DISSERTATION

EVOLUTIONARY AND EPIDEMIC POTENTIAL OF DENGUE VIRUSES

Submitted by

Francisco J. Diaz-Castrillon

Department of Microbiology, Immunology and Pathology

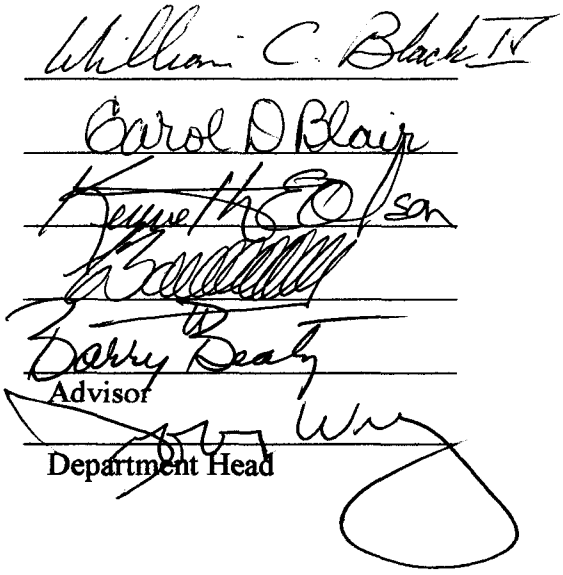
In partial fulfillment of the requirements
For the Degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
Summer 2004

COLORADO STATE UNIVERSITY

March 30, 2004

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY FRANCISCO JAVIER DIAZ CASTRILLON ENTITLED "EVOLUTIONARY AND EPIDEMIC POTENTIAL OF DENGUE VIRUS" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work



Advisor

Department Head

ABSTRACT OF DISSERTATION

EVOLUTIONARY AND EPIDEMIC POTENTIAL OF DENGUE VIRUSES

The epidemiology of infectious diseases caused by RNA virus is largely determined by the evolutionary potential of this kind of agents, characterized by high mutation rates and occasional exchange of genomic segments between strains. Dengue virus (DENV), one of such rapidly evolving viruses, has emerged as the most extended vector-borne viral pathogen in the world in the last decades. In this study, the forces shaping DENV evolution were investigated in the laboratory and in a natural population.

Phylogenetic analyses were performed in the prM and E genes of a collection of DENV isolates from Mexico and in selected sequences of viruses isolated around the world. The results showed a continuous appearance of substitutions mostly in synonymous sites, along with the frequent introduction, expansion and extinction of old and new viral strains coming from different continents. The introductions of some of these strains were temporally and geographically related to changes in the incidence and severity of DENV infection. DENV serotype-2 (DENV-2) exhibited a larger number of variants circulating in close temporal relationship and seems to have had a major role in the observed changes in the clinical outcome of the disease.

Recombination between different strains of the same serotype was investigated *in vitro* and in the same collection of viruses. While no clear evidence of this phenomenon was obtained in cell culture, two probable episodes of recombination

were detected in the natural population. All DENV-1 isolates from Mexico and the Americas seem to be derived from a strain originated in a remote recombination event between strains of the Eastern Hemisphere. The sequence of a single Mexican DENV-2 isolated in 1983 is better explained as a mosaic between the two major lineages of this serotype circulating in the Americas at that time.

Finally, the evolutionary consequences of alternation between arthropod and vertebrate hosts were explored *in vitro* by serial passages of four DENV-2 strains in a simplified cell culture model. It was found that different, mostly non-synonymous substitutions accumulated depending on the cell-type used to serially propagate the viruses. Some of these amino acid replacements were consistent across strains and were predicted to generate major changes in the structure of the envelope protein of the virus. They include substitutions that disrupt a glycosylation site or that may alter the conformational change occurring during the fusion between viral and cell membranes. Cell culture-adapted viruses reduced their infectivity for mosquitoes in a cell-type dependent manner but retained their infectiousness for the bypassed cell culture system.

Thus, different forces seem to govern DENV evolution *in vivo* and *in vitro*. While in cell culture host-dependent directional selection is evident, in natural populations purifying selection and gene flow are the dominant forces.

Francisco J. Diaz-Castrillon
Microbiology, Immunology and Pathology Department
Colorado State University
Fort Collins, CO 80523
Summer 2004

ACKNOWLEDGEMENTS

I would like to thank my graduate advisor, Dr. Barry Beaty for providing scientific guidance, support and mentorship. I also wish to thank the other members of my committee Drs. Carol Blair, William Black, Carolina Barillas and Ken Olson, for their enlighten ideas and effective advice during the last five years.

I am in debt with most of the staff, personnel and fellow students at the Arthropod-Borne and Infectious Diseases Laboratory (AIDL), but it will be unfair not to mention some of them that were especially helpful to my progress during these years. Drs. Alfredo Rayms-Keller, Kristine Bennett, Irma Sanchez, Dennis Pierro taught me many of the techniques I used in this work. I thank Dr. Joel Hutcheson for his opportune advices and for allowing me to use his fast computer, in which all the analyses and most of the writing of this dissertation were done. I am grateful with Dr. Charles Calisher who was a constant source of historical information.

Drs. Goro Kuno and Gary Clark, from the Dengue Branch from the Centers for Disease Control and Prevention gave me the idea and the support, respectively, to apply for graduate studies in Colorado State University. Dr. Jorge Osorio helped me to crystallize that decision and introduced me to the research team at AIDL. Thank you very much to them.

This work would have not been possible without the collaboration of our colleagues Dr. Jose Farfan and Maria Alba Lorono at the Centro de Estudios Regionales "Dr. Hideyo Noguchi", Merida, Mexico, who provided most of isolates used in this research. Drs. Robert Tesh from the University of Texas, Galveston and Duane Gubler, from CDC, Fort Collins, Colorado, kindly provided additional contributions to this virus collection.

I am also grateful with the following authorities of the Universidad de Antioquia, Medellin, Colombia, for their unconditional and generous support to my study project: Dr. Jaime Restrepo, president; Dr. Luis Fernando Mejia, secretary general; Dr. Alberto Uribe, dean, Faculty of Medicine; Dr. Fabio Nelson Zuluaga, Head, Department of Microbiology, Ana Eugenia Arango MS, coordinator, Virology Section and Dr. Jorge Ossa, head of Biogenesis research group.

Funding for the research presented here was provided by several grants from the National Institute of Health. Economic support for my life expenses and studies during these years was provided by Universidad de Antioquia, Fundacion para el Futuro de Colombia (Colfuturo) and by several assistantships and fellowships awarded through Colorado State University by Drs. Barry Beaty and Carol Blair.

FJD, June, 2004

TABLE OF CONTENTS

Abstract of Dissertation..... iii
Acknowledgements.....v
List of Tables..... ix
List of Figures x

Chapter 1. A Literature Review on Dengue..... 1
The problem..... 1
The disease..... 1
The agent..... 2
Pathogenesis of DENV infections..... 5
Pathophysiology of DHF and DSS..... 8
Ecology and vectors.....12
Epidemiology.....14
Prevention and control.....17

Chapter 2. Genetic variation within the prM coding region of DENV from the
Yucatan peninsula of Mexico..... 20
Materials and Methods.....23
Results..... 27
Discussion.....30

Chapter 3. The molecular epidemiology of DENV in Mexico.....37
Materials and Methods.....44
Results.....50
Discussion..... 59

Chapter 4. A survey for recombinants among DENV isolates from Mexico..... 66
Materials and Methods..... 72
Results..... 77
Discussion..... 89

Chapter 5. Development of an *in vitro* model for DENV recombination..... 95
Materials and Methods..... 103
Results..... 112
Discussion..... 127

Chapter 6. Evolutionary forces shaping the adaptation of an arbovirus to vertebrate and/or mosquito cells..... 132
Materials and Methods.....135
Results..... 140
Phenotypic changes.....140
Genotypic changes..... 146
Discussion.....150

Chapter 7. Summary and conclusions.....160

References.....164

Appendix: Alignments of sequences.....189
DENV-1.....190
DENV-2..... 195
DENV-3.....202
DENV-4..... 206

LIST OF TABLES

Table #		Page
2-1	Mexican isolates used in the phylogenetic analysis of the prM gene	23
2-2	Primers used to amplify and sequence the prM gene	25
2-3	Reference strains used in the phylogenetic analysis of the prM gene	26
3-1	Mexican isolates analyzed in the E gene	45
3-2	Primers used for E gene amplification and sequencing	47
3-3	Diversity indexes and test for selection	50
4-1	Cases of recombination in natural populations of RNA viruses	71
4-2	Primers used in the full genome sequencing of DENV-2	73
5-1	Primers used in recombinant-specific PCR	109
6-1	Plaque assay and mosquito infectivity results at the beginning and the end of the adaptive process	143
6-2	Nucleotide and amino acid substitutions observed in cell culture-adapted viruses	147
6-3	Results of the tests for selection in the adapted strains	149

LIST OF FIGURES

Figure #		Page
1-1	Histogram of dengue virus genome	2
2-1	Maximum likelihood trees of the prM gene	29
3-1	Incidence of DF and DHF in Mexico 1978-2002	39
3-2	Patterns of transmission and sampling sites	40
3-3	Phylogenetic tree for DENV-1	52
3-4	Phylogenetic tree for DENV-2	53
3-5	Phylogenetic tree for DENV-3	55
3-6	Phylogenetic tree for DENV-4	56
3-7	Phylogenetic tree for all DENV serotypes	58
4-1	Strategy for full genome sequencing	74
4-2	Schematic representation of the bootscanning test	75
4-3	Schematic representation of the Chimaera test	76
4-4	Demonstration of methods for detecting recombination in WEEV	78
4-5	Similarity plots and bootscans in intraserotypic analyses	79
4-6	Similarity plots and bootscans for DENV-3 and DENV-4	81
4-7	Similarity plots and bootscans for DENV-1	83
4-8	Bootscans of DENV-1 isolates reported as recombinants	84
4-9	Additional tests for recombination in DENV-1	85
4-10	Similarity plots and bootscans for DENV-2	87
4-11	Additional tests for recombination in DENV-2	88
4-12	Bootscans of the complete genome of Mexican DENV-2	90
5-1	Strategy and procedures used to search for recombination	104
5-2	Amplification of recombinant molecules in RS-PCR	105
5-3	Testing of the specificity of strain-specific primers	112
5-4	Results of the first round RS-PCR	114
5-5	Similarity plots of the products SLxJAM	116
5-6	Similarity plots of the products PRxJAM	117
5-7	Breakpoint locations and parental origin of RS-PCR products	119

LIST OF FIGURES (continued)

Figure #		Page
5-8	Results of RS-PCRs with different polymerases.....	122
5-9	Results of RS-PCRs with different temperatures	124
5-10	Results of RS-PCRs using cDNAs synthesized with Thermoscript....	126
6-1	Schematic view of the adaptive process.....	137
6-2	Effect of serial passage in plaque size and morphology.....142	
6-3	Infectivity in mosquitoes of the adapted viruses.....	144
6-4	Fitness assessment of the adapted viruses.....	145
6-5	Mapping of the substitutions in the 3D-structure of E protein.....	148

CHAPTER 1.

A LITERATURE REVIEW ON DENGUE

The Problem

Dengue (DEN) is the most important arthropod-borne viral infection of humans with about 100 million cases and 25,000 DEN-related deaths reported annually. Mosquitoes of the genus *Aedes*, most importantly *Ae aegypti,* transmit DEN. About one half of humanity lives in areas infested with this vector and therefore are at risk of acquiring DEN (Monath, 1994, WHO, 1999). Neither anti-DEN vaccines nor effective antiviral drugs are available to prevent or treat the disease.

The Disease

DEN virus infection can result in clinical outcomes ranging from asymptomatic to fatal disease. At least half of the infections are inapparent or are not symptomatic enough to require medical attention (Burke et al., 1988, Kliks et al., 1989). Among symptomatic cases most are classified as dengue fever (DF), an acute and self-limited condition characterized by fever, rash, lymphadenopathy, minor hemorrhages and pains localized in the head, eyes, muscles and joints. These pains can be very incapacitating and have earned the disease its pseudonym "break-bone fever". In some settings a portion of cases have a more serious outcome designated dengue hemorrhagic fever (DHF). DHF characterized by hemostatic disorders, hepatic

involvement and, more importantly, plasma leakage resulting from increased vascular permeability. Dengue Shock Syndrome (DSS), the extreme form of DHF, happens when the plasma loss is severe enough to lead to potentially fatal circulatory failure (Gubler, 1998, Guzman & Kouri, 2002, WHO, 1999).

The Agent

Dengue virus (DENV) is a member of the genus *Flavivirus*, family *Flaviviridae*. There are four antigenically related DENV serotypes (DENV-1 to DENV-4), which have the same transmission cycles and cause similar clinical manifestations. The virion is spherical, about 50 nm in diameter, has a lipid envelope, a protein capsid and a RNA genome (Kuhn et al., 2002). The genome (Figure 1-1) consists of a positive-sense, single-stranded RNA approximately 10,700 nucleotides in length. It has 5' and 3' untranslated regions (UTR) and a single open reading frame (ORF). The 5' UTR is capped; the 3' UTR is not polyadenylated. The ORF encodes a polyprotein about 3,400 amino acids long. The polyprotein is cleaved to produce three structural proteins, capsid (C), membrane precursor (prM), envelope (E), and seven nonstructural (NS) proteins: NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Chambers et al., 1990, Chang, 1997).

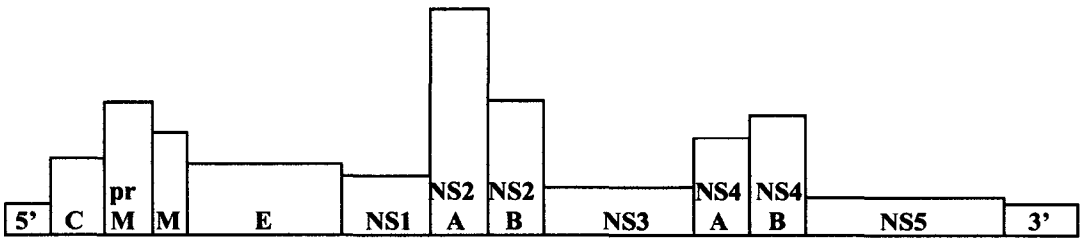


Figure 1-1. Histogram showing the order, relative length (horizontal axis), and relative variation (vertical axis) of genes in DENV genome.

The DENV life cycle begins with the adsorption of the virion to the cell membrane. Several cellular molecules have been reported to bind the viral envelope (E) glycoprotein and serve as putative receptors for DENV. Heparan sulfate could function as a primary adhesion molecule (Chen et al., 1997), but several different membrane proteins have been reported to bind to the viral envelope (E) glycoprotein (Martinez-Barragan & del Angel, 2001, Moreno-Altamirano et al., 2002, Tassaneetrithep et al., 2003, Wei et al., 2003, Yazi Mendoza et al., 2002). Interestingly, these putative receptors are different depending on the cell type and the viral strain used in the study, suggesting that viral attachment is a complex phenomenon that involves several receptors and/or that receptors vary depending on the tissue being infected (Bielefeldt-Ohmann et al., 2001). DENV enters the cell mainly by endocytosis and a low pH-dependent conformational change of E glycoprotein enables the fusion of the viral envelope with the membrane of the endosome (Modis et al., 2003). Mosquito cell generated DENV particles may enter cells by direct fusion with the plasma membrane (Chang, 1997).

In order to proceed with its life cycle, the DENV genome (or vRNA) must be translated by the protein synthesis machinery of the cell. The single ORF is translated into a polyprotein that is co- and post-translationally cleaved to produce the ten viral proteins. The cleavage is partially carried out by a viral protease encoded in the NS3 gene.

The DENV genome is replicated by a putative replication complex composed of viral NS1, NS3 and NS5 proteins and hypothetical host molecules. This process begins with the synthesis of a negative strand RNA complementary to the vRNA resulting in a double stranded replicative form (RF), which functions as a recycling template. This is

followed by an asymmetric and semi-conservative synthesis of more positive strand RNAs that serve as new messengers or are packaged as the vRNA in progeny virions. The NS5 protein functions as the viral RNA-dependent RNA polymerase (RdRp) in the synthesis of both negative and positive sense strands (Chambers et al., 1990, Chang, 1997). This replicative process takes place in membranous compartments called vesicle packets that supposedly protect the RF from dsRNA-induced host defenses such as protein kinase R, RNase L and RNA interference (Uchil & Satchidanandam, 2003). Like in other RNA viruses, DENV replication is an error-prone process that generates diversity of the viral populations and favors rapid evolution. These two phenomena will be discussed in chapters 3 and 6.

DENV assembly and maturation have not been extensively studied. It is known that prM, E and NS1 proteins are translocated into the lumen of the rough endoplasmic reticulum and remain anchored to its membranes. They are glycosylated and carbohydrates trimmed as they advance in the secretory pathway (Chang, 1997). This step results in a major difference in the virions depending on the cell in which they replicate. In insect cells, the oligosaccharide assembled is of the high-mannose type, but in vertebrate cells it is a more complex glycan that also contains galactose and sialic acid (Marz L, 1995). Assembly of the viral capsids and packaging of the vRNA could take place close to the membranes of this system and virions presumably acquire their envelopes while budding through the membranes of secretory vesicles. Before viral release, the prM protein is cleaved to produce the membrane (M) protein observed in mature viral particles. This cleavage is thought to prime the conformational change of E glycoprotein that enables viral envelope fusion with the endosome membrane (Modis et al., 2003). The virus is released from the cell when the secretory vesicles empty their contents into the extracelular space.

Pathogenesis of DENV Infections.

The study of the pathogenesis of DENV infections has been impaired by the lack of animal models that reproduce the clinical and pathologic findings in humans. Mice can be infected intravenously but neither viremia nor disease develops. Intracerebral injection of mouse adapted strains may produce an encephalitis, which is not a component of disease in humans (Rothman, 1997). Monkeys develop an infection with a time course similar to that observed in humans but viremias are low and no symptoms or pathological findings are observed (Halstead et al., 1973, Marchette et al., 1973). The use of human monocytes and endothelial cells, either in primary culture or as cell lines, has allowed the *in vitro* study of specific phenomena like antibody-dependent enhancement (Brandt et al., 1982, Halstead et al., 1983, Halstead et al., 1984, Morens et al., 1987a) or the changes in the permeability of human endothelium (Bonner & O'Sullivan, 1998, Jacobs & Levin, 2002). However these models cannot substitute for *in vivo* studies.

Thus, most information on DEN pathogenesis has been obtained from studies in humans. Early experiments with human volunteers were instrumental in elucidation of the viral agent, arthropod vectors, extrinsic and intrinsic incubation periods, multiplicity of serotypes, development of immunity and lack of long-term cross protection in DENV infections (Sabin, 1952). However, no DHF/DSS cases were observed and further experimentation with humans was halted after the recognition of the severe forms of DEN. Because of these limitations the study of DEN pathogenesis has relied to a large extent on epidemiological, clinical and pathological studies in naturally infected persons.

Human infection begins with the intradermal or subcutaneous inoculation of DENV during the bite of an infected mosquito. DENV replicates at the site of inoculation, probably in dendritic cells or tissue macrophages (Marovich et al., 2001, Wu et al., 2000). A few days after the inoculation, the virus is detectable in the regional lymph nodes and from there the virus spreads to other organs through the blood stream (Rothman, 1997). Detectable viremia coincides with the febril period. In the blood, virus is detectable in the plasma and in peripheral blood mononuclear cells (PBMC). Both monocytes and B-lymphocytes have been reported as the main cells supporting viral replication in blood (Halstead et al., 1977, King et al., 1999). In fatal cases the virus has been isolated more consistently from liver but occasionally from other organs. Macrophages and Kupffer cells have been identified as targets in solid tissues (Rothman, 1997). Infection of vascular endothelial cells *in vivo* has been controversial although DENV readily infects such cells *in vitro* (Warke et al., 2003).

Although *in vitro* DENV induces cytopathic effects (CPE) in many cell lines, most studies have failed to find DENV induced CPE *in vivo*. Physiologic changes have been detected instead. For example, human monocytes infected with DENV release interleukin (IL)-1α, tumor necrosis factor (TNF)α and plasminogen-activator inhibitor (Chang & Shaio, 1994, Hober et al., 1998). Infected human endothelial cells release IL-6 and IL-8 and tissue plaminogen activator (Huang et al., 2003, Huang et al., 2000). In the latter cells DENV infection induces the expression of multiple genes related to stress, defense, cell adhesion, wounding, inflammatory, and antiviral pathways (Warke et al., 2003). DENV also can induce apoptosis of hepatic cells and $CD8^+$ lymphocytes, and this mechanism could be playing a role in the pathogenesis of DHF (Huerre et al., 2001, Marianneau et al., 1998, Mongkolsapaya et al., 2003).

Antibody response to DENV infection is notoriously different in primary and subsequent infections. In the former, a strong IgM response appears by the end of the 3 to 6 day-long febrile period, peaks within two weeks and lasts for two to three months. IgG appears shortly after IgM, elevates at medium titer and lasts for years. In secondary and tertiary infections an IgG increase usually precedes IgM and it is much stronger and more durable than in primary infections. IgM response is weaker and declines quickly (Innis et al., 1989). These differences permit differentiation between primary and subsequent (heterotypic) infections. Antibodies are protective against the homologous serotype as demonstrated by passive protection to DENV challenge in animals (Henchal et al., 1988) and in infants born to DENV immune mothers (Kliks et al., 1988). Antibodies can also negatively affect the clinical outcome of DEN as explained below. Antibodies can be serotype specific, or DENV complex or flavivirus group-reactive and this specificity correlates roughly with the neutralizing ability of the antibody. Antibodies can recognize both structural and non-structural DENV proteins, but only those reactive with E, M and NS1 proteins have a protective function (Roehrig, 2003).

