5	5'- GGTGACACAGCCTGGGATTT-3’	DENV-2
d2s6	5'- YATGACAGGAGACATCAAAGGA-3’	DENV-2
d2s7	5'- WCAACACAACTAYAGACCAGGCT-3’	DENV-2
d2s8	5'- TGGGCGTGACTTATCTTGC-3’	DENV-2
d2s9	5'- GCATTTTRGCCAGTTCTCTCCTA-3’	DENV-2
d2s10	5'- GYGCTGTYCTAATGCATAAAGG-3’	DENV-2
d2s11	5'- YAGAGTCGTGGCAGCTGAA-3’	DENV-2
d2s12	5'- GGAAGACYTTTGATTCTGAGTATGT-3’	DENV-2
d2s13	5'- GCAGACAGAAGGTGGTGTTTT-3’	DENV-2
d2s14	5'- CCACACTGGATAGCAGCTTCAATA-3’	DENV-2
d2s15	5'- GACTYCAAGCAAAAGCAACC-3’	DENV-2
d2s16	5'- CAGGAAGTGGATAGAACCTTAGCA-3’	DENV-2
d2s17	5'- CTCTCACGRAACTCCACACAT-3’	DENV-2
d2s18	5'- RGCAGAGTGGCTKTGGAAA-3’	DENV-2
d2s19	5'- GGGACACAAGAATCACACTAGAAG-3’	DENV-2
d2s20	5'- GCCYTTYTGTTCACACCATTTCCA-3’	DENV-2
d2s21	5'- AGGAATACACAGATTACATGCCA-3’	DENV-2
d2s22	5'- GGAATGGTGCTGTTGAATCAAC-3’	DENV-2
d2s23	5’- AGTWGTTAGTCTACGTGGAC-3’	DENV-2
d3a1	5'- GGTTTCTCACGCGTTTCAG-3’	DENV-3
d3a2	5'- TTTTAACGTCCTTGGACGG-3’	DENV-3
d3a3	5'- GGATGCTAGTCTRAGATCTCTTCTG-3’	DENV-3
d3a4	5'- CTGCCTCTTTGGTCTTTCCT-3’	DENV-3

d3a5	5'- CGTTCTCTGTCCACAAGTTTCC-3’	DENV-3
d3a6	5'- GCATTRACATGTCGRGTTCC-3’	DENV-3
d3a7	5'- TCCTCGCACTTCTGTRACTTT-3’	DENV-3
d3a8	5'- TTGAACTGCACACARAACCAG-3’	DENV-3
d3a9	5'- CACCTGGYTCYTTAGACATTCCTA-3’	DENV-3
d3a10	5'- GCYGCAAARTCCTTGAATTCCT-3’	DENV-3
d3a11	5'- TTGGTCCAGCCAGGATCA-3’	DENV-3
d3a12	5'- GTGAAATGRGCCTCATCCAT-3’	DENV-3
d3a13	5'- CCTGGCATGGTTTGAAAGTT-3’	DENV-3
d3a14	5'- ACTGTGATCATTAARTTGTGGGA-3’	DENV-3
d3a15	5'- CCCCARAGCRATTCCATT-3’	DENV-3
d3a16	5'- GGCAACACCATTCGTGTATCA-3’	DENV-3
d3a17	5'- CACTTGGACACTCCGGTGT-3’	DENV-3
d3a18	5'- GATTCCTATCGCAATGCATG-3’	DENV-3
d3a19	5'- GCGTTTCKGAGACTTCTTTCTTC-3’	DENV-3
d3a20	5'- GACGGTGTATTTGAGGTTCTCA-3’	DENV-3
d3a21	5'- GGCTAGTATGGTRAACCCTGG-3’	DENV-3
d3a22	5'- CCTTCTTGAAGCCTTTYARGACCT-3’	DENV-3
d3a23	5’- AGAACCTGTTGATTCAACAG-3’	DENV-3
d3s1	5'- CAGTTTCGACTCGGAAGCTT-3’	DENV-3
d3s2	5'- CAACATGTGCACACTCATAGCC-3’	DENV-3
d3s3	5'- GACTACCATGGCTAAGAACAAGC-3’	DENV-3
d3s4	5'- GAAGAACAAAGCATGGATGGTA-3’	DENV-3
d3s5	5'- TGAACCTCCTTTTGGGGAA-3’	DENV-3
d3s6	5'- CCMAAAAGATTGGCAACAGC-3’	DENV-3
d3s7	5'- CATGGGCTATTGGATAGAAAGC-3’	DENV-3
d3s8	5'- GGTGATGAGAGGAAAATTTGGG-3’	DENV-3
d3s9	5'- GAAAACAGATTGGCTCCCAA-3’	DENV-3
d3s10	5'- CCCCCCAGAGACACAGAAAG-3’	DENV-3
d3s11	5'- CGACACCAGAGTTGGAAGAAG-3’	DENV-3
d3s12	5'- GCTCATGGAATTCAGGCAAT-3’	DENV-3
d3s13	5'- CCAGCTCTCTTTGAACCAGAAA-3’	DENV-3
d3s14	5'- CTCYTGGGACTGATGATCTTGT-3’	DENV-3
d3s15	5'- CTGATGGGTTTRGACAAAGGA-3’	DENV-3
d3s16	5'- TTTTTCTATYATGAAATCAGTTGGA-3’	DENV-3
d3s17	5'- CAACAGTGGAAGAAAGCAGAAC-3’	DENV-3
d3s18	5'- ACAAAACCATGGGATGTGG-3’	DENV-3
d3s19	5'- CTGGTTCTCGCGTGAAAAC-3’	DENV-3
d3s20	5'- GGGATGATTGCGTAGTGAAA-3’	DENV-3
d3s21	5'- TCCAGTCACAACGTGGGAA-3’	DENV-3
d3s22	5'- TGTACCTCCTTGCAAAGGACTA-3’	DENV-3
d3s23	5’- AGTTGTTAGTCTACGTGGAC-3’	DENV-3
AR1	5’-CTTGCCTTYGGTCAACACCC-3’	DENV-4
BR1	5’-CCTCGTTAAGRGGCCARGATC-3’	DENV-4
CR1	5’-GGCTTCAGTCCTGTCCACTTCTAG-3’	DENV-4
DR1	5’-ATCCATCTTGCGGCGCTCTGTG-3’	DENV-4

Appendix D

Supplementary Tables described in Chapter 4: Genomic Epidemiology of a Dengue Virus Epidemic in Urban Singapore.