DENV also elicits specific $CD4^+$ and $CD8^+$ T cellular immune responses. After primary infection, most circulating T lymphocytes are serotype specific, but following secondary infection serotype cross-reactive memory T cells predominate and more frequently recognize the highly conserved non-structural proteins, especially NS3. Both $CD4^+$ and $CD8^+$ T lymphocytes are able to lyse DENV-infected cells by perforin-dependent and Fas-dependent mechanisms. DENV-specific $CD4^+$ cells also release many cytokines, which could be important factors in viral clearance as well as in the pathogenesis of DHF (Rothman & Ennis, 1999).

The duration of protection against reinfection with DENV has been the object of few studies. Human volunteers challenged with the homologous strain of virus were completely immune to reinfection for as long as 18 months. Partial cross-protection to the challenge with a heterologous serotype was also demonstrable two to nine months after the primary infection (Sabin, 1952). Neutralizing antibodies have been observed more than forty years after infection in some individuals (Papaevangelou & Halstead, 1977), but the protective capacity of such immunity has not been proven. No experimental reinoculation with homologous virus has been attempted after intervals longer than 18 months, and no challenge study using divergent strains of the same serotype has ever been reported.

Pathophysiology of DHF and DSS.

The main pathologic features of DHF/DSS are plasma leakage, bleeding tendency and hepatic involvement (WHO, 1999). Factors predisposing to the severe forms of dengue and mechanisms involved in their production have been the objects of many studies and considerable debate. The first hint came from the observation that most DHF/DSS cases were associated with secondary serologic responses (Halstead & Yamarat, 1965). Several prospective studies conducted in Southeast Asia confirmed that finding and estimated that the relative risk for DHF was at least 15-fold and for DSS between 50 and 100-fold in patients undergoing a secondary infection (Burke et al., 1988, Sangkawibha et al., 1984, Thein et al., 1997). The same association between secondary infection and DHF has been observed in the Americas (Bravo et al., 1987, Guzman et al., 1990, LaFleur et al., 2002).

These observations were linked to an *in vitro* phenomenon described years before and referred as antibody-dependent enhancement (ADE), which refers to an increase in replication of some flaviviruses in cell culture by the addition of immune sera at dilutions beyond the neutralization endpoint (Hawkes, 1964). *In vitro* models were then developed to test if ADE could also occur with DENV in monocytes and macrophages, cells that were thought to be the main targets for DENV replication. These models demonstrated that increased virus yield can be obtained by treatment with either immune sera at subneutralizing dilutions or monoclonal antibodies directed against flavivirus group or serotype specific determinants (Brandt et al., 1982, Halstead et al., 1983, Halstead et al., 1984, Littaua et al., 1990, Morens et al., 1987a). Perhaps the most convincing evidence of the importance of ADE in DHF/DSS was the demonstration of higher ADE activity in sera of children that later developed DHF in comparison to those of children than later developed DF (Kliks et al., 1989). Furthermore sera from mothers of infants who developed DHF in their first year exhibited ADE in vitro (Kliks et al., 1988).

The joint evidence of epidemiologic studies and *in vitro* findings led to the "immune enhancement hypothesis", which postulates that antibodies remaining after the first dengue episode bind to shared epitopes on a different infecting serotype but do not neutralize it. This would facilitate DENV contact with and enhanced penetration into monocytes and macrophages through the interaction of the Fc portion of IgG and its receptor (FcγR), which is abundant in these types of cells. Thus, the FcγR would act as an alternative receptor for the virus. This in turn would lead to greater viral replication and release of hypothetical soluble factors that somehow would mediate the increased vascular permeability and the hemostatic disorders (Halstead et al., 1973).

Indeed DENV infection has been shown to induce human monocytes in culture to produce TNFα and IL-1β (Chang & Shaio, 1994, Hober et al., 1993) and supernatants of these tissue cultures can activate endothelial cells (Anderson et al., 1997).

Significant insight into the mechanisms of immune enhancement has been provided by studies on cellular immunity in dengue. Cytokines play a role in this phenomenon and they could be the soluble factors mediating plasma leakage and bleeding (Rothman & Ennis, 1999). Dengue specific $CD4^+$ cells release cytokines like interferon (IFN)-γ, IL-2, TNFα and TNFβ (Gagnon et al., 2002). TNFα and IL-2 can induce plasma leakage and IFN-γ enhances TNFα production by monocytes. IFN-γ also enhances the expression of class II HLA antigens and FcγR in monocytes, which could augment DENV antigen presentation and antibody-mediated uptake of virus respectively. Higher levels of TNFα and IL-6 have been found in DHF than in DF patients (Green et al., 1999, Hober et al., 1993, Kuno & Bailey, 1994).

However, the immune enhancement hypothesis fails to explain the sporadic occurrence of severe forms of the disease in primary infections (Barnes & Rosen, 1974, Morens et al., 1987b), and the absence of DHF/DSS in some epidemic settings in spite of frequent occurrence of sequential infections by different serotypes. The latter was the case in the Americas, where DHF/DSS outbreaks were not observed before 1981 although several serotypes had been circulating (Gubler & Clark, 1995, Watts et al., 1999). This moved other researchers to look for risk factors in addition to the immune memory in the pathogenesis of DHF.

Viral risk factors were then explored. It was well known that DHF/DSS was more frequently caused by DENV-2 than by other serotypes (Sangkawibha et al., 1984). Indeed the emergence and rapid spread of epidemic DHF/DSS in the Americas in the eighties and nineties coincided with the replacement of the American DENV-2 strain by another DENV-2 strain of Asian origin (Guzman et al., 1995, Rico-Hesse, 1990, Rico-Hesse et al., 1997). Several molecular differences were detected between the strains associated and non-associated with DHF/DSS, and they were postulated to account for the different clinical outcomes (Cologna & Rico-Hesse, 2003, Leitmeyer et al., 1999, Pandey & Igarashi, 2000). The introduction and dispersion of DENV strains in the Americas and their influence in the emergence of DHF in this hemisphere will be described and discussed in detail in chapter 3.

A third, much less studied factor that influences the frequency of severe forms of the disease is the genetic background of the infected individual. A puzzling fact is the virtual absence of DHF/DSS in Africa in spite of circulation of multiple serotypes and strains (Gubler & Clark, 1995, Halstead, 2002). A similar situation is observed in Haiti where most of the population is black (Halstead et al., 2001). A study on risk factors in Cuba showed that patients of African descent developed DHF with much lower frequency than whites (Bravo et al., 1987). These observations suggest that host genetic factors can be important determinants of dengue severity.

Most studies on genetic factors in DHF have focused on the major histocompatibility complex (MHC). Several class I and class II MHC antigens have been found associated with either susceptibility or protection to DHF but those associations have not been consistent among studies (LaFleur et al., 2002, Loke et al., 2001, Stephens et al., 2002). A recent study on non-HLA genetic factors found

significant protection against DHF in carriers of a rare allele of the vitamin D receptor gene and in individuals homozygous for the 131R variant of the FcγR-IIA gene (Loke et al., 2002). Interestingly FcγR-II mediates ADE *in vitro* (Littaua et al., 1990) and the 131R variant has lower capacity to bind IgG-2 antibodies, which theoretically could decrease the enhancing capacity of heterotypic antibodies in monocytes and macrophages. This latter study (Loke et al., 2002) is the first to link polymorphisms in human genes and the ADE phenomenon.

Ecology and Vectors

DENV is a mosquito-borne pathogen. Several species of the genus *Aedes* are the vectors and different primates are the vertebrate hosts. There are at least three cycles of transmission of DENV in nature: the forest or enzootic cycle takes place in the forests of Asia and Africa and involves primates and canopy-dwelling *Aedes* of the subgenus *Diceromyia*, *Stegomyia* and *Finlaya*. Vertical transmission could play a role in the maintenance of this cycle. In rural areas or small towns close to these forests there is a rural epidemic cycle in which humans and *Aedes* (*Stegomyia*) mosquitoes are the hosts; *Ae. albopictus*, *Ae. polynesiensis*, *Ae aegypti* and perhaps *Ae. (Gymnometopa) mediovittatus* may act as vectors in these settings (Gubler, 1998). This cycle is limited by the rapid exhaustion of susceptible humans. Finally, and more important for human health, an urban, endemic or epidemic cycle take place in cities around the tropical and subtropical areas of the entire world; *Ae. aegypti* is the vector in this cycle. The endemic or epidemic character of the cycle depends mainly on the size of the population of the city. The critical size for sustained endemic transmission is estimated to be between 150,000 and one million inhabitants (Kuno, 1995), but it also likely depends on the mobility of the population and the vectorial capacity of the vector.

DENV infection of the mosquito begins with a blood meal from a viremic human or monkey. Initial viral replication occurs in the mosquito midgut and it takes several days to be disseminated to other organs. Then, virus spreads to hemocytes, proventriculus, fat body, brain, and nervous ganglia. Later in the infection, the virus is found in the Malpighian tubules, eye, legs, antennae and, more important, in the salivary glands. DENV also invades the oviduct epithelium but not the ovarioles (Rodhain, 1997). Virus persists in all these organs and remains infectious throughout the life of the mosquito. The extrinsic incubation period varies between one to three weeks and depends on the amount of virus ingested, environmental conditions (e. g. the temperature) and very likely on the infecting viral strain and mosquito species and strain (Rodhain, 1997). Vertical transmission of the virus varies among mosquito species being maximal in *Ae. mediovittatus*, medial in *Ae. Albopictus,* and very rare in *Ae. aegypti*. Sexual transmission from males to females has also been demonstrated. The importance of vertical and sexual transmission in the maintenance of DENV in nature is still uncertain.

Little is known about the level of viremia necessary to initiate an infection in the mosquito in nature but it is believed to be several orders of magnitude lower than the viral titers routinely used to infect mosquitoes with artificial feeding mixtures, the method routinely used to estimate vector susceptibility and competence (Rodhain, 1997). It is assumed that the susceptibility measured in the laboratory parallels that found in nature. Using these methods it has been shown that vector competence widely vary among mosquito species and among mosquito strains within species (Bennett et al., 2002, Rosen et al., 1985). It was previously believed than *Ae. aegypti* was less susceptible to infection than *Ae. albopictus* and that this relative resistance in

the former species could select for viral strains that produce high viremias and therefore are more virulent (Rosen et al., 1985). However, experiments using recently colonized mosquito strains demonstrated that *Ae. albopictus* is less competent, and suggested that the higher susceptibility found before was an artifact of using of strains that had been colonized for many years (Vazeille et al., 2003).

Except in Africa, where it breeds as a feral species independent of humans, *Ae. aegypti* is a highly domesticated mosquito that usually oviposits in artificial containers like cisterns, tanks, plastic recipients, automobile tires, flower pots and any other objects that collect rainwater (Gubler, 1998, Rodhain, 1997). Females prefer to live inside human dwellings where they can feed all day long. They are very nervous feeders that can stop and resume biting several times on the same or different persons before completing a blood meal (Putnam & Scott, 1995, Scott et al., 1993). This makes them very efficient vectors that can transmit the virus to several people in a single gonotrophic cycle. *Ae aegypti* has undergone dramatic expansions since the middle of the 20th century in most of the tropical and subtropical regions of the world (Gubler & Clark, 1995, Halstead, 1997).

Epidemiology

Until recently, epidemiological data on DEN were limited to clinical records of epidemic activity of DEN-like illnesses. However, the clinical manifestations of DF and DHF are not specific. Specialized laboratory tests were developed only by the mid 20th century. Such tests are still not widely available and efficient diagnostic tests are still lacking.

The earliest clinical description of a potential dengue outbreak discovered to date was found in a Chinese encyclopedia most likely published in the 4th century AD (Nobuchi, 1979). The first pandemic on record ocurred in 1779-1780 in countries as distant as Indonesia, Egypt and the United States (Gubler, 1998), and this was when Benjamin Rush made the classical clinical description of DF (Rush, 1789). Before 1940 the presence of DEN was recorded many times involving rather infrequent although often large outbreaks throughout the world.

The current pandemic of DEN began in Southeast Asia during World War II. In the following decades the illness exhibited a steady increase in that part of the world and the problem extended to the Pacific Islands in the 1970's, to China, India, Sri Lanka and Africa in the 1980's, and more recently to the Americas (Gubler & Clark, 1995). DHF was first recognized in the 1950's in the Philippines and soon after in all countries in Southeast Asia; however, retrospective studies suggest that it could have existed as early as 1897 in northern Australia (Halstead, 1997). DHF has become a leading cause of hospitalization and death in children of Southeast Asia and with increasing frequency in other parts of the world (Gubler & Clark, 1995).

In the Americas DEN was probably present as early as the 17th century since outbreaks of a disease similar to DF were described in French West Indies in 1635 and Panama in 1699 (Gubler, 1998). Beginning in the 1950s a huge campaign to combat urban yellow fever eradicated *Ae. aegypti* from most Central and South American countries until the 1970s. However the program failed to eradicate the mosquito in some countries and DEN occurred sporadically in the Caribbean Islands during that time with DENV-2 and DENV-3 being the etiologic agents (Gubler & Clark, 1995). During the 1970s the program was discontinued or deteriorated in most countries and

DEN outbreaks reappeared. In 1977 and 1981 DENV-1 and DENV-4 were introduced respectively, and new strains of DENV-2 were also imported (Gubler, 1998, Guzman & Kouri, 2003). DENV-3 was not detected since 1977, but a different strain of that serotype was introduced in Central America in 1994 (Guzman et al., 1996b). The introduction and dispersal of these DENV strains and their influence in the epidemiology and severity of DEN in the Americas will be discussed in detail in another chapter of this dissertation. In the last decades DF incidence has increased and its distribution has expanded. By 2002 more than 30 Latin American countries reported over 1,000,000 DEN cases. (Guzman & Kouri, 2003).

The first DHF outbreak recorded in the Americas happened in 1981 in Cuba; more than 10,000 cases were classified as DHF and more than 150 persons died (Guzman et al., 1984). Beginning in 1989 DHF appeared and rapidly dispersed in the countries of the Atlantic coast of South America. In Central America and Mexico DHF was not present until 1994. In 2002 DHF occurred in 20 countries with more than 17,000 DHF reported cases, including 225 fatalities (Guzman & Kouri, 2003).

Thus, epidemic DF and DHF have emerged as major public health problems in tropical and subtropical areas throughout the world. Distribution, seasonality and clinical outcomes vary by region but a general tendency to increasing of frequency and severity of dengue outbreaks is pervasive. Reasons invoked for this emergence are demographic, cultural, environmental and political (Monath, 1994). The unprecedented global population growth with the crowding and suboptimal water and sewer management resulting from unplanned urbanization is likely the main factor. The introduction of many disposable, mainly plastic, products has multiplied the availability of mosquito breeding sites in areas with inadequate waste management systems

(Gubler, 1998). The progressive decay in public health infrastructure resulting from privatization, deregulation and globalization policies imposed on developing countries in the last decades has eroded mosquito control programs. Increased air travel has become an important method of trafficking virus strains among continents and has resulted in the simultaneous or closely sequential circulation of multiple serotypes, a condition referred to as hyperendemicity, in regions where these serotypes were widely separated before. The controversial influence of global warming could be an additional conditioning factor in DEN expansion (Hales, 2002).

Prevention and Control.

This far, most campaigns or programs aimed to control or eradicate DENV infection have relied on vector control. One strategy has been to spray ultra-low volume insecticides, a measure of relatively low efficacy in the Americas. Larvicides are also commonly used with similar results (Gubler, 1998, Gubler & Clark, 1995). More recently, education and community participation have been stressed in order to control the vector by eliminating potential mosquito breeding sites. The efficacy of such strategies is still unknown (Gubler & Clark, 1995).

The successes obtained with the use of vaccines in the control of other flavivirus infections, e.g. yellow fever and Japanese encephalitis, and the recent advances in bio-engineering techniques, including the development of reverse genetic approaches for RNA viruses, offer promise for vaccination as a viable alternative to control DEN (Halstead & Deen, 2002). In spite of the appeal of these strategies, many obstacles lie in the way. The lack of appropriate animal models for DEN is a major practical limitation (Rothman, 2003). The multiplicity of serotypes entail the potential of immune enhancement of the infection if only partial protection is achieved. This has always

been a major concern that mandates development tetravalent vaccines (Halstead & Deen, 2002). Nevertheless, many researchers have moved into the field of DENV vaccine development. Most strategies to produce vaccines for other viruses have been tried for DENV. These include candidates vaccines attenuated by cell culture passage, attenuated by genetic manipulation including chimeric vaccines, chemical inactivation as well as recombinant subunit and DNA vaccines (Chang et al., 2001, Chang et al., 2004, Eckels et al., 2003, Kuno et al., 1981, Lai & Monath, 2003, Pugachev et al., 2003, Putnak et al., 2003b).

Most non-replicating DENV candidate vaccines have proved to be immunogenic in animal models and in some cases this correlated with protection to challenge (Chang et al., 2001, Eckels & Putnak, 2003, Mune et al., 2003, Wu et al., 2003). However several doses are required and protection is not long lasting, which suggests that frequent revaccination would be necessary (Eckels & Putnak, 2003, Putnak et al., 2003a, Wu et al., 2003). Human clinical trials with such vaccines have not been reported.

Attenuated DENV vaccine candidates are in a more advanced stage. Two tetravalent cell culture-attenuated vaccines had been evaluated in phase I clinical trials with moderate success (Innis & Eckels, 2003, Sabchareon et al., 2002). The vaccines are still considerably reactogenic and multiple doses are necessary to achieve the desired response (Sabchareon et al., 2004, Sun et al., 2003). They have been evaluated for safety and immunogenicity but not yet for efficacy in the field.

Other approaches to produce live attenuated vaccines include artificial mutation of the virus. This far, deletions in the C gene (Kofler et al., 2004) or in the 3' UTR

(Troyer et al., 2001) have produced adequate attenuation while retaining immunogenicity (Kofler et al., 2004, Whitehead et al., 2003). Some successfully attenuated flavivirus strains like DENV-4 strain 814669 (Bray et al., 1996), DENV-2 PDK-53 candidate vaccine (Butrapet et al., 2002) and YF 17D vaccine (Guirakhoo et al., 2001) have been used as genetic backbones to construct chimeric flaviviruses that express the target antigens and have the required attenuation. In all these cases, the homologous genes of the virus of interest replace structural genes of the attenuated strains. Preclinical and clinical studies have shown that several chimeric flavivirus vaccines have the safety and immunogenic profile to warrant further evaluation in humans (Lai & Monath, 2003).

CHAPTER 2

GENETIC VARIATION WITHIN THE PREMEMBRANE CODING REGION OF DENGUE VIRUSES FROM THE YUCATAN PENINSULA OF MEXICO.

Epidemic dengue has emerged as a major public health problem in tropical and subtropical areas throughout the world (Gubler & Clark, 1995, Monath, 1994). Since the middle of the 20th century, the disease, its agents and its vectors have undergone dramatic expansions. World Health Organization (WHO) receives reports of about 500,000 DF cases each year, but estimates that as many as 50 million people are infected annually (Clark, 2002). 2,5 billion people live in areas where DEN is endemic (WHO, 1999). DHF and DSS, the severe forms of the disease, are also increasing in frequency and expanding their territory; in 1998, 3442 DEN related deaths were reported worldwide (Guzman & Kouri, 2002).

In the Americas, epidemic DHF/DSS appeared for the first time in Cuba during 1981 in an explosive outbreak that lasted for four months (Kouri et al., 1986). Then it occurred sporadically in several countries until 1989, when epidemic DHF emerged again in Venezuela. Other countries of South America, Central America and the

Caribbean islands were soon affected, and DHF became a permanent threat to this region (Gubler, 1998, Gubler & Clark, 1995).

The changes in the epidemiology of DENV in the Western Hemisphere are illustrated by events in the Yucatan peninsula of Mexico. During the 1950's and 1960's, intensive campaigns to eradicate *Ae. aegypti* led to the disappearance of DF. However, *Ae. aegypti* recolonized most of the Mexican Atlantic coast during the 1970's and DF was reported again in 1978 (Gomez-Dantes, 1992). In Yucatan, from 1979 to 1981, thousands of cases were reported each year and DENV-1 virus was identified as the etiologic agent. DENV-4 virus caused a major epidemic in 1984 with more than 5,000 reported cases but only one case met all WHO criteria for DHF (Lorono-Pino et al., 1993). In 1991 DENV-2 was detected in Yucatan and 364 cases were confirmed. DENV-3 was detected for the first time in 1995. Epidemic DHF was detected beginning in 1996 and 163 DHF cases were reported in 1997 (Reported by Sistema Nacional de Vigilancia Epidemiológica, México, 1998). DENV-3 virus was the predominant serotype in the latter year, representing 92% of the DENV virus isolates. Thus, DF has been diagnosed in the Yucatan peninsula of México every year since 1979, but epidemic DHF/DSS occurred only after 1995, when the incidence of DENV infection increased abruptly (Briseno-Garcia et al., 1996).

Molecular epidemiological approaches have revealed extensive genetic variability within the DENV serotypes and have led to the recognition of two to five mayor clades or "genotypes" within the respective serotypes (Chungue et al., 1995, Chungue et al., 1993, Farfan et al., 1997, Lanciotti et al., 1997, Lanciotti et al., 1994, Lewis et al., 1993, Rico-Hesse, 1990, Vorndam et al., 1994). Genotypes differ in their ability to cause severe DENV disease. For example in the Americas, infection with the

"native" DENV-2 American genotype typically results in DF cases, whereas infection with recently introduced, Southeast Asian genotype more frequently results in DHF/DSS (Rico-Hesse et al., 1997, Watts et al., 1999).

By the time this study was done (year 2000), phylogenetic studies had suggested that at least two different genotypes of both, DENV-1 and -2, had circulated in México (Rico-Hesse, 1990, Rico-Hesse et al., 1997), two DENV-4 isolates of 1984 had been classified in a single genotype (Chungue et al., 1995, Lanciotti et al., 1997) and sequences of DENV-3 viruses from Mexico had not been studied. The city or state from which most of these isolates came was not known or not-reported. Thus, knowledge of the phylogenetic relationships of DENV viruses circulating in Yucatan was lacking. This information is important to understand the trafficking and evolutionary potential of the viruses. Timely reporting on specific genotypes circulating in a region would also be of great value to public health authorities.

Previously, single-stranded conformational polymorphism (SSCP) was used to characterize diversity in the prM gene of a collection of 140 Mexican DENV isolates mostly from the Yucatan peninsula (Farfan JA, PhD Dissertation, Colorado State University, 2001). That virus collection was comprised of 23 DENV-1, 10 DENV-2, 62 DENV-3 and 45 DENV-4 isolates. The prM gene was selected because it proved to be more diverse by SSCP and presumably more informative for genetic analyses than loci at E and NS5 genes of DENV-2 (Farfan et al., 1997). Six, four, four and eight banding patterns or "haplotypes" were found for DENV-1 to -4 respectively. That information was used in this work to select a sub-sample of isolates for a more in depth study of their genetic diversity and phylogenetic relationships; at least one isolate of each

haplotype was sequenced in the prM gene to characterize the DENV viruses circulating in the Yucatán peninsula of México during an 18 year period (1980-1997).

MATERIALS AND METHODS

Viruses. DENV viruses from the Yucatán peninsula were obtained from collections provided by the Centro de Investigaciones Regionales "Dr. Hideyo Noguchi" of the Universidad Autónoma de Yucatán and the Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention (CDC), San Juan Laboratories, San Juan, Puerto Rico. DENV viruses in these collections had been isolated from human serum, passed one to several times in C6/36 (*Ae. albopictus*) cells, and the serotype was determined by immunofluorescence analysis using serotype specific antibodies (Henchal et al., 1983). The sub-sample of isolates selected for this study was comprised of eight DENV-1, six DENV-2, eleven DENV-3 and seventeen DENV-4. Table 2-1 shows the year of isolation, municipality where the patient lived and SSCP haplotype designation for the 42 isolates. All of them were obtained from patients suffering DF.

RNA extraction and cDNA amplification. Viruses were propagated in C6/36 cells once prior to isolation of viral RNA. Total RNA was extracted from the infected cells by a guanidine thiocyanate/phenol-chloroform method (Lanciotti et al., 1992) or with the QIAamp® system according to the manufacturer's instructions (Qiagen Inc, Valencia, CA). Primers were designed to amplify cDNA segments from the prM region that were between 400 to 500 nucleotides (Table 2-2). The nucleotide sequences used in primer design were from the following strains: DENV-1 S275/Singapore90, DENV-2 NGC/New Guinea44, DENV-3 H-87/Philippines56 and DENV-4 814669/Dominica81.

Table 2-1. Place of origin and year of isolation of DENV isolates sequenced in this study. Isolates in bold were included in the phylogenetic analyses.

SEROTYPE	STATE	CITY	YEAR	ISOLATE DESIGNATION.
DENV-1	YUCATÁN	MÉRIDA	1980	**BC159,**
	UNKNOWN	UNKNOWN	1980	**1298**
	UNKNOWN	UNKNOWN	1983	**1378**
		MÉRIDA	1994	BC136, BC137, BC131, **BC119,** BC121, **BC126, BC128**
DENV-2	YUCATÁN	MÉRIDA	1994	**BC134**
		MÉRIDA	1996	**BC17**
	Q. ROO	UNKNOWN	1994	**BC133**, BC139
	JALISCO	UNKNOWN	1992	**BC122,**
	UNKNOWN	UNKNOWN	1983	**200787**
	UNKNOWN	UNKNOWN	1983	**SJL1421**
	UNKNOWN	UNKNOWN	1984	**SJL1482**
DENV-3	YUCATÁN	MÉRIDA	1995	**BC4**
		MÉRIDA	1996	BC21, BC22, BC29,
		HUNUCMA	1996	**BC285**
		HOCABA	1997	**BC192**, BC196
		TIMUCUY	1997	BC178
		TECOH	1997	**BC195**
	Q. ROO	CANCUN	1997	BC177
		UNKNOWN	1997	**BC197**
DENV-4	YUCATÁN	MÉRIDA	1984	**BC146**, BC147, **BC149, BC158,** BC143, BC195, BC196, BC207
		MÉRIDA	1995	**BC9**
		OXKUTZCAB	1995	BC5, BC10, **BC24**
		CHICXULUB	1996	**BC293**
		ABALA	1995	**BC6**
		TZUCACAB	1997	**BC287**
	Q. ROO	JUAREZ	1995	**BC8**

Viral RNA was reverse transcribed to cDNA using the respective serotype-specific reverse primer and SuperScript II RNaseH-reverse transcriptase (Life Technologies, Gaithersburg, MD) following the instructions of the manufacturers. Two µl of first-strand cDNA reaction were amplified in a total volume of 25 µl containing 0.5 mM of each forward and reverse primers in 10 mM Tris-HCl (pH 8.3), 50mM KCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP and 0.05u/ul of Taq polymerase (Promega, Madison, WI). PCR was performed as follows: an initial denaturation of one minute at 94°C, 30

cycles of 94°C for 1 min, annealing at 53-58°C (Table 2-2) for 1 min, and extension at 72°C for 2 min. Finally, the samples were incubated at 72°C for 10 min. The expected sizes of the amplified products were verified by electrophoresis in 2% agarose.

Table 2-2. Primer sequences, length of the amplified cDNAs and annealing temperatures used in the PCR of each DENV serotype.

Primer Name	Sequence (5'→3')	Product length (bp)	Annealing Temp. (°C)
DENV1F-314	CAAAGTGCTACGGGGTTTCAA	445	56.4
DENV1R-738	GGACATCCACGTTTCGGTTCT		
DENV2F-367	ATGCTGAACATCTTGAACAGG	504	53.9
DENV2R-850	TATGGTGTATGCCAGGATTGC		
DENV3F-385	ACGGAAAAAGACATCGCTCTG	392	56.2
DENV3R-758	CAGCCGACATCCAGGTTTG		
DENV4F-357	GAGATAGGCCGCATGCTGAAC	486	54.3
DENV4R-822	GTGGTGTTAAAGCTACTGAGC		

Sequencing and Sequence Analysis. Direct sequencing of both strands of the amplified products was performed with the same primers used for the PCR in ABI 377 DNA sequencers (Perkin-Elmer Inc., Foster City, CA). cDNA sequences were aligned using Seqman software (Dnastar Inc. Madison, WI). Sequences used in the phylogenetic analysis corresponded to the 273 nucleotides of the pr portion and 51-120 nucleotides of the M portion of prM gene. Sequences of relevant reference strains available in GenBank (Table 2-3) were trimmed accordingly and included in the alignment. Unpublished sequences of DENV-1 Mexican isolates 1298/Mexico80 and 1378/Mexico83 were kindly provided by Dr. F. Liprandi (Instituto Venezolano de Investigaciones Científicas, IVIC, Venezuela). Two additional DENV-2 reference strains were sequenced in the prM locus as part of this work (10/Somalia84 and 1592/Sri Lanka85).

Table 2-3. Reference strains used in the phylogenetic analyses. The genotype classification was taken from the following references: for DENV-1, Chunge *et al* 1995; for DENV-2, Lewis *et al* 1993; for DENV-3, Lanciotti *et al* 1994; for DENV-4, Lanciotti *et al* 1997. N.A.: non-available or non-classified.

Type	Strain	Country of Origin	Year of Isolation	Genotype	Genbank Accession No.
DENV-1	Mochizuki	Japan	1943	N. A.	S75335
DENV-1	17646	Nauru	1974	1	M23027
DENV-1	S275-90	Singapore	1990	1	M87512
DENV-1	AHF82-80	Thailand	1980	2	D00502
DENV-1	Abidjan/1056	Ivory Coast	1998	N. A.	AF298807
DENV-1	DJIB/98/606	Djibouti	1998	N. A.	AF298808
DENV-1	FGA/89	Fr. Guyana	1989	1	AF226687
DENV-1	GZ80	China	N. A.	N. A.	AF350498
DENV-1	836-1	Philippines	1984	3	D00503
DENV-1	CV1636-77	Jamaica	1977	2	D00501
DENV-1	40514	Brazil	1990	2	S64849
DENV-1	1298	Mexico	1980	N. A.	N. A.
DENV-1	1378	Mexico	1983	N. A.	N. A.
DENV-2	PR159	Puerto Rico	1969	V	M19197
DENV-2	200787	México	1983	N. A.	L04561
DENV-2	VEN2	Venezuela	1987	N. A.	AF100465
DENV-2	FJ-10	China	N. A.	N. A.	AF276619
DENV-2	1592	Sri Lanka	1985	IV	N. A.
DENV-2	#10	Somalia	1984	IV	N. A.
DENV-2	NGC	New Guinea	1944	I	M29095
DENV-2	16681	Thailand	1964	IIIa	M84727
DENV-2	1409	Jamaica	1983	IIIb	M20558
DENV-2	PUO218	Thailand	1980	IIIa	D00345
DENV-3	H-87	Philippines	1956	I	L11423
DENV-3	29472	Fiji	1992	I	L11422
DENV-3	85-159	Indonesia	1985	I	L11425
DENV-3	29586	Malaysia	1981	I	L11427
DENV-3	MK315	Thailand	1987	II	L11442
DENV-3	CH3489D	Thailand	1973	II	L11620
DENV-3	5987	Thailand	1962	II	L11440
DENV-3	1416	India	1984	III	L11424
DENV-3	2783	Sri Lanka	1991	III	L11438
DENV-3	1558	Mozambique	1985	III	L11430
DENV-3	1696	Samoa	1986	III	L11435
DENV-3	1327	Tahiti	1965	IV	L11439
DENV-3	1340	Puerto Rico	1977	IV	L11434
DENV-3	PR6	Puerto Rico	1963	IV	L11433
DENV-4	H241-P	Philippines	1956	I	S66064
DENV-4	B5	China	N. A.	N. A.	AF289029
DENV-4	814669	Dominica	1981	II	M14931

Separate maximum likelihood phylogenetic trees were constructed for each DENV serotype using the tree-bisection and reconnection (TBR) algorithm and the Hasegawa-Kishino-Yano (HKY) model of nucleotide substitution. Each tree was rooted with sequences of the three other serotypes. Strains used as outgroups were: DENV-1 17646/Nauru 74, DENV-2 PR159/Puerto Rico 69, DENV-3 H-87/Philippines 56 and DENV-4 814669/Dominica 81. Five hundred bootstrap replicates were used to assess the robustness of the inferred topology. These analyses were conducted with the PAUP* program version 4.04 (Swofford, 1998).

RESULTS

A number of sequenced Mexican isolates exhibited identical or very similar sequences and were excluded from the phylogenetic analysis. The remaining sequences were submitted to GenBank (Accession numbers AF459605 to AF459627). The trees resulting from the phylogenetic analysis are shown in Figure 2-1 and are described in more detail below.

DENV-1. The 324 nucleotides analyzed (positions 437-760 in the DENV-1 genome) contained 103 phylogenetically informative characters. The phylogenetic tree showed three major branches but none of them is supported by a high bootstrap value. One of them had a single isolate from West Africa (Ivory Coast 98); a second branch formed a clade with isolates from the Pacific Islands, East Asia and East Africa; the last branch contained all the isolates from the Americas, as well as a single isolate from the Philippines. The 6 isolates from Yucatán and the two other Mexican isolates differed by 1% or less and clustered together in the phylogenetic tree with the isolates 0514/Brazil90 and FGA/French Guyana89. They were more distantly related to

CV1636/Jamaica77 and 836-1/Philippines84 (Figure 2-1A). Two sequences from 1980 (BC159/Mérida80 and 1298/México80) were identical. These could represent the same isolate present in two different collections.

DENV-2. The 357 nucleotides analyzed (positions 439-795) contained 104 phylogenetically informative characters. The phylogenetic three exhibited three major clades corresponding to the previously named American, Sri Lanka and Southeast Asian genotypes but the latter is not well supported. Six viruses in the Mexican collection that were isolated before 1995 diverged by less than 2% and clustered with older viruses of the Americas (genotype V in Lewis *et al.,* 1993), such as PR159/Puerto Rico69, VEN2/Venezuela/7 and 200787/México83 (Figure 2-1B). The SJL1421/México83 isolate was identical to VEN2/Venezuela87. The most recent Mexican isolate, BC17/Mérida96, diverged between 8.9% and 10.2% from the other Mexican isolates and clustered with strains like 1592/Sri Lanka85, FJ-10/China and 10/Somalia84 in the Sri Lanka genotype (genotype IV in Lewis *et al.,* 1993).

DENV-3. The 330 nucleotides analyzed (positions 437-766) contained 93 phylogenetically informative characters. None of eleven sequenced Mexican isolates differed by more than one nucleotide (0.3%) from the consensus sequence of the group. The phylogenetic tree showed four major branches (Figure 2-1C). All the Mexican isolates clustered with those in the "Sri Lanka/India" genotype (group III in Lanciotti *et al.,* 1994) but this grouping is not well supported in the bootstrap. The greatest similarity was found with the 1558/Mozambique85 isolate, which diverged by less than 1% from the Mexican viruses.

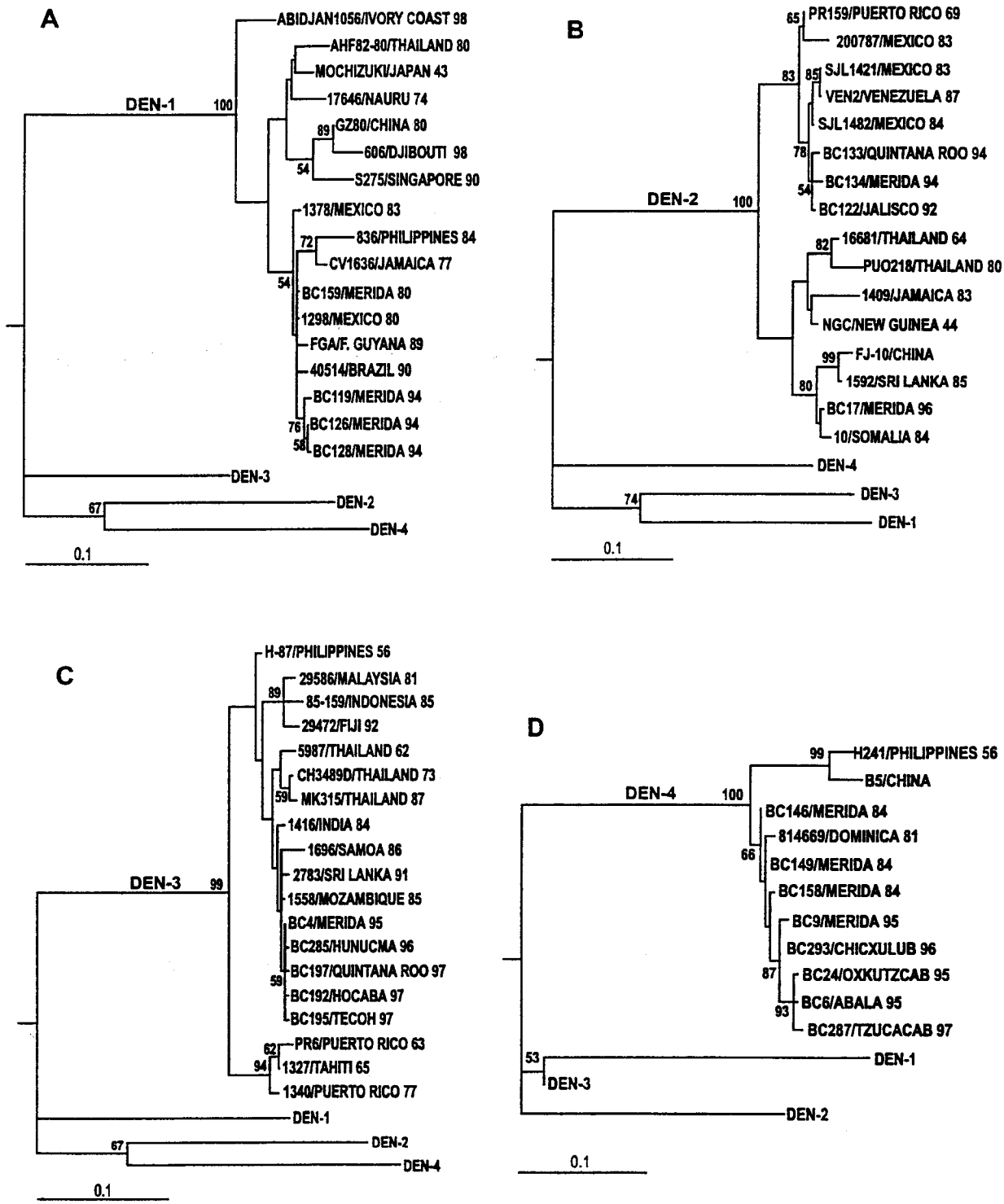


Figure 2-1. Maximum likelihood phylogenetic trees for the prM locus of (A) DENV-1, (B) DENV-2, (C) DENV-3 and (D) DENV-4 viruses. The numbers displayed next to the nodes correspond to the bootstrap values (500 replicates) supporting that clade. Trees were constructed with PAUP* 4.0 (Swofford, 1998) and drawn using Drawgram (Felsenstein, 1993).

DENV-4. The 393 nucleotides analyzed (positions 441-833) contained 93 phylogenetically informative characters. Only two major branches appeared in the phylogenetic tree (Figura 2-1D). The Mexican isolates of 1984 clustered in a single branch with the only other American strain 814669/Dominica81 from which they differed less than 1%. Viruses isolated from 1995 to 1997 formed a separated clade within the same branch and differed between 1.8% and 3.2% from the earlier isolates. The prototype strain H241-P/Philippines56 differed between 6.0% and 7.8% from the Mexican isolates and clustered with the B5/China isolate.

DISCUSSION

In agreement with the results obtained by SSCP (Farfan JA, PhD Dissertation, Colorado State University, 2001) the genetic diversity at the sequence level in this collection was maximal for DENV-2, minimal for DENV-3 and intermediate for DENV-1 and -4. These differences in diversity are partially explained by the different history of the serotypes in Yucatan. DENV-3 was isolated for the first time in 1995 and the isolates analyzed encompassed only three years of circulation. It could be speculated that the DENV-3 strain originated from single introduction from Central America and it was still in an early time in the process of diversification. The higher diversity in the other serotypes could be due to repeated introductions of the same genotype or, alternatively, to their persistence for a longer time, which could generate some degree of *in situ* evolution. DENV-1 and DENV-4 were originally identified in 1979 and 1984, respectively. DENV-2 was detected in 1991 in Yucatan but the collection analyzed included isolates obtained during the 1980's and 1990's in different Mexican states. However, the much higher diversity found in DENV-2, which include dissimilarities of up to 10%, cannot be thoroughly explained but by the introduction of different genotypes in the region, a situation not seen in any other serotype.

Sequence data were used to determine phylogenetic relationships of the viruses within each serotype. Previously, two phylogenetic studies analyzing a large number of DENV-1 genomic sequences have been published (Chungue et al., 1995, Rico-Hesse, 1990). They analyzed either a 240 nt sequence in the E-NS1 junction region of the genome(Rico-Hesse, 1990) or a 180 nt sequence in the E gene (Chungue et al., 1995) and suggested that there were potentially 5 or 3 DENV-1 genotypes, respectively. The phylogenetic relationships in those studies were not comparable. Our analysis of a 324 nt sequence in the prM region of DENV-1 viruses generated a different tree with three major branches (Figure 2-1A). The different outcomes in this and the other studies may be partially attributable to the different isolates studied and different genomic segments analyzed. All the DENV-1 Mexican isolates that we analyzed clustered in a single branch. The possibility that a different DENV-1 genotype had circulated there in 1980, suggested by the location of the isolate 1298/México80 in the phylogenetic tree of the E-NS1 junction (Rico-Hesse, 1990), was not supported in this work, since that isolate appeared in the same branch containing Mexican DENV-1 isolates of 1983 and thereafter. The apparent different phylogenetic history of the E-NS1 genomic region of this isolate is not due to a recombinant event; a recent study showed that the 1298/Mexico80 isolate is also closely related to other American DENV-1 viruses in the E gene (Goncalvez et al., 2002).