Supplementary Table 1: List of reverse transcription primers.

Fragment	Primer	Sequence	Position
DENV-1			
F1	d1a17	CCAATGGCYGCTGAYAGTCT	2540-2559
F2	d1a13a	CCTACTTGGGACCTGCCCA	4611-4629
	d1a13b	GGGACCTGCCCAACAGTCCT	4603-4622
	d1a13c	TGCCCAACAGTCCTCTCTGC	4597-4616
F3	d1a9	CCAGTYARCACAGCTATCAAAGC	6551-6573
F4	d1a4	CACTCCACTGAGTGAATTCTCTCT	9068-9091
F5	d1a23	AGAACCTGTTGATTCAACAG	10716-10735
DENV-3			
F1	d3a18	GATTCCTATCGCAATGCATG	2361-2380
F2	d3a14	ACTGTGATCATTAARTTGTGGGA	4334-4356
F3	d3a10	GCYGCAAARTCCTTGAATTCCT	6339-6360
F4	d3a6	GCATTRACATGTCGRGTTCC	8342-8361
F5	d3a23	AGAACCTGTTGATTCAACAG	10688-10707

Supplementary Table 2: Reverse Transcription reaction conditions.

Step	Temperature (°C)	Time
1	65	5 min
2	0 (ice)	1 min
3	25	5 min
4	42	60 min
5	70	15 min
6	4	Hold

Supplementary Table 3: List of PCR primers.

Fragment	Primer	Sequence	Position
DENV-1			
F1	d1a17	CCAATGGCYGCTGAYAGTCT	2540-2559
	d1s22	AGTTGTTAGTCTACGTGGAC	1-20
F2	d1a13	TTCCACTTCYGGAGGGCT	4544-4561
	d1s6	GGYTCTATAGGAGGRGTGTTCAC	2201-2223
F3	d1a9	CCAGTYARCACAGCTATCAAAGC	6551-6573
	d1s10	RGCYGGSCCACTAATAGCT	4213-4231
F4	d1a4	CACTCCACTGAGTGAATTCTCTCT	9068-9091
	d1s14	ATGGRGAAAGGAACAACCAG	6216-6235
F5	d1a23	AGAACCTGTTGATTCAACAG	10716-10735
	d1s18c	GGTGAGTCCTCTCCAAACCC	8015-8034
DENV-3			
F1	d3a18	GATTCCTATCGCAATGCATG	2361-2380
	d3s23	AGTTGTTAGTCTACGTGGAC	1-20
F2	d3a14	ACTGTGATCATTAARTTGTGGGA	4334-4356
	d3s5	TGAACCTCCTTTTGGGGAA	2035-2053
F3	d3a10	GCYGCAAARTCCTTGAATTCCT	6339-6360
	d3s9	GAAAACAGATTGGCTCCCAA	4030-4049
F4	d3a6	GCATTRACATGTCGRGTTCC	8342-8361
	d3s13	CCAGCTCTCTTTGAACCAGAAA	6032-6053
F5	d3a23	AGAACCTGTTGATTCAACAG	10688-10707
	d3s17	CAACAGTGGAAGAAAGCAGAAC	8025-8046

Supplementary Table 4: PCR Program

Serotype	Fragment	Cycles	Initial	Cycle			Extend	Hold
DENV-1	F1	35	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F2	35	95°C 2 min	95°C 30 sec	58°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F3	35	95°C 2 min	95°C 30 sec	63.1°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F4	35	95°C 2 min	95°C 30 sec	58°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F5	40	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 4 min	72°C 2 min	4°C
DENV-3	F1	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F2	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F3	45	95°C 2 min	95°C 30 sec	63.1°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F4	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F5	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C

Supplementary Table 5: DENV Isolates Used in this Study

GenBank Accession	Standardized Strain Name
A75711	DENV-1/SG/S275-90/1990
AB074760	DENV-1/JP/Mochizuki/1943
AB074761	DENV-1/ID/A88/1988
AB178040	DENV-1/JP/20-Feb/XXXX
AB189120	DENV-1/ID/98901518 DHF DV-1/1998
AB189121	DENV-1/ID/98901530 DF DV-1/1998
AB195673	DENV-1/SC/D1-hu-Seychelles-NIID41-2003/2003
AB204803	DENV-1/JP/D1-hu-Yap-NIID27-2004/2004
AF180817	DENV-1/TH/16007/1964
AF180818	DENV-1/XX/PDK-13/XXXX
AF226685	DENV-1/BR/BR-90/1990
AF226686	DENV-1/XX/FGA-NA d1d/XXXX
AF226687	DENV-1/GF/FGA-89/1989
AF298807	DENV-1/CI/DI-H-IMTSSA-ABID-99-1056/1998
AF298808	DENV-1/DJ/DI-H-IMTSSA-DJIB-98-606/1998
AF309641	DENV-1/KH/DI-H-IMTSSA-CAMB-98-658/1998
AF311956	DENV-1/BR/BR-97-111/1997
AF311957	DENV-1/BR/BR-97-409/1997
AF311958	DENV-1/BR/BR-97-233/1997
AF350498	DENV-1/CN/GZ-80/1980