DENV-2 viruses are genetically the most diverse among the DENV serotypes. At least four different genotypes have been identified in previous studies (Lewis et al., 1993, Rico-Hesse, 1990, Rico-Hesse et al., 1997).They are the "sylvatic", "Americas", "Sri Lanka" and "Southeast Asia" genotypes. Most cases of DHF/DSS have been associated with this latter genotype. It includes not only viruses isolated in Southeast

Asia but also many of the DENV-2 isolates found in the Americas since 1981. Previous studies of the E-NS1 gene junction have shown that DENV-2 Mexican isolates belong either to the Americas or to the Southeast Asia genotypes depending on whether they were isolated before 1994 or thereafter (Rico-Hesse et al., 1997) In this study all the Mexican viruses isolated until 1994 were classified in the Americas genotype, but the BC17/Mérida96 isolate of 1996, surprisingly clustered in the DENV-2 Sri Lanka genotype (Figure 2-1B). This finding was confirmed by sequencing and phylogenetic analysis of the E-NS1 region (data not shown). This is the first DENV-2 isolate from the Western Hemisphere that segregated with this genotype. This isolate was most closely related to the 10/Somalia84 isolate, from which it differed in only four nucleotides (1.1 %) in the prM gene. It diverged from the viruses in the Southeast Asia genotype by 4.6% to 8.1%. Since five Mexican DENV-2 viruses of 1995 studied by Rico-Hesse and others (Rico-Hesse et al., 1997) including four from the neighbor state of Chiapas, belong to the Southeast Asia genotype, it seems that at least three different DENV-2 genotypes have circulated in southern México, two of which arrived there by the same time.

Phylogenetic relationships of DENV-3 viruses have also been investigated. Chungue *et al* (1993) compared 194 bases of the E gene of 27 isolates mostly from the Pacific and Southeast Asia and found four genotypes diverging by 6% or more. Lanciotti *et al.* (1994) compared the entire sequence of prM/M/E genes of 23 viruses, and also found four genotypes. The two studies only concurred in genotype IV, which contained old American and old Polynesian isolates. Differences in the selected isolates and length of the segments studied might account for the differences. Our DENV-3 phylogenetic tree, based on 330 bases of the prM gene, resembles that of the latter study (Lanciotti et al., 1994).

DENV-3 was absent from the Western Hemisphere for 17 years (1978-1993) and reappeared simultaneously in Nicaragua and Panama in 1994 (CDC, 1995, Guzman et al., 1996a). It was soon dispersed to other countries in Central America and Mexico and later to the Antilles and South America. An early isolate from these outbreaks (Nicaragua 1994) was partially sequenced and included in the phylogenetic tree of Chungue *et al.* (1993), where it clustered with viruses of genotype 2, which includes isolates from Southeast Asia (Guzman et al., 1996a). Another early DENV-3 isolate from Panama 1994 was genotyped at CDC and classified in the Sri Lanka/India genotype (CDC, 1995). This analysis of isolates from Yucatan, as well as the analysis of isolates from Nicaragua 1998 (Harris et al., 2000), supports the grouping of the new DENV-3 strain now circulating in the Americas in the DENV-3 Sri Lanka/India genotype (Figure 2-1C).

Two phylogenetic analyses of a significant number of DENV-4 isolates produced similar results (Chungue et al., 1995, Lanciotti et al., 1997); both studies showed two well-defined groups of DENV-4 isolates. The first included viruses from Southeast Asia, Philippines, Sri Lanka and Senegal, and the second comprised viruses from the South Pacific, the Americas and some from Indonesia. Although our phylogenetic analysis of DENV-4 was significantly limited because there were only three sequences of the DENV-4 prM region in GenBank, it also shows two major branches that could correspond to the previously described genotypes. Both the earlier (1984) and the later (1994-1997) Mexican DENV-4 viruses, cluster in the same branch with the 814669/Dominica81 isolate (Figure 2-1D). The low divergence between the two subgroups of Mexican DENV-4 isolates (1.8-3.2%) is compatible with the rate of nucleotide substitution reported for this virus (Lanciotti et al., 1997), which suggests

that the virus introduced in Yucatan in 1994 is the same strain that caused the 1984 outbreak (Lorono-Pino et al., 1993) and that both are derived from the original DENV-4 virus that appeared for the first time in the Americas in Dominica in 1981.

Major limitations in this study were the low number of non-Mexican isolates included in the analyses and the low bootstrap support found for many of the clades, even those previously defined as genotypes. The first problem was due to the scarcity of sequences of the prM gene available in GenBank or other databases. This limitation was more critical for DENV-1 and -4, for which the previously described genotypes are barely recognizable. The phylogenetic trees of DENV-2 and -3 presented here showed topologies that resemble those previously published, but some genotypes, like the sylvatic DENV-2 genotype, are absent. The low bootstrap values found in many clades could be due to the shortness of the analyzed segment or to the presence of homoplasies. Although the number of phylogenetically informative characters (93-104) seems sufficient for a robust phylogeny, many of these characters differ only among serotypes used as outgroups and are not discriminative among the sequences in the ingroups.

The most important change in the epidemiology of DENV in México in the last two decades is the abrupt increase in the number of cases of DHF/DSS in 1995 (Anonymous, 1996, Briseno-Garcia et al., 1996). That year all four DENV serotypes simultaneously circulated, DENV-3 was detected for the first time in the country, and the Southeast Asian genotype of DENV-2 also appeared (Rico-Hesse et al., 1997), closely followed by the DENV-2 Sri Lanka genotype. Which of these changes in DENV circulation conditioned the worsening of the epidemiological situation? The data are limited but provide some hints. Twenty of 25 isolates obtained from DHF/DSS cases

were DENV-2 (Briseno-Garcia et al., 1996). Some of them were classified as DENV-2 genotype Southeast Asia (Rico-Hesse et al., 1997), which is well known to contain virulent viruses. The role of the DENV-2 genotype Sri Lanka described in this study is difficult to assess since the only virus available (BC17/Merida96) was isolated from a ten year old patient who suffered a primary infection and clinically presented as a DF case (J. A. Farfan, unpublished data). It is provocative, however, that the year of introduction (1996), coincided with the emergence of DHF in Yucatan. The eventual contribution of the new DENV-3 strain to DHF in the Yucatan also remains to be determined. This DENV-3 Sri Lanka/India genotype had been associated with mild disease in the early 1980's, but it was identified later in outbreaks of severe cases in Mozambique in 1984-1985 (Gubler et al., 1986), Sri Lanka 1989 and India 1990 (Lanciotti et al., 1994). Its emergence in 1994 in the Americas was temporally associated with cases of DHF in Nicaragua but data are lacking about its role in more recent DHF outbreaks in the Americas.

Another issue to consider is the route of arrival of the new DENV-2 and DENV-3 viruses found in Yucatan. It is noteworthy that both viruses are phylogenetically related to viruses that have circulated in similar geographic regions. The DENV-3 Sri Lanka/India genotype viruses have been reported in Sri Lanka 1981-1991, India 1984, Mozambique 1985 and Samoa 1986 (Lanciotti et al., 1994). The DENV-2 Sri Lanka genotype viruses have been found in Malaysia 1969, Indonesia 1973-1978, Seychelles Islands 1997, Sri Lanka 1981-1990, Burkina Faso 1982-1986 and Somalia 1984 (Lewis et al., 1993, Rico-Hesse, 1990). It is also interesting that both new viruses exhibit the closest relationship in the prM gene with strains from East Africa (DENV-3 Mozambique 1985 and DENV-2 Somalia 1984) and that they were introduced to Mexico about the same time (1995-1996). This suggests trafficking of viruses from the

Indian Ocean or Africa to the Americas. Commerce and travel between India and eastern African countries and Pacific Islands have been implicated in the dispersal of DENV viruses to Africa and Polynesia, respectively (Lanciotti et al., 1994, Rico-Hesse, 1990). To understand the origin of the new DENV-2 virus in Yucatan, more information is needed about the presence of the Sri Lanka genotype viruses in recent years in Africa, Southeast Asia and the Pacific Islands, as well as a more detailed study of the DENV-2 viruses prevalent in Mexico and Central America in recent years.

CHAPTER 3

THE MOLECULAR EPIDEMIOLOGY OF DENGUE VIRUS IN MEXICO

Although dengue fever (DF) had been known for a long time and major epidemics had been described since the 18th century, it was not until the middle of the 20th century that DEN epidemics emerged as major public health problems in most tropical and subtropical areas throughout the world (Gubler, 1998, Gubler & Clark, 1995, Guzman & Kouri, 2002). Beginning in World War II, a generalized increase in the incidence of DF occurred in Southeast Asia and the Pacific Islands. Soon after, it moved to the Philippines, India, Sri Lanka and other islands in the Indian Ocean. The epidemic wave then extended north to China and Taiwan and west to Africa, Pakistan and the middle east (Gubler & Clark, 1995). In 1953 a more severe form of the illness known as dengue hemorrhagic fever (DHF) was recognized for the first time in the Philippines (Hammon et al., 1960). Soon thereafter DHF began to be diagnosed in most Southeast Asian countries and has steadily increased in frequency and has become a leading cause of hospitalization and death among children of that region. DHF also radiated from Southeast Asia in most directions including countries around the Indian Ocean, the Pacific and Australia (Gubler, 1998).

In the Americas DEN has been present since the 17th century, but most outbreaks were isolated in time and space (Gubler, 1998, Pinheiro, 1989) and DHF

was not known until recently. During the 1950's and 1960's, the Pan American Health Organization (PAHO) led a campaign to eradicate *Ae. aegypti* and to prevent urban yellow fever in the Americas. Control of the mosquito vector also resulted in the disappearance of DF in most Central and South American countries. The program was not as successful in the Caribbean islands where DF was detected sporadically during the times of the eradication. DENV-2 and DENV-3 were isolated in a few cases (Gubler, 1998, Gubler & Clark, 1995). During the 1970s the PAHO campaign gradually eroded and *Ae. aegypti* reappeared in most countries. DEN followed soon after, and in 1977 and 1981 DENV-1 and DENV-4, respectively, were introduced and caused repeated outbreaks through the region (Pinheiro, 1989). DF was the most frequent clinical presentation although a few severe cases were reported. DHF/DSS emerged abruptly in 1981 in Cuba and caused about 150,000 hospitalizations and more than 150 deaths in less than four months (Kouri et al., 1986). An intensive eradication program stopped the outbreak and DHF did not move beyond that island. DHF appeared in Venezuela in 1989, and soon after it was detected in other countries along the Atlantic coast of South America from Colombia to Brazil. DHF was not diagnosed in Central America until 1994, concurrently with the reintroduction of DENV-3, which had not been detected in the Americas since 1977 (Gubler & Clark, 1995).

In Mexico, from 1947 to the 1960's, health authorities conducted intensive campaigns to eradicate *Ae. aegypti*, and the mosquito was officially declared eradicated from the country in 1963 (Navarrete, 2002). However, *Ae. aegypti* reappeared two years later. Major epidemics of DF caused by DENV-1 occurred on the eastern coast of Mexico during 1979 and 1980. Approximately 17,000 cases of DEN were reported in 1981 (Gomez-Dantes, 1992). DENV-4 caused a major epidemic in 1984 in the Yucatan peninsula; there were more than 5,000 reported cases and four

fatalities, but only one case met all WHO criteria for DHF (Lorono-Pino et al., 1993). In 1984 and 1985, DF was diagnosed in 25 of 32 states and DENV-1, 2, and 4 were reported throughout Mexico (Pinheiro, 1989). In 1995 DENV-3 was isolated for the first time and a few cases of DHF were confirmed (Briseno-Garcia et al., 1996). DHF appeared in epidemic proportions in 1996 and 1997. DENV-3 was the predominant serotype isolated in these years but all the others were also present. Following a period of relative calm, DHF emerged again in 2001 and 2002 with incidences similar to those of the mid-1990s and a greater proportion of severe cases (Navarrete, 2002). The incidence of DF and DHF in Mexico in the last decades is summarized in Figure 3-1.

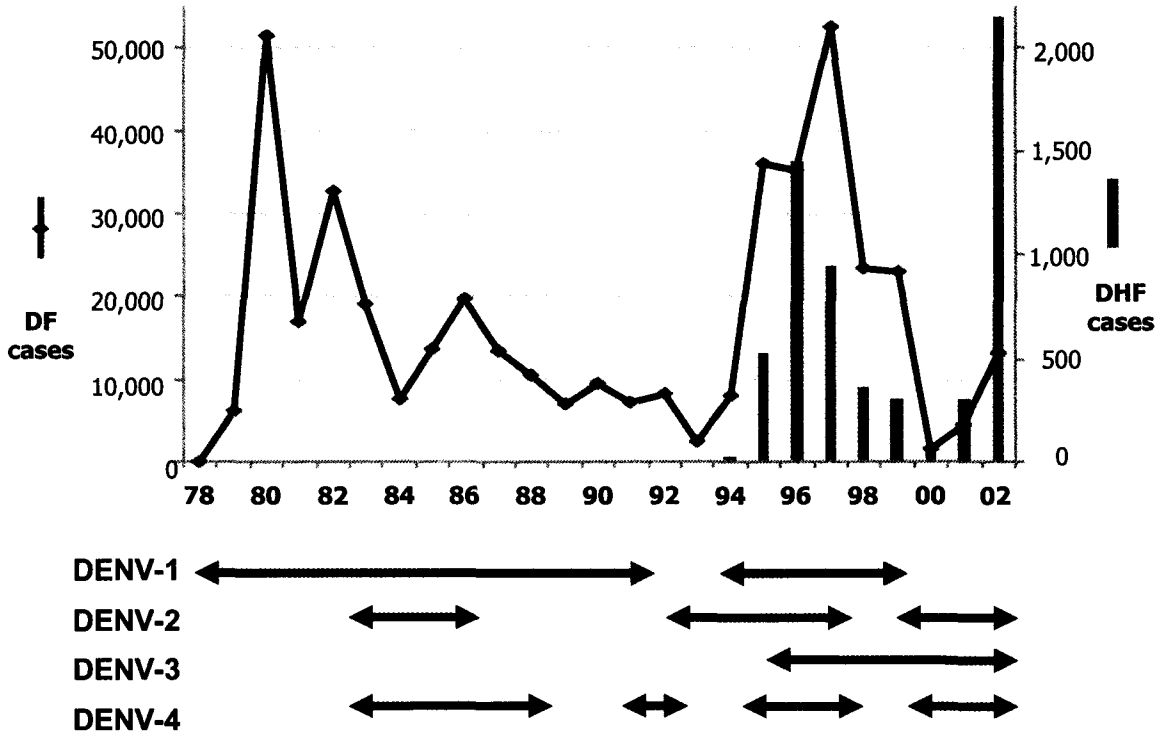


Figure 3-1. Number of cases of DF and DHF reported in Mexico between 1978 and 2002 and serotypes circulating by year. (Source: Secretaria de Salud, Mexico)

It is usually thought that DEN is periodically introduced into Mexico via the frontier with Guatemala, because the southern states usually have the most human cases (Navarrete, 2002). However, the disease has been frequent in most states along the Pacific and Atlantic (gulf) coasts. Two different epidemiological patterns of DENV transmission have been identified in Mexico: 1) an endemic pattern in coastal states of the Gulf of Mexico and 2) a seasonal pattern in the Pacific coast and the Yucatan peninsula (Figure 2-2) (Navarrete, 2002). However, a recent study in the western state of Colima demonstrated focalized, low-level transmission during the interepidemic periods (Espinoza-Gomez et al., 2003).

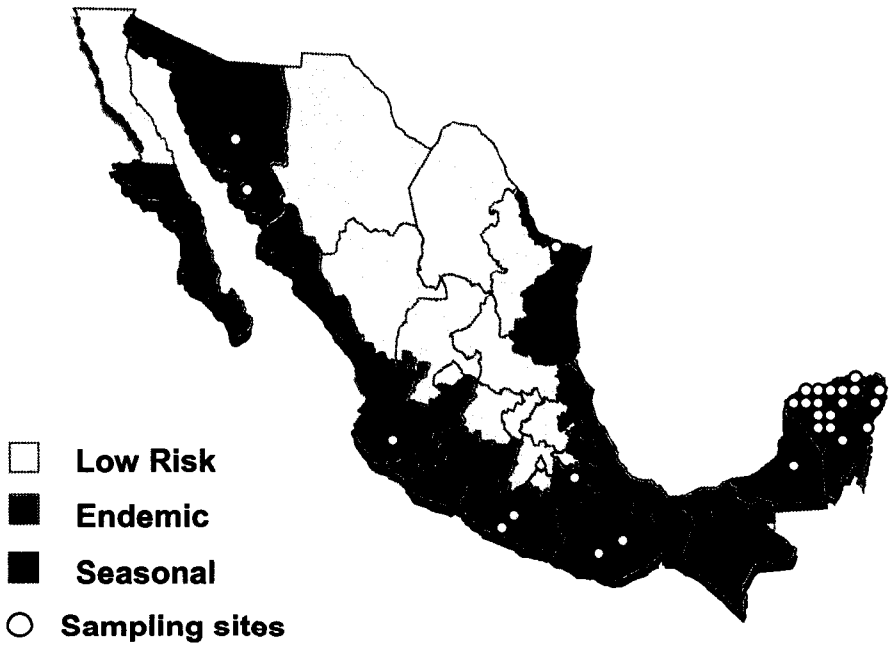


Figure 3-2. Distribution of the patterns of transmission of DENV in Mexico and sites of isolation of the viruses analyzed in this study. The place of isolation was not known for thirteen isolates.

The causes of the increased incidence of dengue in the last century have been reviewed and include demographic, cultural, environmental and political changes that were already mentioned in the first chapter of this dissertation (Gubler, 1998, Kuno, 1995, Monath, 1994). The causes for the more severe clinical outcome of DENV infections are not so clear. One hypothesis is that simultaneous or closely sequential epidemics are now resulting from the increase in virus trafficking of different DENV serotypes and strains. This would lead to an increase in the frequency of secondary infections and more common occurrence of the immune-enhancement phenomenon (Monath, 1994), which is a major risk factors for DHF/DSS (Halstead, 1988, Monath, 1994). The other hypothesis is that increased travel and commerce have facilitated the dispersal of DENV genotypes that are more fit and more virulent at the same time (Armstrong & Rico-Hesse, 2003, Rosen, 1977). Such strains would displace other less virulent, indigenous strains that would have otherwise immunized the populations without causing significant mortality. There is evidence to support either hypothesis; however, it seems more likely that both factors are working in concert to produce the current situation of more frequent and more severe DENV infections.

To determine the role of virus strains in epidemic DF and DHF, it is necessary to have techniques that detect differences between strains within serotypes and to have collections of isolates representative of the specific epidemic or endemic situation of interest. Immunological techniques have been used to track the movement of some viruses (Henchal et al., 1986, Russell & McCown, 1972), but they have low resolution below the serotype level.

Molecular genetic methods are more powerful for detecting subtle differences among DENV strains and can be utilized to trace the origin, dispersion and evolution of microorganisms causing isolated infections or major epidemics. The first technique used with DENV was the ribonuclease T_1 oligonucleotide fingerprinting, which allowed classification of strains within serotypes into "topotypes". With this procedure a rough estimation of the genetic diversity in each serotype was obtained, and the trafficking of some strains was inferred (Repik et al., 1983, Trent et al., 1983). However this method is laborious, has limited sensitivity, and was rapidly displaced by genome sequence analysis. Other molecular epidemiologic techniques that have been utilized include amplified fragment length polymorphism (AFLP) (Vorndam et al., 1994), single-strand conformational polymorphism (SSCP) (Diaz et al., 2002, Farfan et al., 1997) and restriction site-specific polymerase chain reaction (RSS-PCR) (Harris et al., 1999, Miagostovich et al., 2000). These techniques can still be used as screening methods, but none of them compare to sequencing in sensitivity and versatility. Coupled with traditional and new analytical methods, sequencing has revealed DENV ancestral relationships, virus trafficking and evolutionary processes that were undetectable using the aforementioned methods. Sequence analysis can also be used to investigate other important evolutionary parameters and statistics like, substitution rate, nucleotide diversity and test for natural selection.

By the late 1980's sequence analysis was applied to a limited number of DENV strains (Blok et al., 1989, Chu et al., 1989). The first phylogenetic analysis of a significant number of DENV serotypes 1 and 2 strains was attempted by Rico-Hesse (1990). Using a 6% divergence cut-off, she found that both serotypes segregated into five groups or "genotypes" that correlated with geographic origin. Phylogenetic studies soon revealed four genotypes of DENV-3 and one genotype of DENV-4 using the

arbitrary 6% cut-off (Chungue et al., 1995, Chungue et al., 1993). All these early works suffered from two limitations: 1) the segments analyzed were short (180-240 nucleotides) and 2) branch support methods were not used to assess the robustness of the resulting tree topology. In more recent years analysis of the entire E gene has become the standard for DENV phylogenies (Goncalvez et al., 2002, Lanciotti et al., 1997, Lanciotti et al., 1994, Twiddy et al., 2002, Wang et al., 2000). Of utmost importance was the discovery of the sylvatic DEN genotypes, which allowed inferring the ancestral nature of the enzootic cycle in the evolution of DENVs (Wang et al., 2000). These studies have revealed extensive genetic variability among isolates of every DENV serotype and have led to the recognition of three to six different genotypes within the each serotypes. Genotypes can differ significantly in their ability to cause severe DEN. For example in the Americas, infection with the "native" DENV-2 American genotype typically results in DF, whereas infection with recently introduced Asian genotypes frequently results in DHF/DSS (Rico-Hesse et al., 1997, Watts et al., 1999).

Knowledge of the phylogenetic relationships of DENV is important to understand the trafficking and epidemic potential of the viruses. Previously, an attempt to determine the phylogenetic origin of DENV strains circulating in the Yucatan peninsula of Mexico using sequences of the prM gene produced mixed results (Diaz et al., 2002 and previous chapter). Information regarding the diversity and phylogenetic relationships of DENV strains that circulated between 1980 and 1997 was obtained, allowing identification of some of the genotypes involved and, in some cases, inferences to be made about possible routes of their introduction. However, those phylogenies resolved neither the routes of DENV trafficking in the Western Hemisphere nor ancestral relationships with genotypes and strains that were not represented in

those analyses. The scarcity of prM sequences available and the limited number of informative characters in that gene generated phylogenies with limited resolution and poor bootstrap support (Figure 2-1).

In this chapter, sequencing and analyzing the full E gene of all four DENV serotypes addressed the limitations found in the prM analysis. The plethora of sequences available for comparison in this genomic region and the larger sequence of this gene provide a large number of phylogenetically informative characters that enabled robust analyses. The geographic representation of the Mexican DENV collection was expanded to at least nine states and was updated until 2002. Thus, it was possible to describe the origin, geographical dispersion and extinction of DENV strains in Mexico and other countries of the Americas. It was also possible to associate recent changes in the incidence and severity of DEN with the introduction and circulation of different serotypes, genotypes and strains of DENV. Additionally, some evolutionary parameters and statistics were computed and used to derive useful information about forces acting in DENV evolution.

MATERIALS AND METHODS

Viruses. Most of the viruses analyzed (Tabla 3-1) were obtained from the "Centro de Estudios Regionales Dr. Hideyo Noguchi" of the Universidad Autonoma de Yucatan, Merida, Mexico. Other isolates were obtained from collections belonging to the University of Texas Medical Branch, Galveston (UTMB), the Dengue Branch Center for Disease Control and Prevention, San Juan Laboratories (SJL), Puerto Rico, and the Instituto Nacional de Higiene (INH), Mexico City. All viruses analyzed were isolated from DF cases.

Table 3-1. DENV isolates from Mexico sequenced and/or analyzed in this study. * Isolates sequenced in one of the following studies: Goncalvez et al 2002 (DENV-1), Leitmeyer et al or Armstrong and Rico-Hesse (DENV-2), Uzcategui et al (2003), Foster et al (2003)

Isolate ID	Alternative ID	Serotype	State	Municipality	Year of Isolation
1298*	BC159/95	1	Yucatan	Merida	1980
1379	BC163/95	1	Unknown	Unknown	1982
1378*	BC162//95	1	Unknown	Unknown	1983
1462	BC166/95	1	Puebla	Unknown	1984
1463	BC167/95	1	Sonora	Unknown	1984
1756	BC175/95	1	Unknown	Unknown	1986
BC126	BC126/94	1	Yucatan	Merida	1994
BC7	BC7/96	1	QuintanaRoo	Cozumel	1995
BC264	BC264/95	1	Campeche	Campeche	1995
1421	BC165/95	2	Unknown	Unknown	1983
1482	BC200/95	2	Unknown	Unknown	1983
0131*		2	Sonora	Navojoa	1992
BC139	BC139/94	2	QuintanaRoo	Unknown	1994
328298*		2	Tamaulipas	Unknown	1995
BC17	BC17/97	2	Yucatan	Merida	1996
C932		2	Guerrero	Acapulco	1997
C1077		2	Guerrero	Chilpancingo	1997
Oax468*		2	Oaxaca	Juchitan	2000
11936		2	Yucatan	St. Elena	2001
12021		2	Yucatan	Oxcutzcab	2001
12914		2	Yucatan	Tekax	2001
13381		2	Yucatan	Chochola	2002
13382		2	Yucatan	Tizimin	2002
13404		2	Yucatan	Opichen	2002
6097*		3	Unknown	Unknown	1995
BC4	BC4/96	3	Yucatan	Merida	1995
BC285	BC285/97	3	Yucatan	Hunucma	1996
BC192	BC192/97	3	Yucatan	Hocaba	1997
BC177	BC177/97	3	QuintanaRoo	Cancun	1997
BC197	BC197/97	3	QuintanaRoo	Unknown	1997
Oaxaca00		3	Oaxaca	Unknown	2000
1414	BC181/95	4	Unknown	Unknown	1983
1420	BC182/95	4	Unknown	Unknown	1983
1503	BC196/95	4	Unknown	Unknown	1984
1551	BC192/95	4	Unknown	Unknown	1985
1554	BC193/95	4	Unknown	Unknown	1985
Mex91*		4	Unknown	Unknown	1991
Mex95*	111	4	Unknown	Unknown	1995
BC5	BC5/96	4	Yucatan	Oxcutzcab	1995
BC8	BC8/96	4	QuintanaRoo	B. Juarez	1995
BC26	BC26/97	4	Yucatan	Izamal	1996
BC287	BC287/97	4	Yucatan	Tzucacab	1997

Isolates used in the analysis were selected to represent most years and regions with DEN activity in Mexico. The origins of the samples that were analyzed in this study are shown in Figure 3-2 and Table 2-1 This information was unknown for thirteen isolates. There were geographic regions for which no strains were available, and there were a few strains that represented a single or few years. Most strains were between second and fourth cell culture passage in *Aedes albopictus* C6/36 cells. An additional passage in the same cell system was necessary to obtain viral RNA.

Molecular procedures. Viral genomic RNA was extracted from cell culture supernatant using a commercial silica-based method (QIAamp viral RNA, Qiagen, Valencia, CA). RNAs were copied into cDNAs using either random hexamers or a reverse primer annealing in the NS1 gene (Table 3-2) and RAV-2 reverse transcriptase (Amersham, Piscataway NJ). cDNA segments between 2,474 and 2,577 nucleotides encompassing the prM and E genes of DENV were amplified by PCR using the consensus forward primer D1 and serotype-specific reverse primers (Table 3-2). DNA polymerase used was *Taq* DNApol (Promega, Madison WI).

PCR product length and uniqueness was verified by 0.8% agarose gel elecrophoresis. When additional DNA bands were present, the correct PCR product was isolated in crystal violet-stained 0.8% agarose gels. The PCR products were then purified using either the QIAquick PCR purification kit or the QIAquick gel extraction kit (Qiagen). Both strands of the amplicons were sequenced using the primers listed in Table 3-2 and the Big Dye Terminator Cycle Sequencing System (ABI, Foster City, CA).

Table 3-2. Primers used for E gene amplification and sequencing

Primer	Sequence 5' → 3'	Use
D1	TCAATATGCTGAAACGCGCGAGAAACCG	PCR
D1-682F	AACCGGYGAACACCGACGAGA	Sequencing
D1-1064F	GAACTCTTGAAGACGGAGGTCACGAA	Sequencing
D1-1167R	TTGTTCTTCCACCAGTGTAGCCTCTC	Sequencing
D1-1488F	GCTCACCTAGAACAGGGCTGGACTTT	Sequencing
D1-1649R	TTCTTTGCATGAGCTGTCTTGAATGT	Sequencing
D1-1972F	CCAGAATGGGAGATTGATAACA	Sequencing
D1-2125R	CGGTTGCTTCGAACATTTTCCCTATG	Sequencing
D1-2726R	ATGGGTTGTGGCCTAATCAT	RT, PCR, Sequencing
D2-739F	ATGGGATTGGAGACACGAACTGAA	Sequencing
D2-1184F	ATGAAGAGCAGGACAAAAGGTT	Sequencing
D2-1225R	CCATTTCCCCATCCTCTGTCTAC	Sequencing
D2-1540F	GAAGACAAAGCTTGGCTGGTG	Sequencing
D2-1648R	GCATGGGGATTTTTGAARGTGAC	Sequencing
D2-2028F	GGAGGTTCTGCTTCTATGTTGACT	Sequencing
D2-2120R	TCAAACATTTGGCCGATRGAACTTC	Sequencing
D2-2588R	TCTTGTTACTGAGCGGATTC	RT, PCR, Sequencing
D3-385F	ACGGAAAAAGACATCGCTCTG	Sequencing
D3-644F	CTTACATCAACATGGGTGACTTAT	Sequencing
D3-758R	CAGCCGACATCCAGGTTTG	Sequencing
D3-1343F	TACACCGTCATCATCACAGTG	Sequencing
D3-1390R	CTTACATCAACATGGGTGACTTAT	Sequencing
D3-1671F	AAGAAGTAGTTGTCCTTGGAT	Sequencing
D3-1953R	CATTGTGAGCTTTCCCTTGTC	Sequencing
D3-2429R	TTCTTTGCCTTTCCAGTTTAT	RT, PCR, Sequencing
D4-762F	GAGACATGGATGTCATCGGAAGG	Sequencing
D4-1240F	GGGGCAATGGCTGTGGCTTGTT	Sequencing
D4-1308R	TTGGACCAAATTGCCTGTTATCTT	Sequencing
D4-1667F	YCATGCCAAGAGACAGGATGTGAC	Sequencing
D4-1702R	GCAAGAATGCATGGCTCCTTCCTGAG	Sequencing
D4-2158R	GAATGGCCATTCGTTTTGCACCTC	Sequencing
D4-2052F	CCYTTTGGGGACAGCGCTACATA	Sequencing
D4-2649R	TGTCCTCCTTCCCAGAGAACATAGTT	RT, PCR, Sequencing

Phylogenetic Analysis. Sequences corresponding to the entire E gene (1,485 nucleotides for DENV-1, -2 and -4 or 1,979 for DENV-3) were aligned based upon codons with sequences previously obtained in other studies and submitted to GenBank (Armstrong & Rico-Hesse, 2001, Foster et al., 2003, Goncalvez et al., 2002, Lanciotti et al., 1997, Lanciotti et al., 1994, Leitmeyer et al., 1999, Uzcategui et al., 2001, Uzcategui et al., 2003). Strains belonging to all major branches of previously published DENV phylogenies were included in the analyses and were selected to maximize the representation of genotypes, countries and years of isolation. Sequences that were very closely related to one another in the data set as well as those believed to represent recombinant events were removed from the analyses.

To obtain high-resolution phylogenies, separate Bayesian likelihood analyses for each serotype were performed using MrBayes 3.0 program (Ronquist & Huelsenbeck, 2003). Since this is a recently introduced phylogenetic method a brief description of it follows.

Likelihood approaches to phylogenetic reconstruction were proposed as an alternative to more traditional parsimony or distance-based methods (Felsenstein, 1981). However, the use of the maximum likelihood criteria in the analysis of large datasets is limited by the considerable time consumed in the search of the most likely tree and the even larger time used to compute the branch support by resampling processes like bootstrap. Bayesian likelihood phylogenetic inference method was introduced as a more computationally efficient alternative (Rannala & Yang, 1996). Bayesian likelihood method estimates posterior probabilities of different trees given a dataset and a previously selected model of nucleotide substitution. Instead of looking

for the most likely tree, Bayesian likelihood looks for a set of progressively more likely trees by a sampling method based on a Markov chain Monte Carlo (MCMC) algorithm. Given a large enough number of sampled trees, the frequency with which each branch appears in the subset of sampled trees is the posterior probability that it represents a true clade. A consensus tree is then computed out of the last sampled trees and posterior probabilities (or clade credibility values) are used as a measure of branch support. This makes performing bootstrap analysis unnecessary saving enormous amounts of time (Hall, 2001, Holder & Lewis, 2003).

In this study Bayesian likelihood analyses were performed independently for each DENV serotype with the following parameters and settings: general time reversibility (GTR) substitution model, gamma distribution of the substitution rates and α shape parameter estimated directly from the data. Four MCMC tree searches of 1,000,000 generations each were run simultaneously, sampling one in every 100 trees and computing a 50% majority-rule consensus out of the last 8,000 sampled trees. The consensus tree was drawn with the TreeView program.

To obtain a wider picture of phylogenetic relationships and genetic distances among and within serotypes in DENV complex, a single neigbor-joining tree was constructed with all the sequences used in this study using PAUP* 4.10 package (Swofford, 1998). Additional analyses performed in the sequences of each serotype included measures of nucleotide diversity and tests of neutrality. These latter analyses were performed with the DnaSP 4.0 software package (Rozas et al., 2003).

RESULTS

The alignments of the sequences included in this study had between 256 and 433 parsimony informative sites, which permitted robust phylogenetic analyses. These alignments can be seen in the Appendix at the end of this dissertation.

General data about the amount and distribution of variability in the four alignments, including parameters and statistics relevant to the phylogenetic analysis and useful in detecting natural selection, is presented in Table 3-3. All measurements of diversity were consistently higher for DENV-2 and lower for DENV-4. Fu and Li's F* and Tajima's D statistics were negative in all serotypes indicating strong purifying selection.

Table 3-3. Diversity indexes and tests for selection in the four sequence datasets.

Parameter or Statistic	DENV-1	DENV-2	DENV-3	DENV-4
# of sequences in the alignment	42	52	40	46
Length of the alignment (nt)	1485	1485	1479	1485
# of variable sites	441	564	367	354
# of parsimony informative sites	307	433	256	285
Average # of nt differences (K)	88.031	119.786	73.211	59.685
θ per gene	115.038	154.465	90.278	88.283
θ per site	0.077	0.104	0.061	0.059
Nucleotide diversity (π) total	0.059	0.081	0.050	0.040
π at synonymous sites	0.212	0.307	0.178	0.148
π at non-synonymous sites	0.010	0.013	0.010	0.007
ratio (π sy/π ns)	20.765	23.010	17.211	22.463
Fu & Li's F* statistic	-1.048	-0.840	-0.801	-0.400
Tajima' D statistic	-0.874	-0.816	-0.709	-1.190

DENV-1. The alignment for DENV-1 contains 42 sequences of 1485 nt each. The inferred phylogeny of serotype 1 is shown in Figure 3-3. There were five major branches that correspond to the genotypes previously described by Goncalvez et al (2002). The only known sylvatic strain of this serotype (Malaysia/72 P72) branched independently. Genotype V included three different branches, containing isolates from Asia, Africa and the Americas, respectively. The American isolates formed a monophyletic group that was resolved as a big polytomy in the consensus tree. Several isolates branched independently and four well-supported clades originated from this polytomy. Two of these clades were comprised of South American viruses and the other two corresponded to Mexican viruses from the 1980's and 1990's, respectively, suggesting there were two different introductions of DENV-1 into this country.

DENV-2. The phylogenetic tree for DENV-2 is shown in Figure 3-4. The six genotypes previously described by Twiddy et al (2002) are present and strongly supported. Most of the American strains are located either in the American or the American-Asian genotypes. However, the Mexican isolates segregated into four different genotypes. Strains isolated in 1980's and early 1990's are in the American genotype and are grouped in two separate clades depending on the decade. No isolate made after 1995 was in that group suggesting that the lineage went to extinction. One isolate from 1996 segregated with viruses in the "Cosmopolitan genotype" (Twiddy et al., 2002). This isolate, MX/Yucatan/96/BC17, is not closely related to any strain studied thus far. Two isolates from 1997 clustered with isolates in the Asian-2 genotype. The DENV-2 strain that caused the big epidemic in Cuba 1981 also clustered in the Asian-2 genotype, but the two 1997 Mexican isolates are more similar to the prototype DENV-2 New Guinea C/44 (NGC). Finally, all DENV-2 isolates from 2000 to 2002 grouped in the American-Asian genotype.

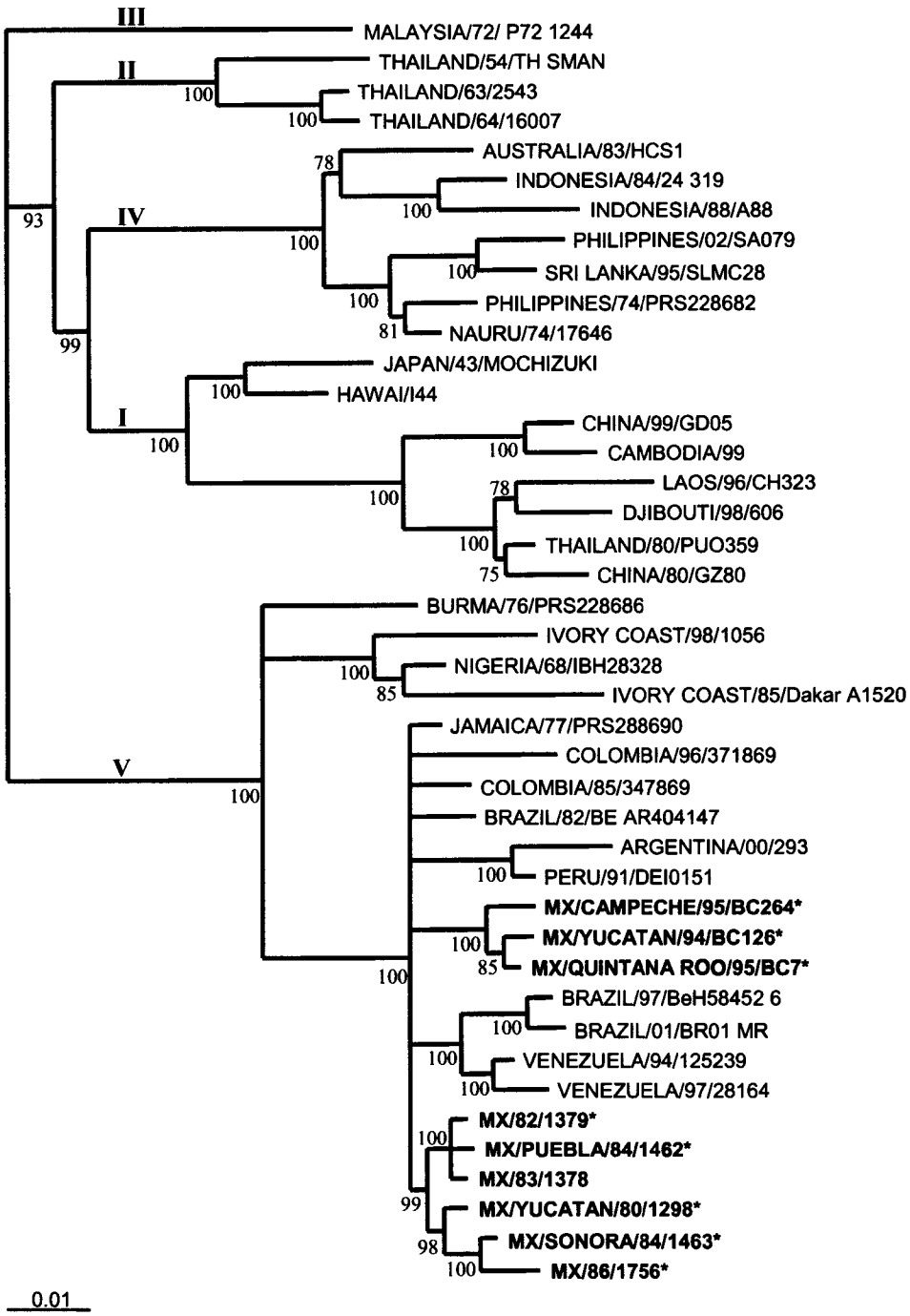


Figure 3-3. Phylogenetic tree for the E gene of DENV-1 estimated by the Bayesian likelihood method. Numbers close to nodes are posterior probabilities. Genotypes are designated in roman numerals according to Goncalvez *et al* (2003). The tree was rooted with the sylvatic Malaysia/72/P72 1244 strain. Sequences from Mexican viruses are shown in bold. An asterisk indicates that the sequence was obtained in this study. The analysis was performed with MrBayes 3.0

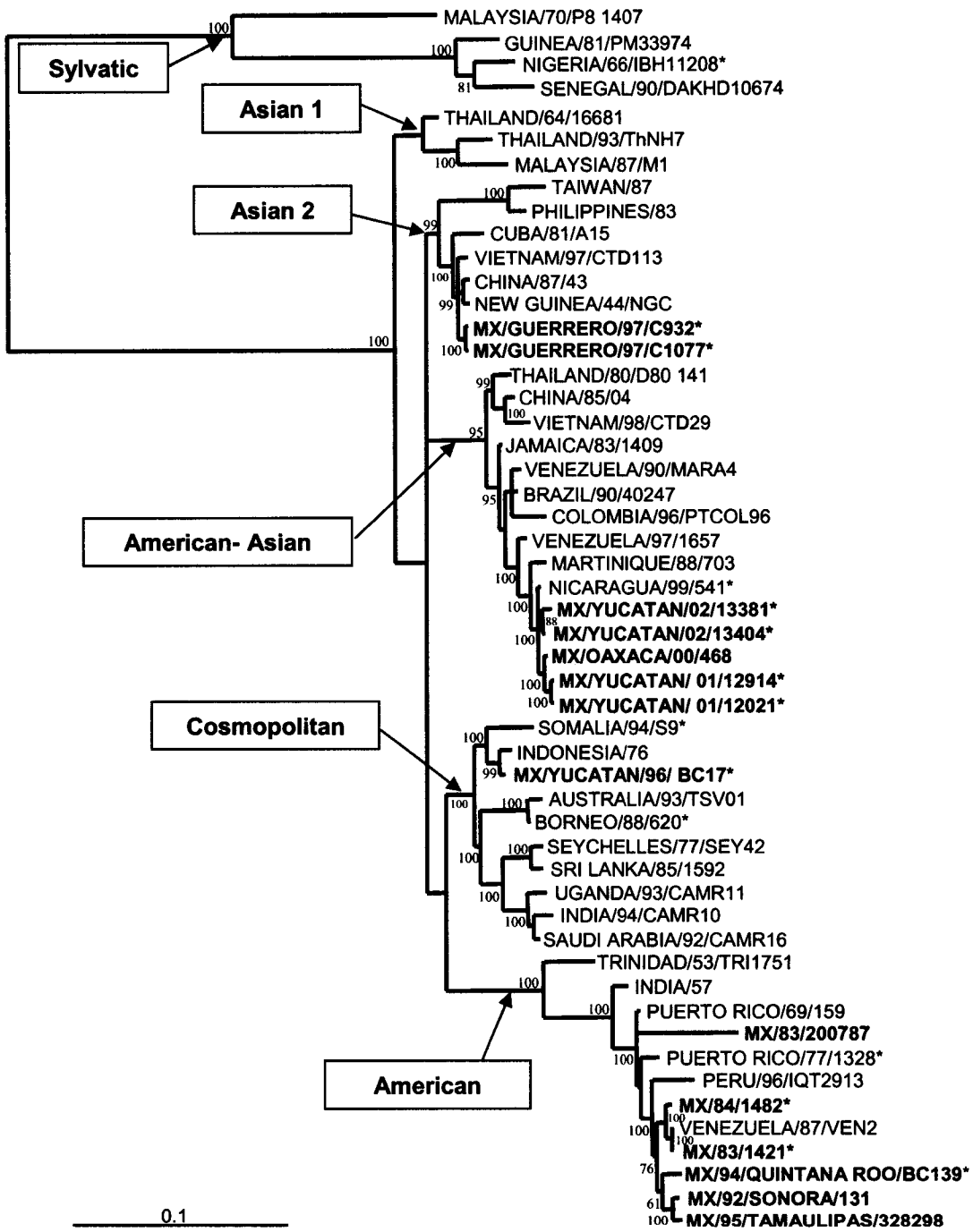


Figure 3-4. Phylogenetic tree for the E-gene of **DENV-2** estimated by the Bayesian likelihood approach. Posterior probabilities of relevant clades are shown close to the respective node. Genotypes are designated in boxed letters according to Twiddy et al (2002). The tree was rooted with the five isolates in the sylvatic genotype. Sequences from Mexican viruses are shown in bold. An asterisk indicates that the sequence was obtained in this study. The analysis was performed with MrBayes 3.0.