AF513110	DENV-1/BR/BR-01-MR/2001
AF514876	DENV-1/AR/301arg00/2000
AF514878	DENV-1/PY/280par00/2000
AF514883	DENV-1/PY/259par00/2000
AF514885	DENV-1/AR/295arg00/2000
AF514889	DENV-1/AR/297arg00/2000
AY145121	DENV-1/XX/Western Pacific (WP)/XXXX
AY145122	DENV-1/XX/rDEN1mutF/XXXX
AY145123	DENV-1/XX/rDEN1 30/XXXX
AY206457	DENV-1/AR/293arg00/2000
AY277664	DENV-1/AR/ARG9920/1999
AY277665	DENV-1/AR/ARG0028/2000
AY277666	DENV-1/AR/ARG0048/2000
AY373427	DENV-1/CN/GD23-95/1995
AY376737	DENV-1/CN/GD14-97/1997
AY376738	DENV-1/CN/GD05-99/1999
AY708047	DENV-1/MM/D1.Myanmar.059-01/2001
AY713473	DENV-1/MM/D1.Myanmar.40553-71/1971
AY713474	DENV-1/MM/D1.Myanmar.194-01/2001
AY713475	DENV-1/MM/D1.Myanmar.206-01/2001
AY713476	DENV-1/MM/D1.Myanmar.305-01/2001
AY722801	DENV-1/MM/D1.Myanmar.40568-76/1976

AY722802	DENV-1/MM/D1.Myanmar.23819-96/1996
AY722803	DENV-1/MM/D1.Myanmar.32514-98/1998
AY726549	DENV-1/MM/D1.Myanmar.37726-01/2001
AY726550	DENV-1/MM/D1.Myanmar.38862-01/2001
AY726551	DENV-1/MM/D1.Myanmar.44168-01/2001
AY726552	DENV-1/MM/D1.Myanmar.44988-02/2002
AY726553	DENV-1/MM/D1.Myanmar.49440-02/2002
AY726554	DENV-1/MM/D1.Myanmar.31987-98/1998
AY726555	DENV-1/MM/D1.Myanmar.31459-98/1998
AY732474	DENV-1/TH/ThD1_0673_80/1980
AY732475	DENV-1/TH/ThD1_0488_94/1994
AY732476	DENV-1/TH/ThD1_0442_80/1980
AY732477	DENV-1/TH/ThD1_0336_91/1991
AY732478	DENV-1/TH/ThD1_0323_91/1991
AY732479	DENV-1/TH/ThD1_0102_01/2001
AY732480	DENV-1/TH/ThD1_0097_94/1994
AY732481	DENV-1/TH/ThD1_0081_82/1982
AY732482	DENV-1/TH/ThD1_0049_01/2001
AY732483	DENV-1/TH/ThD1_0008_81/1981
AY762084	DENV-1/SG/Singapore 8114-93/1993
CS477263	DENV-1/XX/WO2006134433-1/XXXX
CS477264	DENV-1/XX/WO2006134433-2/XXXX

CS477265	DENV-1/XX/WO2006134433-3/XXXX
CS479203	DENV-1/XX/WO2006134443-39/XXXX
CS479204	DENV-1/XX/WO2006134443-40/XXXX
M87512	DENV-1/SG/S275-90/1990
DQ193572	DENV-1/CN/Fj231-04/2004
DQ285558	DENV-1/RE/Reunion 185-04/2004
DQ285559	DENV-1/RE/Reunion 191-04/2004
DQ285560	DENV-1/RE/Reunion 257-04/2004
DQ285561	DENV-1/SC/Seychelles 1480-04/2004
DQ285562	DENV-1/KM/Comoros 04.329-93/1993
DQ672556	DENV-1/FP/FP0203/XXXX
DQ672557	DENV-1/FP/FP0705/XXXX
DQ672558	DENV-1/FP/FP0908/XXXX
DQ672559	DENV-1/FP/FP1104/XXXX
DQ672560	DENV-1/US/HawM2516/2001
DQ672561	DENV-1/US/HawM3430/2001
DQ672562	DENV-1/US/HawM2540/2001
DQ672563	DENV-1/US/HawO3758/2001
DQ672564	DENV-1/US/HawO3663/2001
U88535	DENV-1/NR/WestPac/1974
U88536	DENV-1/XX/45AZ5 PDK-0/XXXX
U88537	DENV-1/XX/45AZ5 PDK27/XXXX

EF025110	DENV-1/CN/71-02GZ/1971
EF032590	DENV-1/CN/GZ01-95/1995
EU081226	DENV-1/SG/05K814DK1/2005
EU081227	DENV-1/SG/05K847DK1/2005
EU081228	DENV-1/SG/05K872DK1/2005
EU081229	DENV-1/SG/05K2398DK1/2005
EU081230	DENV-1/SG/05K2402DK1/2005
EU081231	DENV-1/SG/05K2887DK1/2005
EU081232	DENV-1/SG/05K2895DK1/2005
EU081233	DENV-1/SG/05K2901DK1/2005
EU081234	DENV-1/SG/05K2916DK1/2005
EU081235	DENV-1/SG/05K2928DK1/2005
EU081236	DENV-1/SG/05K3297DK1/2005
EU081237	DENV-1/SG/05K3300DK1/2005
EU081238	DENV-1/SG/05K3301DK1/2005
EU081239	DENV-1/SG/05K3318DK1/2005
EU081240	DENV-1/SG/05K3886DK1/2005
EU081241	DENV-1/SG/05K3894DK1/2005
EU081242	DENV-1/SG/05K3903DK1/2005
EU081243	DENV-1/SG/05K3904DK1/2005
EU081244	DENV-1/SG/05K3905DK1/2005
EU081245	DENV-1/SG/05K3908DK1/2005

EU081246	DENV-1/SG/05K3910DK1/2005
EU081247	DENV-1/SG/05K3911DK1/2005
EU081248	DENV-1/SG/05K3915DK1/2005
EU081249	DENV-1/SG/05K3916DK1/2005
EU081250	DENV-1/SG/05K3930DK1/2005
EU081251	DENV-1/SG/05K3933DK1/2005
EU081252	DENV-1/SG/05K3934DK1/2005
EU081253	DENV-1/SG/05K3935DK1/2005
EU081254	DENV-1/SG/05K4138DK1/2005
EU081255	DENV-1/SG/05K4139DK1/2005
EU081256	DENV-1/SG/05K4140DK1/2005
EU081257	DENV-1/SG/05K4142DK1/2005
EU081258	DENV-1/SG/05K4147DK1/2005
EU081259	DENV-1/SG/05K4152DK1/2005
EU081260	DENV-1/SG/05K4154DK1/2005
EU081261	DENV-1/SG/05K4172DK1/2005
EU081262	DENV-1/SG/05K4173DK1/2005
EU081263	DENV-1/SG/05K4174DK1/2005
EU081264	DENV-1/SG/05K4175DK1/2005
EU081265	DENV-1/SG/05K4183DK1/2005
EU081266	DENV-1/SG/05K4441DK1/2005
EU081267	DENV-1/SG/05K4443DK1/2005