DENV-3. The phylogenetic tree of DENV-3 isolates is presented in Figure 3-5. The tree is composed of five major branches, corresponding to genotypes previously described (Lanciotti et al., 1994, Wittke et al., 2002). The American DENV-3 isolates subsequent to the reintroduction of DENV-3 in 1994 form a monophyletic group inside genotype III more closely associated with the only African strain (Somalia/93/S142) than with the other isolates from Asia and the Pacific. Most Mexican isolates from 1995 to 2000 form a monophyletic group but the MX/Yucatan/95/BC4 clusters with Panama/94, one of the two original isolates of the reintroduced strain (CDC, 1995, Guzman et al., 1996b). There are two additional independent clades, one containing viruses from Martinique and Brazil and the other from Venezuela. Overall, the topology suggests a single introduction of DENV-3 in Mexico but probably two different introductions into South America.

DENV-4. The result of the phylogenetic analysis of DENV-4 isolates is presented in Figure 3-6. In agreement with a previously published work (Wang et al., 2000), three major branches or genotypes were strongly supported. Three strains isolated from rural mosquitoes or monkeys in Malaysia formed the strongly divergent sylvatic genotype. Most strains from Asia were in the previously designated genotype I (Lanciotti et al., 1997). However, the isolate Indonesia/77/1132 appeared in a basal position of genotype II, suggesting that it is close to the ancestor of all the American DENV-4 isolates. The Tahiti/79/S44754 is also basal to the American isolates, which suggests that DENV-4 could have been introduced through the Pacific. Most of the other viruses in genotype II were isolated in the Americas, except New Caledonia/81/5489.

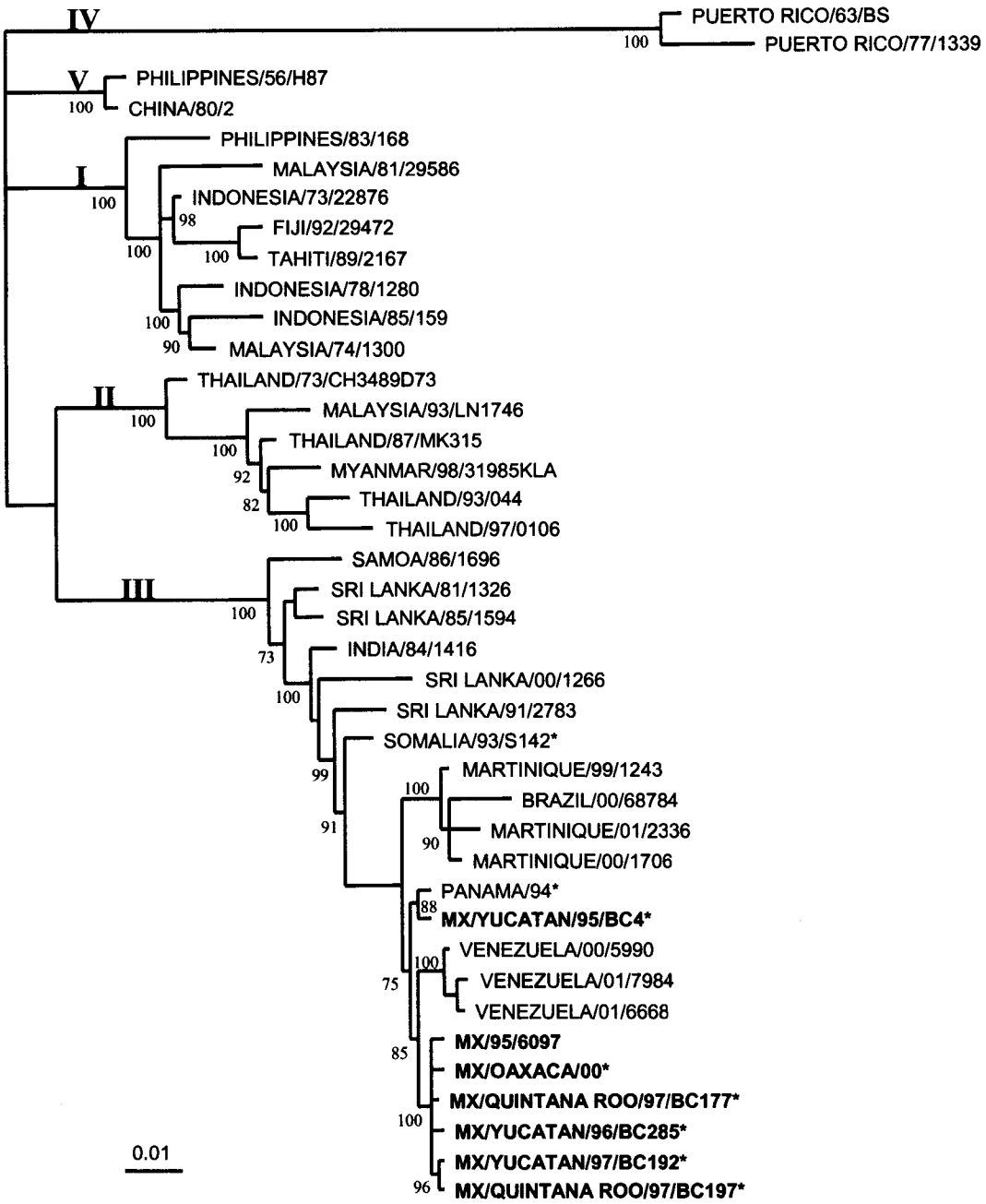


Figure 3-5. Phylogenetic tree for the E-gene of **DENV-3** estimated by the Bayesian likelihood method. Posterior probabilities of relevant clades are shown close to the respective node. Genotypes are designated in roman numerals according to Wittke et al (2002) . The tree was rooted with the two sequences from Puerto Rico (genotype IV). Sequences from Mexican viruses are shown in bold. An asterisk indicates that the sequence was obtained in this study. The analysis was performed with MrBayes 3.0.

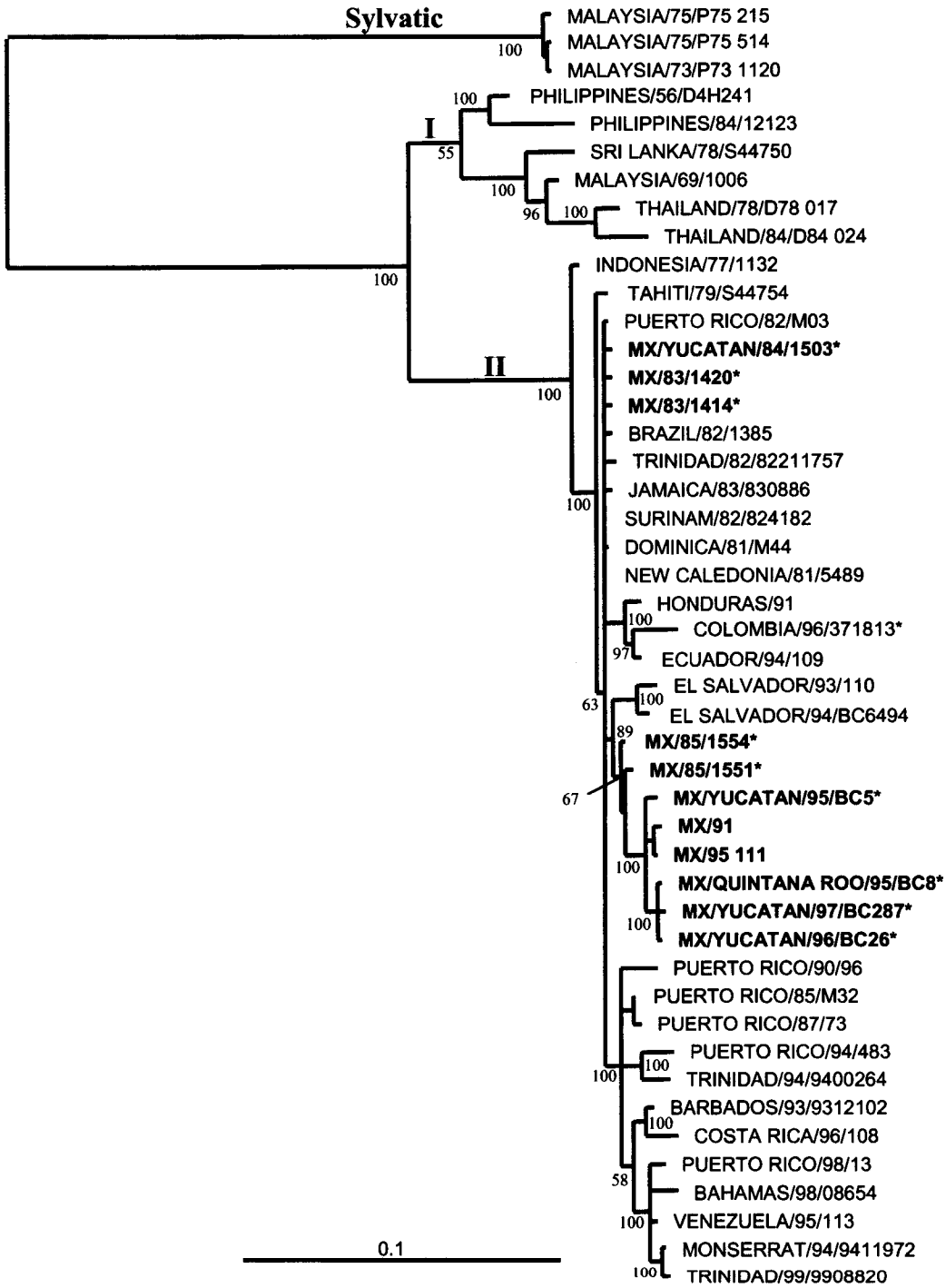


Figure 3-6. Phylogenetic tree for the E gene of **DENV-4** estimated by the Bayesian likelihood approach. Numbers close to nodes are posterior probabilities. Genotypes are designated in roman numerals according to Lanciotti et al (1997) except the sylvatic genotype that is labeled in letters. The tree was mid-point rooted. Sequences from Mexican viruses are shown in bold. An asterisk indicates that the sequence was obtained in this study. The analysis was performed with MrBayes 3.0.

All of the American DENV-4s isolated between 1981 and 1984, including three Mexican viruses, are very similar to the Dominica/81 strain, believed to represent the original introduction of this serotype in the Americas. These "early" American isolates correspond to the paraphyletic "group A" described by Foster et al (2003). Viruses isolated in 1985 or thereafter segregate into three different branches. Most strains isolated in the Caribbean islands, as well as one Venezuelan and one Costa Rican isolates, formed a clade that undergoes further differentiation. This group is the same "clade B" described by Foster et al (2003). Other clade is comprised of isolates from Honduras, Ecuador and Colombia. The remaining clade is included all the Mexican isolates of 1985 and thereafter as well as two isolates from El Salvador. Each country in this clade segregates as a monophyletic group.

Phylogenetic relationships among serotypes. Figure 3-7 shows the neighbor-joining tree of all four serotypes, including all the sequences analyzed in this study. All the previously identified genotypes are recognizable and they contain the same isolates found in the Bayesian analyses for independent serotypes (data not shown). The topology is similar to that obtained in other studies of the entire DENV complex (Wang et al, 2000). DENV-4 is the first to diverge. DENV-1 and -3 are the most closely related. Sylvatic isolates obtained from hosts of the enzootic cycle are the first to diverge in DENV-2 and DENV-4. The only sylvatic strain of DENV-1 forms a clade with genotype II and diverges later than other genotypes in this analysis, but it is still different enough to be considered an independent genotype. However, the branching order among genotypes is not the same. For example, the American genotype of DENV-2 is the last to diverge in the Bayesian tree (Figure 3-4) but it diverges early in the neighbor-joining analysis, second only to the sylvatic genotype.

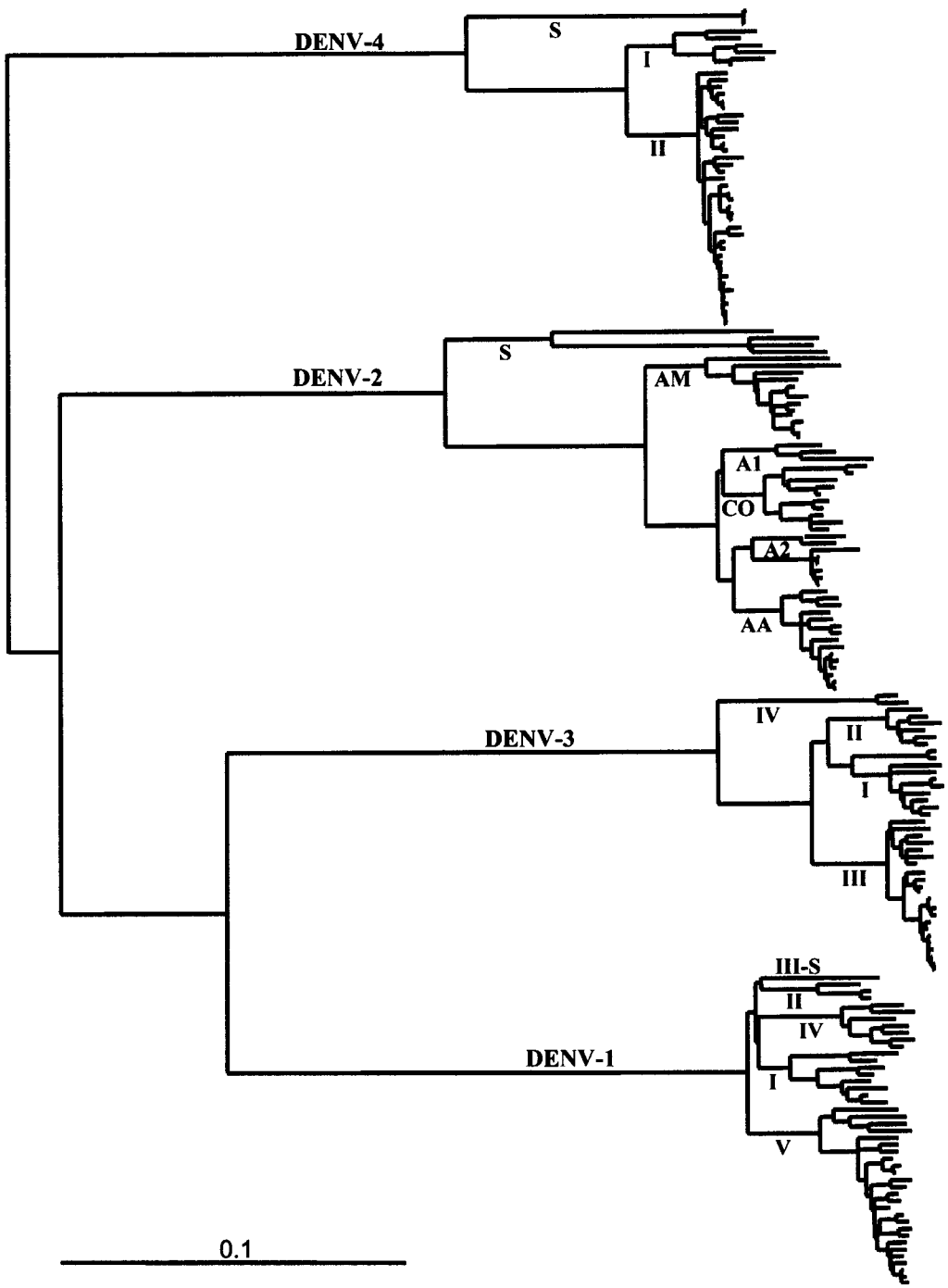


Figure 3-7. Neighbor-joining phylogenetic tree of the E-gene of the 180 sequences of DENV analyzed in this study. The bar in the left lower corner indicate 10% of divergence by the Tamura-Nei method. Genotypes are denoted in roman numbers according to Goncalvez et al (2003), (1993), Lanciotti et al (1994) and Foster et al (2003) for serotypes -1, -3, and -4 , respectively. Genotypes within DENV-2 are named according to Twiddy et al, (2002) as follow: S=sylvatic, AM=American, AA=American-Asian, CO=cosmopolitan, A1=Asian 1, A2= Asian 2.

DISCUSSION

Phylogenetic analysis is a powerful approach to determine the geographical and biological origin of emerging viruses as demonstrated in studies with Human Immunodeficiency virus (HIV), West Nile virus (WNV) and the agent of severe acute respiratory syndrome (SARS-CoV), (Lanciotti et al., 1999, Rambaut et al., 2001, Stavrinides & Guttman, 2004). The pioneering work by Rico-Hesse with DENV revealed the extent of variation and general distribution of lineages of DENV-1 and DENV-2 (Rico-Hesse 1990). It was shown later that analysis based on 240 nucleotides of the E/NS1 junction lacked the power to resolve ancestral relationships of certain strains with close geographic origin (Rico-Hesse et al., 1998). The growing availability of DENV isolates and new sequencing techniques facilitated phylogenetic analyses based on the full E gene, permitting more robust analyses. Thus, it has been possible to clarify some previously unresolved parental relationships (Twiddy et al., 2002), to describe new DENV lineages (Lewis et al., 1993, Wang et al., 2000), and to reveal *in situ* evolution and divergence in viruses isolated in close temporal and geographical proximity (Foster et al., 2003, Goncalvez et al., 2002, Uzcategui et al., 2001).

The phylogenetic analyses conducted in this study support single introductions of DENV-1 and -4 into the Americas in recent decades and no new information regarding their origins was revealed. The introduction of a new DENV-3 strain into Central America in 1994, different to the one circulating during 1960's and 1970's, has been previously documented. However, the prevailing idea of an Asian origin for this genotype III DENV-3 strain (Briseno-Garcia et al., 1996, CDC, 1995, Guzman et al., 1996b, Rico-Hesse, 2003) was not supported. In agreement with a recent work (Messer et al., 2003), this study suggests an African origin for that strain because of the monophyletic clustering of the New World isolates with the Somalia/93/S142 strain

(Figure 3-5). The latter was isolated from an American soldier during the military intervention in that country in 1993-1994. The support for that clade is 95%, but the genetic distance is too large to suggest that the Somalia/93/S142 isolate is a direct ancestor of the new American DENV-3 strain. It seems more likely that a related DENV-3 strain circulating in East Africa was transported to and colonized tropical America. Sequencing of more DENV-3 isolates from the early 1990's from Africa could help to resolve this issue.

The phylogeny of DENV-2 is different from the other serotypes and demonstrates multiple introductions of this serotype in recent years (Figure 3-4). The progressive replacement of the American DENV-2 genotype by viruses coming from the Eastern Hemisphere and the concomitant emergence of DHF has been previously documented (Rico-Hesse, 1990, Rico-Hesse et al., 1997). It is commonly thought that a single strain of DENV-2 was introduced into Jamaica in 1981 and caused the DHF outbreak in Cuba the same year. This so called "Southeast Asian genotype" then dispersed to other countries, first in the Caribbean, then in South America and finally in Central America and Mexico (Armstrong & Rico-Hesse, 2001, Gubler & Clark, 1995, Rico-Hesse, 2003, Vorndam et al., 1994). However, phylogenetic studies in Cuba using the short E/NS1 segment suggested that different strains were introduced into Cuba and Jamaica by the same time (Guzman et al., 1995, Sariol et al., 1999). These results have been largely ignored. This study included, for the first time, the full E gene of the 1981 Cuban strain (GenBank accession No. A91810), and it clustered in a different genotype than the Jamaican strain.

Rico-Hesse et al (1997), studying the short E/NS1 sequence observed that DENV-2 strains associated with DHF cases in the Americas segregated in two different

clades, but they later dismissed these results because of the lack of bootstrap support of this division when a large collection of viruses from Thailand was included in the analysis (Rico-Hesse et al., 1998). However analysis of the entire E gene of an even larger number of viruses from Southeast Asia revealed that they could be clearly separated into at least three well-supported lineages, Asian-1, Asian-2 and American/Asian (Twiddy et al., 2002), suggesting that the lack of support of previous studies was due to the presence of few informative characters in the region analyzed. These three lineages, which are the "Southeast Asian genotype" in the works by Rico-Hesse (2003), did not form a monophyletic group in the Bayesian tree (Figure 3-4) or in the maximum likelihood tree (Twiddy et al., 2002), but they did in a parsimony tree (Lewis et al., 1993).

Mexican DENV-2s segregated into four different genotypes and this grouping was supported by high branch support values. Most isolates appeared in the Americas and the American-Asian genotype, as expected. The isolates appearing in other genotypes deserve special consideration. The MX/Yucatan/96/BC17 virus was isolated in Yucatan state in 1996 from a DF case; it is the only known strain of the Cosmopolitan genotype reported in the Americas. Its origin is uncertain; the closest relative in the phylogenetic analysis was a virus from Indonesia isolated twenty years before. This strain was not isolated again in Mexico, demonstrating that new virus introductions do not always result in extensive epidemics. The viruses C-932 and C-1077 isolated in 1997 from samples collected in the southwestern state of Guerrero clustered in the Asian-2 genotype. They emerged independently in the tree and do not seem to be descendants of the Cuban strain of 1981. They could be derived from a strain found in Honduras in 1986 (Foster et al., 2004), in Venezuela 1994-1995 and in Mexico in 1995 (Rico-Hesse et al., 1997). Some of the 1995 Mexican samples described in the latter study came

from DHF patients in the southern state of Chiapas and the central state of San Luis Potosi (Rico-Hesse et al., 1997), suggesting that this strain circulated widely in Mexico and that it was responsible, at least in part, for the sudden increase in dengue severity in Mexico during 1995-1997 (Figure 3-1).

It is interesting that these DENV-2 Mexican viruses of 1997, as well as isolates from Vietnam, China and, in lesser extent, the Cuban isolate (Figure 3-4), are so close to the prototype DENV-2 strain, New Guinea C (NGC). In the E gene they differ from NGC by only five or six nucleotides. NGC was isolated more than fifty years before, which represents a substitution rate approximately 10-fold lower than what has been reported (Twiddy et al., 2003). It was recently suggested that these isolates could represent contamination of samples with NGC, which is used as a reference strain in many laboratories (Foster et al., 2004, Rico-Hesse, 2003). However, it seems unlikely that different laboratories in different countries experienced contamination with the same strain. An alternative explanation is that the NGC strain could have recently escaped from a laboratory and started to circulate in different countries. Another possibility is that a very similar strain is being maintained in an enzootic cycle where it is evolving very slowly; this strain would have been introduced sporadically to the endemic/epidemic cycle of humans and *Ae aegypti* in Asia and transported later to the Americas. Some viruses like Influenza A evolve more slowly in the natural reservoir than in human populations (Gorman et al., 1992). Analyses of more isolates obtained during a longer period of time are necessary to resolve this intriguing issue.

At a more regional level, DENV evolution appears dominated by frequent episodes of introductions, expansions, extinctions and reintroductions of the same or different viral genotypes. DENV-4 is an exception and evolved in a linear fashion in

Mexico since its introduction (Figure 3-6). DENV-1 transmission seems to have ceased in the early 1990's (Briseno-Garcia et al., 1996). In 1994 a new strain of the same genotype appeared (Figure 3-3). Similarly the American genotype of DENV-2 was prevalent in the early 1980's, disappeared for several years, and reappeared in 1992 (Briseno-Garcia et al., 1996, Navarrete, 2002). These two DENV-2 strains clusters in slightly separated clades (Figure 3-4). This American genotype has not been found in Mexican isolates since 1995 suggesting that it went to extinction as happened earlier in the Caribbean and most of South America (Rico-Hesse 1997 and 2003). The Cosmopolitan and Asian-2 genotypes strains, isolated in Mexico in the mid 1990's, also seem to have disappeared by year 2000. The American-Asian genotype seems to have displaced all the others DENV-2 genotypes in the Americas (Armstrong & Rico-Hesse, 2001, Uzcategui et al., 2001).

DENV-3 was introduced into Central America in 1994 and into Mexico in 1995. It appears to have remained there with few genetic changes until 2000. In this work, sequences from three previous studies were combined into a single phylogenetic tree (Miagostovich et al., 2002, Peyrefitte et al., 2003, Uzcategui et al., 2003). The topology suggests that the arrival of DENV-3 into South America occurred from at least two distinguishable introductions of the Central American strains. One strain was introduced into Brazil in 2000 seemingly through the Antilles (Figure 3-5) and the other appeared in Venezuela also in 2000, but the route of introduction cannot be deduced from the data. Phylogenetic analysis including isolates from other countries during the 1998-2000 interval could resolve this issue.

Previous studies have correlated severe dengue outcomes with infection of certain DENV genotypes or strains (Rico-Hesse et al., 1997). Since no virus in the

Mexican collection was isolated from DHF cases, this study cannot directly test this issue. However, this work shows that the two periods with the highest incidence of DHF in Mexico (Figure 3-2) were preceded by the introduction of new serotypes or genotypes. The critical period of 1995-1997 (Briseno-Garcia et al., 1996, Navarrete, 2002) coincided with the arrival of DENV-3 genotype III and DENV-2 Asian-2 and Cosmopolitan genotypes (Figure 3-4). The new wave of DHF during 2001-2002 (Navarrete, 2002) seems to be related to the introduction of DENV-2 American-Asian genotype (Lorono-Pino et al., submitted and this work). This study provides indirect support to the hypothesis that the introduction of more virulent genotypes has contributed to the increased severity of DEN in the last years in the Americas.

Thus, virus trafficking seems to have played a major role in both the incidence and the severity of the clinical presentation of DEN in Mexico and other countries, although it is clearly not the only factor involved. It is not known, for example, why the introduction of the DENV-2 American-Asian genotype into Jamaica was not accompanied by an outbreak of DHF in that country as happened later in South and Central America and Mexico (Gubler & Clark, 1995, Navarrete, 2002). Differences in the immune status and genetic background of the respective populations must be interacting with the virulence of the agent to determine the clinical outcome and the epidemiological behavior of DEN.

Sequence data collected in this study can also be used to gain insight about the forces acting in DENV evolution. The ratio of diversity at synonymous and non-synonymous sites in the aligned sequences was between 17.2 and 23.0, which indicates that mutations leading to amino acid susbstitutions were strongly constrained. This is in agreement with the results in the Fu and Li's test of neutrality in which a

negative value was obtained for all serotypes (Table 3-3). Synonymous substitutions accounting for most of the variation observed are presumed to be neutral and more likely fixed by random genetic drift.

Nucleotide diversity was higher in sequences of DENV-2, intermediate in DENV-1 and -3 and lower in DENV-4. These could be partially due to a sampling issue. In this study more isolates were included from the Americas than from other continents. Since American strains segregated in four genotypes of DENV-2 but only in two, one and one genotypes of DENV-3, -1 and -4, respectively, it was expected that the average nucleotide diversity would be higher for DENV-2 (Table 3-3). Some serotypes have circulated in the Americas for longer periods than others: they have been present at least since 1954 for DENV-2, 1966 for DENV-3, 1977 for DENV-1 and 1981 for DENV-4. Persistence for long times allows a larger number of substitutions to be accumulated. This factor also accounts for the greater level of differentiation in DENV-2 and the lower in DENV-4. Thus, differences in virus traffiking into the Americas and differences in amounts of time for virus evolution probably account for the different levels of variation observed among serotypes.

In summary, the results of this study indicate that, in the conditions prevailing in nature, strong purifying selection, random genetic drift and gene flow are the predominant forces acting in DENV evolution in the Americas. The last factor, represented by multiple introductions of foreign strains, has played a major role shaping the structure of the virus populations and, very likely, determining the changes in epidemiology of DF and DHF in this continent.

CHAPTER 4

A SURVEY FOR RECOMBINANTS AMONG DENGUE VIRUS ISOLATES FROM MEXICO

RNA viruses can rapidly adapt to the changing environment of their transmission cycles. Such adaptability depends on the rapid generation of a wide array of genetic variants from which the most fitt genotypes can be selected in the new adaptive landscape. This genetic variability is provided by three mechanisms: mutation, recombination and reassortment. Mutation is the fundamental force in RNA virus evolution with frequencies of about one substitution per genome per generation. This results in substitution rates of about 10^{-3} substitutions per site per year in most RNA viral populations, a rate about one-million fold higher than that observed in eukaryote organisms (Drake & Holland, 1999, Holland et al., 1982, Holmes, 2003, Lai, 1992). Reassortment refers to the exchange of RNA segments between viruses with a multipartite genome. Recombination is defined as the formation of chimeric nucleic-acid molecules from sequences present in different genomes or previously separated in the same genome (Lai, 1992).

Recombination is a widespread phenomenon in many organisms. Genomes of sexually reproducing organisms regularly undergo recombination previous to the first meiotic division and exchange genetic information between the paternal and maternal

copies of each chromosome pair. Prokaryotes horizontally exchange DNA sequences during conjugation and transduction, which eventually become integrated in their chromosomes by recombination (Posada et al., 2002).

Recombination can be classified as homologous or non-homologous. Non-homologous recombination occurs when unrelated organisms exchange genomic material or when related individuals exchange sequences at non-homologous regions. In viruses, non-homologous recombination can occur between two unrelated viruses infecting the same cell, between the viral and host cell genomes or even between different RNA segments of the same virus (Lai, 1992, Nagy & Simon, 1997). Non-homologous recombination is a way of acquiring new genes or exchanging functional domains among proteins, and is a fundamental mechanism in the formation of new species. Homologous recombination occurs when the same or closely related species exchange nucleic-acid sequences at homologous sites in the genome. For the remainder of this chapter the focus will be on homologous recombination.

The high substitution rate exhibited by most RNA viruses generates a theoretical problem: most of the progeny of any RNA virus will have at least one mutation that in most cases will be deleterious. Mathematical models predict that, after a few generations, mutations that decrease fitness will accumulate and the population will progressively decline (Muller, 1964). This irreversible accumulation of deleterious mutations (or genetic load) has been referred as the 'Muller's ratchet. Purifying selection is the principal mechanism to purge the genetic load but some authors have suggested that, in the absence of recombination and reassortment, natural selection would be insufficient to maintain the integrity of viable genomes (Chao, 1990, Lai, 1992, Muller, 1964, Worobey & Holmes, 1999).

Recombination is also a way of producing genetic diversity among individuals, which is the raw material for natural selection and evolution. Adaptive mutations appearing in separate individuals can be incorporated into the same genome by recombination. New combinations of favorable mutations take a long time to occur by mutation only (Hartl, 1997). Recombination can also preclude the fixation of deleterious mutations by hitchhiking on closely linked favorable alleles (Felsenstein & Yokoyama, 1976, Hey, 1998). In summary, recombination play two opposite roles: exploration of new combinations of genomic regions and rescuing of viable genomes from debilitated parental genomes (Domingo & Holland, 1997, Hartl, 1997, Hey, 1998, Lai, 1992).

Recombination also has important consequences for phylogenetic analysis. Most phylogenetic methods assume that new species or strains appear by the splitting of populations in two or more branches, which is properly represented as a tree. But recombination violates that assumption. The merging of two different branches generating new variants or new species is a reticulate process that would be more appropriately represented as a network. When recombination has happened among the taxa in a phylogenetic analysis the topology of the tree could change, the length of terminal branches increases and the time to the most recent common ancestor is reduced. This resembles the phylogeny of an exponentially growing population. It also generates the loss of the molecular clock and overestimates of the substitution rate heterogeneity (Posada & Crandall, 2002, Schierup & Hein, 2000a, Schierup & Hein, 2000b).

Recombination between two RNA viruses was first demonstrated experimentally with viruses of the *Picornaviridae* family. Co-infection of cells with conditional lethal

mutants of either poliovirus (Hirst, 1962) or foot-and-mouth disease virus (Pringle, 1965) yielded viable recombinants. Investigations with polioviruses provided the first evidence that the usual mechanism of RNA recombination was the jumping of the RNA polymerase from one template to another, a process usually referred to as "copy-choice" or "template switching" (Alejska et al., 2001, Kirkegaard & Baltimore, 1986, Nagy & Simon, 1997, Romanova et al., 1986). The mechanism of recombination in RNA viruses will be discussed in detail in the next chapter.

Polioviruses were also the first RNA viruses observed as recombinants in natural populations. This resulted from the generalized use of vaccination against poliomyelitis with the oral, live-attenuated, polyvalent vaccine (OPV); interserotypic recombinants were found to be frequently shed in the feces of vaccinated persons. Some recombinants were isolated from vaccine-associated paralytic cases (Agol et al., 1985, Georgescu et al., 1994), suggesting that recombination could enable the virus to regain virulence. Soon after, it was realized that recombinants between vaccine and wild-type strains were circulating freely in nature (Rico-Hesse et al., 1987). In more recent years, several outbreaks of paralytic poliomyelitis have been associated with circulation of viruses that are mosaics between vaccine and wild-type poliovirus strains or between vaccine poliovirus and unrelated enteroviruses (Kew et al., 2002, Liu et al., 2000).

Recombination has been more frequently reported in retroviruses (Lole et al., 1999, Robertson et al., 1995, Salminen et al., 1995) and coronaviruses (Lai, 1992, Stavrinides & Guttman, 2004). Retroviruses are a special case, because they have several mechanisms available to recombine (Negroni & Buc, 2001). There is now a large list of RNA viruses in which recombination has been demonstrated either in natural or experimental environments (Table 4-1). Perhaps the most famous case is

Western equine encephalitis virus, which resulted from recombination between Eastern equine encephalitis virus and a Sindbis-related virus, two arboviruses of the genus *Alphavirus* (Hahn et al., 1988, Weaver et al., 1997). Recombination has been observed mostly in positive-sense RNA viruses (Chare et al., 2003).

Demonstration of recombination between DENV is more recent. In 1999 two papers showed that the nucleotide sequence of several previously characterized DENV isolates could be better explained as the result of recombinant events between different strains within the same serotype. Evidence for recombination was found in seven out of 71 DENV sequences analyzed, and at least one recombinant was recognized in each DENV serotype (Holmes et al., 1999, Worobey et al., 1999). These authors did not find recombination between different serotypes. Soon thereafter, a few more cases of intraserotypic recombination were reported (Tolou et al., 2001, Twiddy & Holmes, 2003, Uzcategui et al., 2001), which suggested that this phenomenon was not rare among DENV populations.

However, the extent of recombination in DENV remains unknown. One reason is that methods to detect recombination are in early stages of development, and they miss most recombination events occurring between closely related genomes (Hudson & Kaplan, 1985). All cases reported so far exhibit crossovers in the structural part of the DENV genome. This is probably due to underrepresentation of the non-structural genes in the sequence databases. For similar reasons the majority of recombinant events detected were in viruses isolated in Southeast Asia and the Caribbean (Worobey et al., 1999). This study was performed to test the hypothesis that recombination is widespread in natural populations of DENV and in particular in populations circulating in Mexico in the last two decades.

Table 4-1. Cases of recombination in natural populations of RNA viruses of animals.

Family	Virus	Comment	Reference
Picornaviridae	Poliovirus	Interserotypic recombination in healthy vaccinees	Bloomqvist S 2003, J Gen Virol 84:573
		Interserotypic in vaccine-associated paralytic cases	Georgescu MM, 1994 J Virology 68:8089. Martin J, 2002 J Virol 76:10921
		Vaccine & wild strains	Rico-Hesse 87, Virology 160:311; Liu HM J Virol2000 74:11153
		Vaccine strain & other enteroviruses	CDC 2001, MMWR 50:874
	Echovirus	Intertypic	Oprisan G, 2002 J Gen Virol 83:2193
	Coxsackie A virus	Intertypic	Brown B, 2003, J Virol 77:8973
	Coxsackie B virus	Intertypic	Lukashev AN 2003, J Virol 77:10423
	Hepatitis A virus	Intertypic	Costa-Mattioli M 2003, Virology 20:51
	Foot & mouth disease virus	Intertypic	Tosh C, 2002, J Gen Virol 83:2455-60
Retroviridae	Human Inmunodeficiency virus	In individual AIDS patient	Liu SL 2002, J Virol 76:10674. Many others
		In populations, Intersubtypic	Robertson DL 1995, J Mol Evol 40:249; many others
	Simian Inmunodeficiency virus	Intertypic	Bailes E 2003, Science 300:1713, many others
	Avian leukemia virus	Intertypic	Lupiani B 2000, Virology 276:37
Coronaviridae	Mouse hepatitis virus	non-homologous with *Influenza C virus*	Luytjes W 1988, Virology 166:415
	Infectious bronchitis virus		Wang L 1993, Virology 192:710
	SARS-CoV	Avian & mammalian viruses	Stravrides J 2004, J Virol 78:76
	Torovirus		Smith SL 2003, J Virol 77:9567
Astroviridae	Human astroviruses	Interserotypic	Walter JE 2001, Arch Virol 134:2357
Arenaviridae	Whitewater arroyo virus	Intertypic	Charrel RN 2002, Bioch Bioph Res Com 296:1118
Togaviridae	Western equine encephalitis virus	Product of EEEV and Sindbis	Hahn CS 1988, PNAS 85:5997
	Rubella virus	Intersubtypic	Zheng DP 2003, Emerg Infec Dis 9:1523
Flaviviridae	Bovine viral diarrhea virus	non-homologous with cellular sequences	Nagai M 2003, J Gen Virol 84:447
	Dengue virus	Intraserotypic	Worobey M 1999, PNAS 96:7352, others
	San Louis encephalitis virus.	Intersubtypic	Twiddy S 2003. J Gen Virol 84:429
	Japanese encephalitis virus	Intersubtypic	Twiddy S 2003. J Gen Virol 84:429
	Hepatitis C Virus	Intersubtypic	Colina R 2004, J Gen Virol 85:31
	Hepatitis G Virus	Intersubtypic	Worobey M 2001, Mol Biol Evol 18:254

MATERIALS AND METHODS

Viruses. Viral sequences analyzed for recombination were, for the most part, those Mexican isolates sequenced for the phylogenetic study in the previous chapter, plus other Mexican viruses for which the complete sequence of the E gene, or a larger sequence, has been determined (Armstrong & Rico-Hesse, 2003, Foster et al., 2003, Goncalvez et al., 2002, Leitmeyer et al., 1999, Ruiz et al., 2000). In order to maximize the probability of finding recombinant breakpoints, the longest sequence available for each serotype was used. For DENV-1 and -4 this sequence was the entire E gene (1,485 nucleotides), for DENV-3 the prM and E genes (1,977 nucleotides), and for most DENV-2 isolates a 2,426 nucleotides sequence (positions 162 to 2,578) encompassing most of C, entire prM and E and 5' end of NS1 genes. Since DENV-2 strains exhibit the largest diversity, they were studied in more detail. Four Mexican isolates of this serotype previously demonstrated to belong to different genotypes (Chapter 3) were selected for near-full genome sequencing. Representatives of all genotypes in each serotype were also analyzed as potential parental sequences.

Molecular Procedures. Viral propagation, RNA extraction, cDNA synthesis, PCR amplification and cycle sequencing were performed as described in the previous chapter. For the near-full genome sequencing of DENV-2 strains, five long overlapping PCR products (Figure 4-1) were amplified from appropriate cDNAs using Platinum Taq High Fidelity (Invitrogen, Carlsbad CA). Sequencing was mostly unidirectional but reverse sequencing was also done for some parts of the genome (Figure 4-1). Overlapping sequences were assembled in single contigs using the Seqman module of DNAstar package (Lasergene, Madison WI). Primers used in cDNA synthesis, PCR and cycle sequencing are listed in Table 4-2.

Table 4-2. Primers used in the full-genome sequencing of DENV-2 strains.