EU081268	DENV-1/SG/05K4468DK1/2005
EU081269	DENV-1/SG/05K4479DK1/2005
EU081270	DENV-1/SG/05K4480DK1/2005
EU081271	DENV-1/SG/05K4604DK1/2005
EU081272	DENV-1/SG/05K4605DK1/2005
EU081273	DENV-1/SG/05K4606DK1/2005
EU081274	DENV-1/SG/05K4609DK1/2005
EU081275	DENV-1/SG/05K4620DK1/2005
EU081276	DENV-1/SG/05K4621DK1/2005
EU081277	DENV-1/SG/05K4622DK1/2005
EU081278	DENV-1/SG/05K4632DK1/2005
EU081279	DENV-1/SG/05K4820DK1/2005
EU081280	DENV-1/SG/06K2236DK1/2006
EU081281	DENV-1/SG/06K2290DK1/2006
FJ469907	DENV-1/SG/S210Y03/2003
FJ469908	DENV-1/SG/S210Y03/2003
FJ469909	DENV-1/SG/S210Y03/2003
AB122020	DENV-2/DO/DR23-01/2001
AB122021	DENV-2/DO/DR31-01/2001
AB122022	DENV-2/DO/DR59-01/2001
AB189122	DENV-2/ID/98900663 DHF DV-2/1998
AB189123	DENV-2/ID/98900665 DF DV-2/1998

AB189124	DENV-2/ID/98900666 DSS DV-2/1998
AF022434	DENV-2/TH/ThNH-7-93/1993
AF022435	DENV-2/TH/ThNH-28-93/1993
AF022436	DENV-2/TH/ThNH-52-93/1993
AF022437	DENV-2/TH/ThNH-p11-93/1993
AF022438	DENV-2/TH/ThNH-p12-93/1993
AF022439	DENV-2/TH/ThNH-p14-93/1993
AF022440	DENV-2/TH/ThNH-p16-93/1993
AF022441	DENV-2/TH/ThNH-p36-93/1993
AF038402	DENV-2/XX/MON310/XXXX
AF038403	DENV-2/XX/MON501/XXXX
AF100459	DENV-2/TH/K0008/1994
AF100460	DENV-2/TH/K0010/1994
AF100461	DENV-2/TH/C0371/1995
AF100462	DENV-2/TH/C0390/1995
AF100463	DENV-2/TH/C0166/1996
AF100464	DENV-2/TH/C0167/1996
AF100465	DENV-2/VE/Ven2/1987
AF100466	DENV-2/VE/Mara4/1990
AF100467	DENV-2/PE/IQT1797/1995
AF100468	DENV-2/PE/IQT2913/1995
AF100469	DENV-2/MX/O131/1992

AF119661	DENV-2/CN/China 04/1985
AF169678	DENV-2/TH/ThNH29-93/1993
AF169679	DENV-2/TH/ThNH36-93/1993
AF169680	DENV-2/TH/ThNH45-93/1993
AF169681	DENV-2/TH/ThNH55-93/1993
AF169682	DENV-2/TH/ThNH54-93/1993
AF169683	DENV-2/TH/ThNH62-93/1993
AF169684	DENV-2/TH/ThNH63-93/1993
AF169685	DENV-2/TH/ThNH69-93/1993
AF169686	DENV-2/TH/ThNH73-93/1993
AF169687	DENV-2/TH/ThNH76-93/1993
AF169688	DENV-2/TH/ThNH81-93/1993
AF204177	DENV-2/CN/44/1989
AF204178	DENV-2/CN/43/1987
AF208496	DENV-2/MQ/DEN2-H-IMTSSA-MART-98-703/1998
AF276619	DENV-2/CN/FJ-10/1999
AF359579	DENV-2/CN/FJ11-99/1999
AF489932	DENV-2/BR/BR64022/1998
AJ968413	DENV-2/TW/PL046/XXXX
AY037116	DENV-2/AU/TSV01/1993
AY702034	DENV-2/CU/Cuba13-97/1997
AY702035	DENV-2/CU/Cuba58-97/1997

AY702036	DENV-2/CU/Cuba115-97/1997
AY702037	DENV-2/CU/Cuba89-97/1997
AY702038	DENV-2/CU/Cuba165-97/1997
AY702039	DENV-2/CU/Cuba205-97/1997
AY702040	DENV-2/CO/I348600/1986
AY744147	DENV-2/TO/Tonga-74/1974
AY776328	DENV-2/TW/Taiwan-1008DHF/XXXX
AY858035	DENV-2/ID/BA05i/2004
AY858036	DENV-2/ID/TB61i/2004
CS477302	DENV-2/XX/WO2006134433-40/XXXX
CS477304	DENV-2/XX/WO2006134433-42/XXXX
CS479165	DENV-2/XX/WO2006134443-1/XXXX
CS479167	DENV-2/XX/WO2006134443-3/XXXX
CS479202	DENV-2/XX/WO2006134443-38/XXXX
M29095	DENV-2/PG/New Guinea-C/1944
M20558	DENV-2/JM/Jamaica-N.1409/1983
M84728	DENV-2/XX/16681-PDK53/XXXX
M84727	DENV-2/TH/16681/1964
M19197	DENV-2/PR/S1/1969
DQ181797	DENV-2/TH/ThD2_0078_01/2001
DQ181798	DENV-2/TH/ThD2_0055_99/1999
DQ181799	DENV-2/TH/ThD2_0017_98/1998

DQ181800	DENV-2/TH/ThD2_0263_95/1995
DQ181801	DENV-2/TH/ThD2_0284_90/1990
DQ181802	DENV-2/TH/ThD2_0026_88/1988
DQ181803	DENV-2/TH/ThD2_0433_85/1985
DQ181804	DENV-2/TH/ThD2_0498_84/1984
DQ181805	DENV-2/TH/ThD2_0168_79/1979
DQ181806	DENV-2/TH/ThD2_0038_74/1974
AJ487271	DENV-2/TH/Bank74/1974
EF051521	DENV-2/CN/ZS01-01/2001
EU081177	DENV-2/SG/05K3295DK1/2005
EU081178	DENV-2/SG/05K3330DK1/2005
EU081179	DENV-2/SG/05K4137DK1/2005
EU081180	DENV-2/SG/05K4155DK1/2005
U87411	DENV-2/TH/16681/1964
U87412	DENV-2/XX/PDK-53/XXXX
AB189125	DENV-3/ID/98901403 DSS DV-3/1998
AB189126	DENV-3/ID/98901437 DSS DV-3/1998
AB189127	DENV-3/ID/98901517 DHF DV-3/1998
AB189128	DENV-3/ID/98902890 DF DV-3/1998
AB214879	DENV-3/TL/D3-Hu-TL018NIID-2005/2005
AB214880	DENV-3/TL/D3-Hu-TL029NIID-2005/2005
AB214881	DENV-3/TL/D3-Hu-TL109NIID-2005/2005

AB214882	DENV-3/TL/D3-Hu-TL129NIID-2005/2005
AF317645	DENV-3/CN/80-2/XXXX
AY099336	DENV-3/LK/D3SriLa1266/2000
AY099337	DENV-3/MQ/D3Mart1243/1999
AY496871	DENV-3/BD/BDH02-01/2002
AY496873	DENV-3/BD/BDH02-03/2002
AY496874	DENV-3/BD/BDH02-04/2002
AY496877	DENV-3/BD/BDH02-07/2002
AY496879	DENV-3/PH/PhMH-J1-97/1997
AY648961	DENV-3/ID/Sleman-78/1978
AY662691	DENV-3/SG/S221-03/2003
AY676348	DENV-3/TH/ThD3_1687_98/1998
AY676349	DENV-3/TH/ThD3_1283_98/1998
AY676350	DENV-3/TH/ThD3_0104_93/1993
AY676351	DENV-3/TH/ThD3_0055_93/1993
AY676352	DENV-3/TH/ThD3_0010_87/1987
AY676353	DENV-3/TH/ThD3_0007_87/1987
AY679147	DENV-3/BR/BR74886-02/2002
AY744677	DENV-3/PF/PF89-27643/1989
AY744678	DENV-3/PF/PF89-320219/1989
AY744679	DENV-3/PF/PF90-3050/1990
AY744680	DENV-3/PF/PF90-3056/1990

AY744681	DENV-3/PF/PF90-6056/1990
AY744682	DENV-3/PF/PF92-2956/1992
AY744683	DENV-3/PF/PF92-2986/1992
AY744684	DENV-3/PF/PF92-4190/1992
AY744685	DENV-3/PF/PF94-136116/1994
AY766104	DENV-3/SG/Singapore 8120-95/1995
AY776329	DENV-3/TW/Taiwan-739079A/XXXX
AY858037	DENV-3/ID/BA51/2004
AY858038	DENV-3/ID/den3_88/1988
AY858039	DENV-3/ID/den3_98/1998
AY858040	DENV-3/ID/FW01/2004
AY858041	DENV-3/ID/FW06/2004
AY858042	DENV-3/ID/KJ30i/2004
AY858043	DENV-3/ID/KJ46/2004
AY858044	DENV-3/ID/KJ71/2004
AY858045	DENV-3/ID/PH86/2004
AY858046	DENV-3/ID/PI64/2004
AY858047	DENV-3/ID/TB16/2004
AY858048	DENV-3/ID/TB55i/2004
AY876494	DENV-3/TH/C0331-94/1994
AY923865	DENV-3/TH/C0360-94/1994
CS477305	DENV-3/XX/WO2006134433-43/XXXX

CS479205	DENV-3/XX/WO2006134443-41/XXXX
M93130	DENV-3/PH/H87/1956
DQ401689	DENV-3/BD/BDH02-01/XXXX
DQ401690	DENV-3/ID/InJ-16-82/XXXX
DQ401691	DENV-3/BD/BDH02-03/XXXX
DQ401692	DENV-3/BD/BDH02-04/XXXX
DQ401693	DENV-3/BD/BDH02-07/XXXX
DQ401694	DENV-3/ID/InJ-16-82/XXXX
DQ401695	DENV-3/PH/PhMH-J1-97/XXXX
DQ675519	DENV-3/TW/95TW466/1995
DQ675520	DENV-3/ID/98TW182/1998
DQ675521	DENV-3/TW/98TW349/1998
DQ675522	DENV-3/TW/98TW358/1998
DQ675523	DENV-3/TW/98TW360/1998
DQ675524	DENV-3/TW/98TW364/1998
DQ675525	DENV-3/TW/98TW368/1998
DQ675526	DENV-3/TW/98TW388/1998
DQ675527	DENV-3/TW/98TW390/1998
DQ675528	DENV-3/TW/98TW407/1998
DQ675529	DENV-3/TW/98TW414/1998
DQ675530	DENV-3/TW/98TW434/1998
DQ675531	DENV-3/TW/98TW503/1999

DQ675532	DENV-3/TW/98TWmosq/1998
DQ675533	DENV-3/TW/99TW628/1999
EU081181	DENV-3/SG/SS710/2004
EU081182	DENV-3/SG/05K791DK1/2005
EU081183	DENV-3/SG/05K797DK1/2005
EU081184	DENV-3/SG/05K802DK1/2005
EU081185	DENV-3/SG/05K805DK1/2005
EU081186	DENV-3/SG/05K827DK1/2005
EU081187	DENV-3/SG/05K843DK1/2005
EU081188	DENV-3/SG/05K845DK1/2005
EU081189	DENV-3/SG/05K852DK1/2005
EU081190	DENV-3/SG/05K863DK1/2005
EU081191	DENV-3/SG/05K868DK1/2005
EU081192	DENV-3/SG/05K871DK1/2005
EU081193	DENV-3/SG/05K2400DK1/2005
EU081194	DENV-3/SG/05K2406DK1/2005
EU081195	DENV-3/SG/05K2418DK1/2005
EU081196	DENV-3/SG/05K2899DK1/2005
EU081197	DENV-3/SG/05K2918DK0/2005
EU081198	DENV-3/SG/05K2933DK1/2005
EU081199	DENV-3/SG/05K3305DK1/2005
EU081200	DENV-3/SG/05K3312DK1/2005

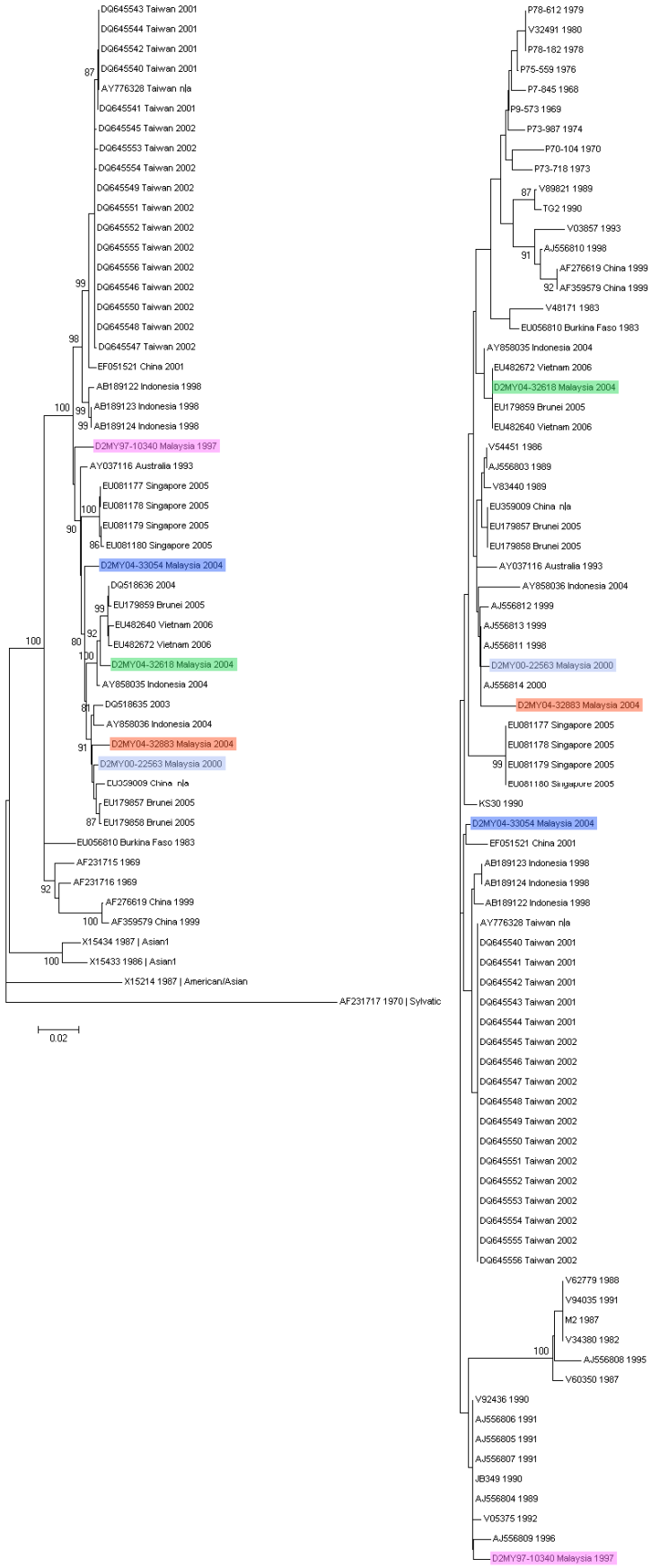
EU081201	DENV-3/SG/05K3314DK1/2005
EU081202	DENV-3/SG/05K3316DK1/2005
EU081203	DENV-3/SG/05K3324DK1/2005
EU081204	DENV-3/SG/05K3325DK1/2005
EU081205	DENV-3/SG/05K3329DK1/2005
EU081206	DENV-3/SG/05K3887DK1/2005
EU081207	DENV-3/SG/05K3897DK1/2005
EU081208	DENV-3/SG/05K3900DK1/2005
EU081209	DENV-3/SG/05K3912DK1/2005
EU081210	DENV-3/SG/05K3913DK1/2005
EU081211	DENV-3/SG/05K3923DK1/2005
EU081212	DENV-3/SG/05K3927DK1/2005
EU081213	DENV-3/SG/05K3928DK1/2005
EU081214	DENV-3/SG/05K4141DK1/2005
EU081215	DENV-3/SG/05K4144DK1/2005
EU081216	DENV-3/SG/05K4157DK1/2005
EU081217	DENV-3/SG/05K4159DK1/2005
EU081218	DENV-3/SG/05K4168DK1/2005
EU081219	DENV-3/SG/05K4176DK1/2005
EU081220	DENV-3/SG/05K4182DK1/2005
EU081221	DENV-3/SG/05K4440DK1/2005
EU081222	DENV-3/SG/05K4454DK1/2005

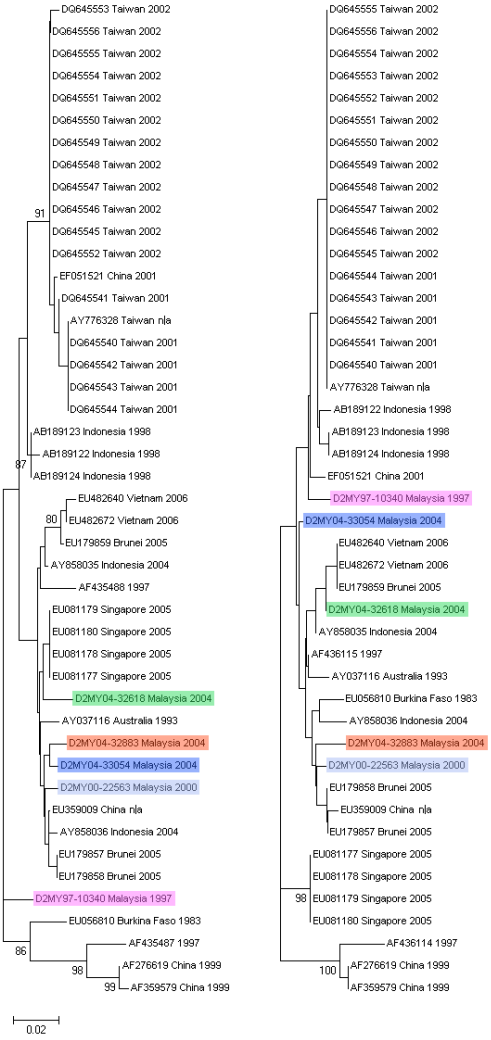
EU081223	DENV-3/SG/05K4477DK1/2005
EU081224	DENV-3/SG/05K4647DK1/2005
EU081225	DENV-3/SG/05K4648DK1/2005

Appendix E

Supplementary Figure described in Chapter 5: Inferring the History of Dengue Virus Diversity in Malaysia from Sequence Data.

Supplementary Figure 1: Montage of four neighbour-joining trees constructed from the Cosmopolitan genotype DENV-2 complete genome and partial sequences. The trees are mid-point rooted, all horizontal branches are drawn to scale, and only nodes with over 80% bootstrap support are highlighted. Complete genome sequences are labeled with their GenBank accession number, country and year of isolation. Partial sequences from Malaysia (listed in Table 5.2) are labeled only with their GenBank accession number and year of isolation. The five DENV-2 complete genomes sequenced in this study are D2MY00-22563, D2MY04-32618, D2MY04-32883, D2MY04-33054 and D2MY97-10340 and are highlighted in different colours. From left, the tree constructed from the sequence of the complete E gene (1485 nucleotides), the E/NS1 junction (240 nucleotides), partial E gene (266 nucleotides) and a shorter E/NS1 junction (225 nucleotides).


DQ645543 Taiwan 2001
DQ645544 Taiwan 2001
DQ645542 Taiwan 2001
DQ645540 Taiwan 2001
AY776328 Taiwan n/a
DQ645541 Taiwan 2001
DQ645545 Taiwan 2002
DQ645553 Taiwan 2002
DQ645554 Taiwan 2002
DQ645549 Taiwan 2002
DQ645551 Taiwan 2002
DQ645552 Taiwan 2002
DQ645555 Taiwan 2002
DQ645556 Taiwan 2002
DQ645546 Taiwan 2002
DQ645550 Taiwan 2002
DQ645548 Taiwan 2002
DQ645547 Taiwan 2002
EF051521 China 2001
AB189122 Indonesia 1998
AB189123 Indonesia 1998
AB189124 Indonesia 1998
D2MY97-10340 Malaysia 1997
AY037116 Australia 1993
EU081177 Singapore 2005
EU081178 Singapore 2005
EU081179 Singapore 2005
EU081180 Singapore 2005
D2MY04-33054 Malaysia 2004
DQ518636 2004
EU179859 Brunei 2005
EU482640 Vietnam 2006
EU482672 Vietnam 2006
D2MY04-32618 Malaysia 2004
AY858035 Indonesia 2004
DQ518635 2003
AY858036 Indonesia 2004
D2MY04-32883 Malaysia 2004
D2MY00-22563 Malaysia 2000
EU359009 China n/a
EU179857 Brunei 2005
EU179858 Brunei 2005
EU056810 Burkina Faso 1983
AF231715 1969
AF231716 1969
AF276619 China 1999
AF359579 China 1999
X15434 1987 | Asian1
X15433 1986 | Asian1
X15214 1987 | American/Asian
AF231717 1970 | Sylvatic
0.02
P78-612 1979
V32491 1980
P78-182 1978
P75-559 1976
P7-845 1968
P9-573 1969
P73-987 1974
P70-104 1970
P73-718 1973
V89821 1989
TG2 1990
V03857 1993
AJ556810 1998
AF276619 China 1999
AF359579 China 1999
V48171 1983
EU056810 Burkina Faso 1983
AY858035 Indonesia 2004
EU482672 Vietnam 2006
D2MY04-32618 Malaysia 2004
EU179859 Brunei 2005
EU482640 Vietnam 2006
V54451 1986
AJ556803 1989
V83440 1989
EU359009 China n/a
EU179857 Brunei 2005
EU179858 Brunei 2005
AY037116 Australia 1993
AY858036 Indonesia 2004
AJ556812 1999
AJ556813 1999
AJ556811 1998
D2MY00-22563 Malaysia 2000
AJ556814 2000
D2MY04-32883 Malaysia 2004
EU081177 Singapore 2005
EU081178 Singapore 2005
EU081179 Singapore 2005
EU081180 Singapore 2005
KS30 1990
D2MY04-33054 Malaysia 2004
EF051521 China 2001
AB189123 Indonesia 1998
AB189124 Indonesia 1998
AB189122 Indonesia 1998
AY776328 Taiwan n/a
DQ645540 Taiwan 2001
DQ645541 Taiwan 2001
DQ645542 Taiwan 2001
DQ645543 Taiwan 2001
DQ645544 Taiwan 2001
DQ645545 Taiwan 2002
DQ645546 Taiwan 2002
DQ645547 Taiwan 2002
DQ645548 Taiwan 2002
DQ645549 Taiwan 2002
DQ645550 Taiwan 2002
DQ645551 Taiwan 2002
DQ645552 Taiwan 2002
DQ645553 Taiwan 2002
DQ645554 Taiwan 2002
DQ645555 Taiwan 2002
DQ645556 Taiwan 2002
V62779 1988
V94035 1991
M2 1987
V34380 1982
AJ556808 1995
V60350 1987
V92436 1990
AJ556806 1991
AJ556805 1991
AJ556807 1991
JB349 1990
AJ556804 1989
V05375 1992
AJ556809 1996
D2MY97-10340 Malaysia 1997


DQ645553 Taiwan 2002
DQ645556 Taiwan 2002
DQ645555 Taiwan 2002
DQ645554 Taiwan 2002
DQ645551 Taiwan 2002
DQ645550 Taiwan 2002
DQ645549 Taiwan 2002
DQ645548 Taiwan 2002
DQ645547 Taiwan 2002
91
DQ645546 Taiwan 2002
DQ645545 Taiwan 2002
DQ645552 Taiwan 2002
EF051521 China 2001
DQ645541 Taiwan 2001
AY776328 Taiwan n/a
DQ645540 Taiwan 2001
DQ645542 Taiwan 2001
DQ645543 Taiwan 2001
DQ645544 Taiwan 2001
AB189123 Indonesia 1998
AB189122 Indonesia 1998
87
AB189124 Indonesia 1998
EU482640 Vietnam 2006
80
EU482672 Vietnam 2006
EU179859 Brunei 2005
AY858035 Indonesia 2004
AF435488 1997
EU081179 Singapore 2005
EU081180 Singapore 2005
EU081178 Singapore 2005
EU081177 Singapore 2005
D2MY04-32618 Malaysia 2004
AY037116 Australia 1993
D2MY04-32883 Malaysia 2004
D2MY04-33054 Malaysia 2004
D2MY00-22563 Malaysia 2000
EU359009 China n/a
AY858036 Indonesia 2004
EU179857 Brunei 2005
EU179858 Brunei 2005
D2MY97-10340 Malaysia 1997
EU056810 Burkina Faso 1983
86
AF435487 1997
98
AF276619 China 1999
99
AF359579 China 1999
0.02
DQ645555 Taiwan 2002
DQ645556 Taiwan 2002
DQ645554 Taiwan 2002
DQ645553 Taiwan 2002
DQ645552 Taiwan 2002
DQ645551 Taiwan 2002
DQ645550 Taiwan 2002
DQ645549 Taiwan 2002
DQ645548 Taiwan 2002
DQ645547 Taiwan 2002
DQ645546 Taiwan 2002
DQ645545 Taiwan 2002
DQ645544 Taiwan 2001
DQ645543 Taiwan 2001
DQ645542 Taiwan 2001
DQ645541 Taiwan 2001
DQ645540 Taiwan 2001
AY776328 Taiwan n/a
AB189122 Indonesia 1998
AB189123 Indonesia 1998
AB189124 Indonesia 1998
EF051521 China 2001
D2MY97-10340 Malaysia 1997
D2MY04-33054 Malaysia 2004
EU482640 Vietnam 2006
EU482672 Vietnam 2006
EU179859 Brunei 2005
D2MY04-32618 Malaysia 2004
AY858035 Indonesia 2004
AF436115 1997
AY037116 Australia 1993
EU056810 Burkina Faso 1983
AY858036 Indonesia 2004
D2MY04-32883 Malaysia 2004
D2MY00-22563 Malaysia 2000
EU179858 Brunei 2005
EU359009 China n/a
EU179857 Brunei 2005
EU081177 Singapore 2005
EU081178 Singapore 2005
98
EU081179 Singapore 2005
EU081180 Singapore 2005
AF436114 1997
100
AF276619 China 1999
AF359579 China 1999

Curriculum vitae

Personal Particulars

Name	Swee Hoe
Family name	Ong
Year of birth	1974
Place of birth	Penang, Malaysia

Academic Qualifications

Ph.D. in Epidemiology
Faculty of Science, University of Basel, Switzerland
March 2007 - May 2009

M.Sc. in Bioinformatics
Faculty of Medicine, National University of Singapore
January 2003 - December 2004

B.Sc. in Cell and Molecular Biology - with first-class honours
Faculty of Life Sciences, National University of Malaysia
July 1994 - November 1997

Publications

1. Schreiber MJ, Holmes EC, **Ong SH**, Soh HSH, Liu W, Tanner L, Aw PPK, Tan HC, Ng LC, Leo YS, Low JGH, Ong A, Ooi EE, Vasudevan SG, Hibberd ML. Genomic Epidemiology of a Dengue Virus Epidemic in Urban Singapore. *J Virol.* 2009 May;83(9):4163-4173.

2. **Ong SH**, Yip JT, Chen YL, Liu W, Harun S, Lystiyaningsih E, Heriyanto B, Beckett CG, Mitchell WP, Hibberd ML, Suwandono A, Vasudevan SG, Schreiber MJ. Periodic re-emergence of endemic strains with strong epidemic potential-A proposed explanation for the 2004 Indonesian dengue epidemic. *Infect Genet Evol.* 2008 Mar; 8(2):191–204.

3. Yang IS, Ryu C, Cho KJ, Kim JK, **Ong SH**, Mitchell WP, Kim BS, Oh HB, Kim KH. IDBD: Infectious Disease Biomarker Database. *Nucleic Acids Res.* 2008 Jan; 36(Database issue):D455-60.

4. Schreiber MJ, **Ong SH**, Holland RC, Hibberd ML, Vasudevan SG, Mitchell WP, Holmes EC. DengueInfo: A web portal to dengue information resources. *Infect Genet Evol.* 2007 Jul; 7(4):540–541.

5. Hibberd ML, Ling L, Tolfvenstam T, Mitchell W, Wong C, Kuznetsov VA, George J, **Ong SH**, Ruan Y, Wei CL, Gu F, Fink J, Yip A, Liu W, Schreiber M, Vasudevan SG. A genomics approach to understanding host response during dengue infection. *Novartis Found Symp.* 2006; 277:206-214.