No.	Name	Sequence 5' → 3'	Uses*
1	F-D25'/JAM	AGTTGTTAGTCTACGTGGACCGACAAAGA	PCR, Seq
2	D1	TCAATATGCTGAAACGCGCGAGAAACCG	PCR
3	D2-367	ATGCTGAACATCTTGAACAGG	Seq
4	D2-616R	TTGCACCAACAGTCAATGTCTTCAGGTTC	PCR Seq
5	D2-739	ATGGGATTGGATACACGAACTGAA	Seq
6	D2-850R	TATGGTGTATGCCAGGATTGC	Seq
7	D2-1184	ATGAAGAGCAGGACAAAAGGTT	Seq
8	D2-1225R	CCATTTCCCCATCCTCTGTCTAC	Seq
9	D2-1540	GAAGACAAAGCTTGGCTGGTG	Seq
10	D2-1648R	GCATGGGGATTTTTGAARGTGAC	Seq
11	D2-2050	CCATTCGGAGACAGCTACATCAT	Seq
12	D2-2120R	TCAAACATTTGGCCGATRGAACTTC	Seq
13	D2-2412	GGTGCAGGCYGATAGTGGTTG	Seq
14	D2-2588R	TCTTGTTACTGAGCGGATTC	Seq
15	D2-2968	CTCATGTCAGCGGCCATAAA	Seq
16	D2-3010R	GCRCTTTCTATCCAATAACCCATCTC	Seq
17	D2-3385	AGATATAGAGGTGAGGATGGATGC	Seq
18	D2-3512R	GAACAGTGCCATTCCCAAGA	PCR, Seq
19	D2-3541	ATGCTCAGGACYCGAGTAGGAACGAA	Seq
20	D2-4295	GCTGCCGATGRAAAATGGGAAG	Seq
21	D2-4420R	TCACCAGCARTCCTGTTCTAATGAGTAT	Seq
22	D2-4756	TATGGATTAGGCTGGAAGCTAGAA	Seq
23	DV1-4918	GGRACKTCAGGWTCTCC	PCR
24	D2-5219	AAGCTCTTAGAGGACTTCCGATAA	Seq
25	DV3-5368R	AARTGIGCYTCRTCCAT	RT, PCR, Seq
26	D2-5720	ATGATTGGGACTTTGTGGTCACAAC	Seq
27	D2-6175	GCTGAAGGCATCAACTACGCAGAC	Seq
28	D2-6259R	CCCCTTCTTTTGTCCAGATTTC	Seq
29	D2-6597	CGGAAAAGGTATAGGGAAGATGAC	Seq
30	D2-7047	GCTAATGGGTCTTGGGAAAGGATG	Seq
31	D2-7493	ACATCTTTAGAGGGAGCTACTTGG	Seq
32	D2-7880	CAAAAGGAGGACCAGGACACGAAGA	PCR
33	D2-8056R	AATTTTCCACTAAGTTGAGGACTCT	PCR, Seq
34	D2-7948	CAGAGCGGAGTTGACGTTTT	Seq
35	D2-8450	ATGACCAAGACCACCCATACAAAA	Seq
36	D2-8868	GCTGGTTGACAGGGAAAGAAATCT	Seq
37	D2-9318	AACACCAAGAGGCACAGTAATGGA	Seq
38	D2-9680	CACAAGTGCCTTTCTGTTCACACC	Seq
39	D2-10146	GGCAAAGAACATCCAAACAGC	Seq
40	R-D23'/JAM	AGAACCTGTTGATTCAACAGCACCATTCCA	RT, PCR, Seq

* RT: reverse transcription, PCR: polymerase chain reaction, Seq: sequencing.

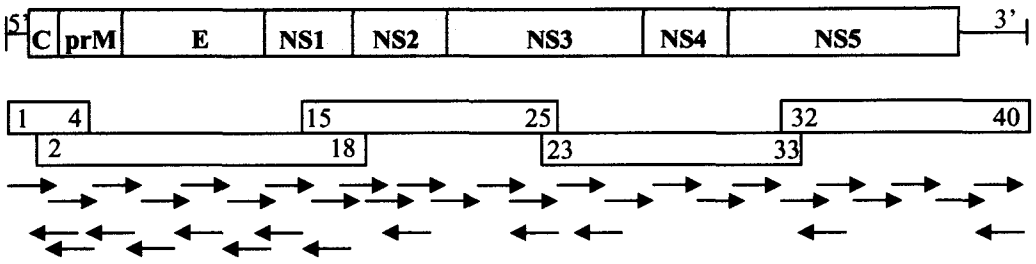


Figure 4-1. Strategy used in the full genome sequencing of selected DENV-2 isolates. Green bar represents DENV genome. Yellow bars represent PCR segments used as templates. Numbers inside bars are the primer used to amplify that segment according to their order in Table 4-2. Arrows indicate direction, approximate length and approximate position of the sequencing reactions performed with the primers listed in Table 4-2.

Recombination Analyses. Sequences representative of each genotype were selected to be tested as potential parent of every Mexican isolate. Sequences were aligned using the Clustal W algorithm (Higgins et al., 1996). Alignments were first screened for potential recombination events using "similarity plots" generated with Stuart Ray's Simplot computer package (Lole et al., 1999). A similarity plot is a graphical representation of the percentage homology among sequences in an alignment, computed in a sliding window. In this study the sliding window used was 200 nucleotides long and was moved in steps of ten nucleotides. Inversions of the similarity patterns were regarded as suggestive of recombination. Some sequences were selected for a more in-depth analysis using bootscanning (Salminen et al., 1995), a method also implemented in Simplot. Bootscanning also uses a sliding window approach, but it performs a phylogenetic analysis with bootstrap support in every step of the run and plots the percentage of permuted trees (or bootstrap value) supporting the clustering of the query sequence with each other sequence included in the data set

as a potential parent. Figure 4-2 depicts the bootscanning procedure in more detail. The program was set to perform neighbor-joining trees with 250 bootstrap replicates for each step. The window was 200 nucleotides wide and was moved in steps of ten. All the genetic distances were estimated using the Kimura's 2-parameter model of nucleotide substitution with a transition/transversion ratio estimated directly from the data. Bootstrap supports higher than 90% in two or more sequences were considered as suggestive of recombination.

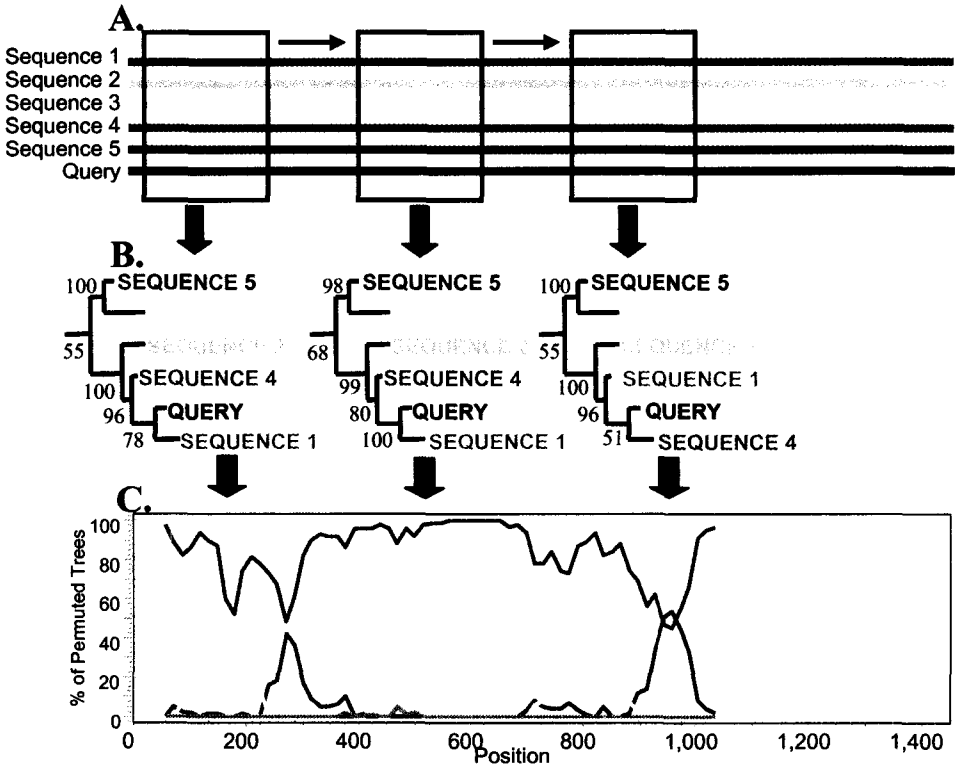
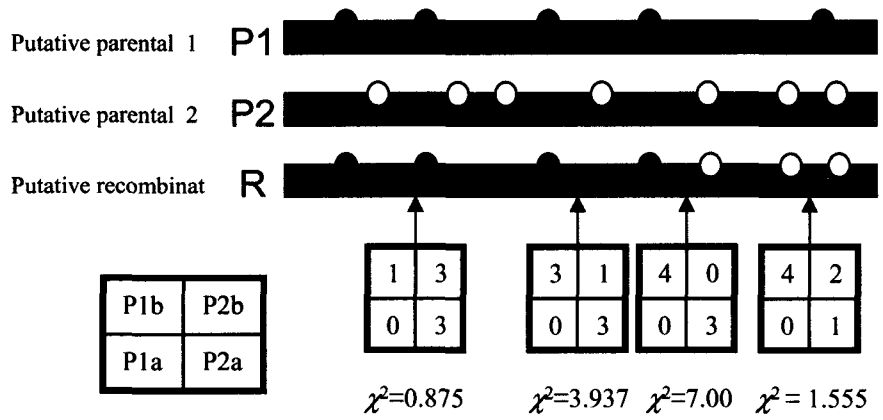


Figure 4-2. Schematic representation of the bootscannig test in a hypothetical alignment. **A.** Sliding window scanning the alignment. **B.** Phylogenetic analysis of each window. **C.** plotting of the bootstrap values. Each color represent the same sequence in all the drawings.

To test the statistical significance of potential recombinant events suggested in the previous methods, the maximum match chi-square test or "Chimaera" (Posada and Crandall, 2001) was used. Chimaera is implemented in the Recombination Detection Program (RDP) available at http://darwin.uvigo.es/rdp/rdp.htlm (Martin & Rybicki, 2000). Chimaera analyzes combinations of three sequences at a time. For each triplet, every sequence is treated as a potential recombinant and the others as potential parents. A hypothetical crossover is moved along every position of the potential recombinant sequence and the number of variable sites matching with either potential parent, before and after the crossover, are counted and analyzed in a 2x2 contingency table. The statistic is the chi-square value of the table. Figure 4-3 shows this procedure in detail. The significance (p-value) of the statistic is calculated by a permutation test. This p-value can be plotted along the sequence to locate the optimal breakpoints.



P1a = # of segregating sites with the same state in P1 and R after the breakpoint
P2a = # of segregating sites with the same state in P2 and R after the breakpoint
P1b = # of segregating sites with the same state in P1 and R before the breakpoint
P1a = # of segregating sites with the same state in P2 and R before the breakpoint

Figure 4-3. Schematic representation of the maximum match chi-square (χ^2) test illustrating how the χ^2 statistic is calculated in a hypothetical example. Black and white dots represent segregating sites in the alignment of 3 sequences. Modified from Posada and Crandall (2001, supporting information).

As a final confirmatory test for recombination, separate maximum likelihood phylogenetic analyses were conducted for each sequence fragment delimited with the previous analyses. The clustering of the putative recombinant sequence was compared in the respective trees and its significance was assessed by 1,000 bootstrap replicates.

RESULTS

Since most recombination detecting methods are relatively new and few people are familiar with them, a test of their performance was set up using a well-known recombinant sequence: Western Equine Encephalitis virus (WEEV). To do this, complete genome sequences of WEEV and four other alphaviruses, including the parental Eastern Equine Encephalitis virus (EEEV) and Sindbis virus, were aligned and analyzed by similarity plots, bootscanning and Chimaera. The results are shown in Figure 4-4. In the similarity plot and in the bootscan the closest relative for most of WEEV sequence is EEEV, but for the last two genes, E2 and E1, there is an inversion of the patterns and Sindbis becomes closer than EEEV. The p-value of the chi-square statistic in Chimaera reaches a very significant value in the same region where the others methods suggest a breakpoint. Other peaks are also observed but they are not so significant. All the methods supported the recombination event originally reported by Hahn et al (1988).

Since for most DENV serotypes no more than one genotype has circulated simultaneously in the Americas, recombination between different strains within the same genotype was first investigated. Sequences of isolates differing between 1-4% at the nucleotide level were examined in the programs as potential parents of the Mexican viruses. Examples of the results obtained are shown in Figure 4-5.

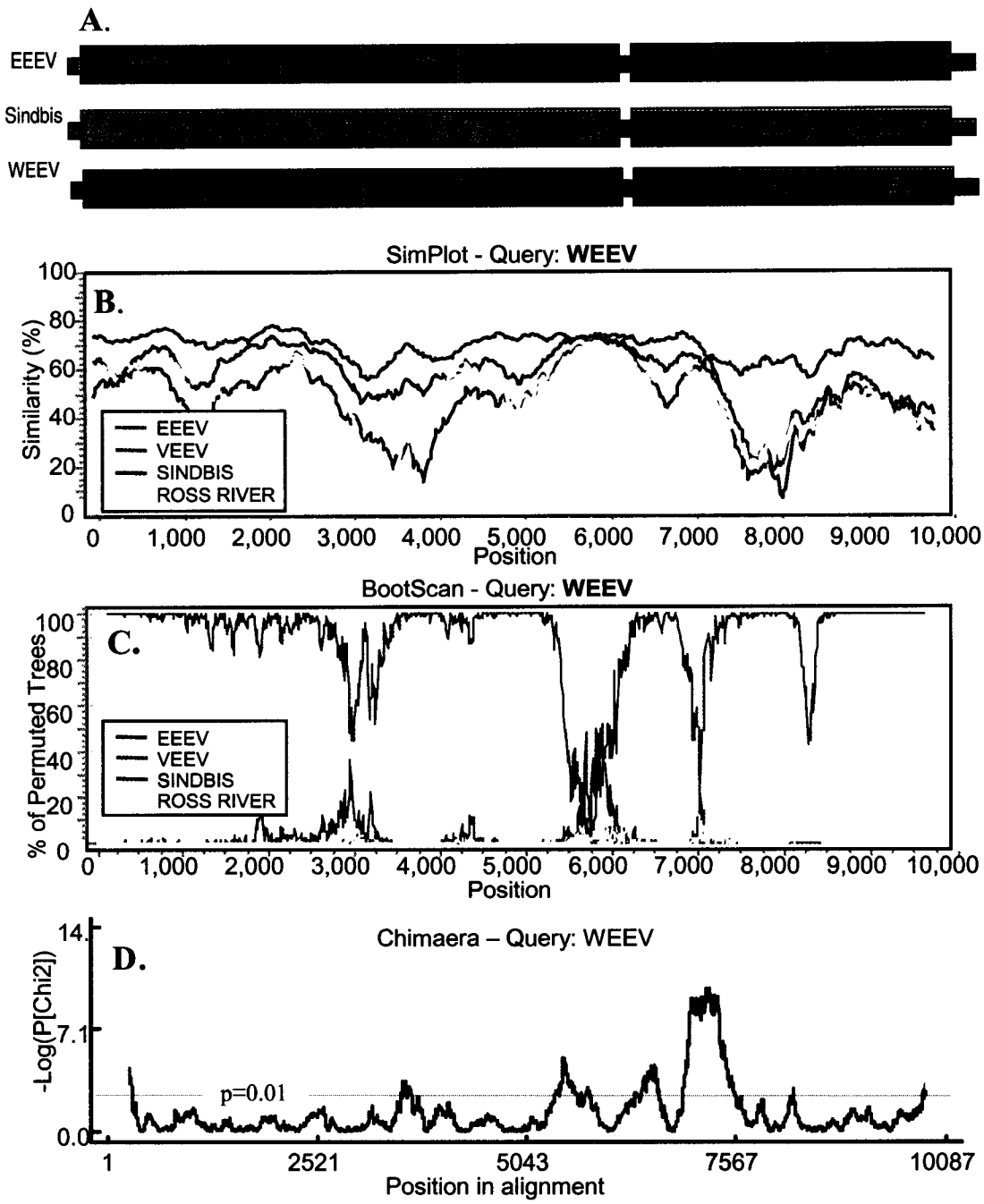


Figure 4-4. Demonstration of the performance of methods used to detect recombination in the entire sequence of several alphaviruses. **A.** Schematic representation of the sequence of Western Equine Encephalitis Virus (WEEV) and its parents according to Hahn et al., (1988). **A. B. C.** Graphic outputs of the recombinational analysis of alphaviruses using similarity plot (**B.**), bootscanning (**C.**) and maximum match chi-square test (Chimaera, **D.**). B. and C. were produced in Simplot (Lole et al, 1999). D. was generated with the Chimaera module of RDP (Martin and Rybicki, 2000). The dashed line indicate significance at the 0.01 level of confidence.

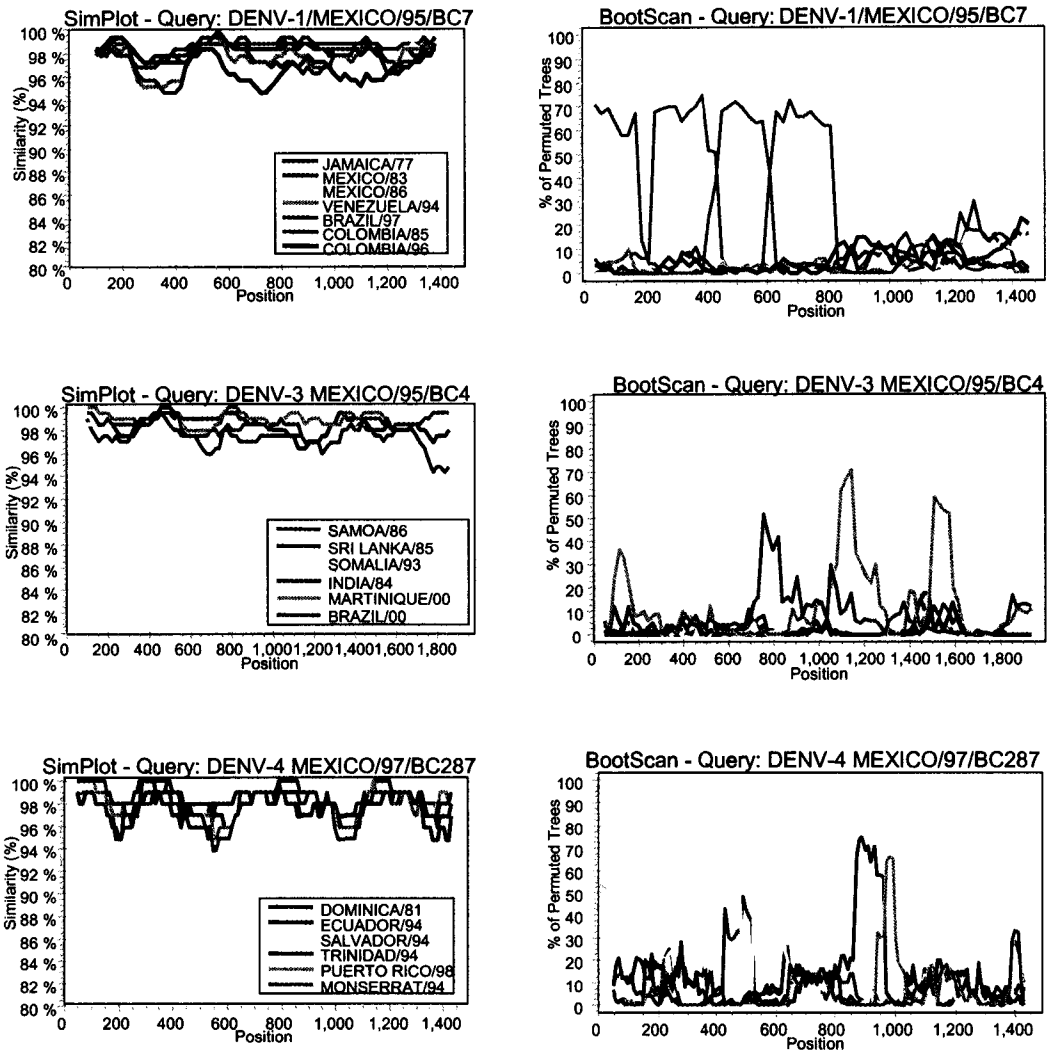


Figure 4-5: Examples of similarity plots (left) and bootscanning (right) of DENV Mexican isolates to demonstrate the lack of resolution of these analysis when very closely related strains are used. Color coding in the bootscanning is the same that for the similarity plot of the same isolate.

No clear pattern of identity was discernible from these similarity plots and the percentages of permuted trees in the bootscans were lower than 70% for most of the length of all sequences (Figure 4-5). Thus, no conclusion could be made about the recombination at this level. For the remainder of this study no sequences with less than 4% difference were included in the same analysis.

Recombination between more divergent strains was then investigated. In most cases, this involved a single sequence of each genotype, but for those genotypes with considerable internal diversity more than one sequence was included in individual runs. Multiple analyses were performed changing the representatives of each genotype. Figure 4-6 shows results of selected analyses of DENV-3 and DENV-4 sequences.

For DENV-3, sequences of genotype III were the closest to the Mexican viruses along all the prM and E genes. No inversions in the identity or phylogenetic patterns were observed in the similarity plots or bootscans respectively (Figure 4-6, upper half). The Somalia/93/S142 isolate was used in those analyses but similar results were obtained when it was replaced by strains from Sri Lanka, India or Samoa as representatives of genotype III (see Figure 3-6 in the previous chapter).

Similar results were obtained with DENV-4. Sequences of two strains in genotype II, either Indonesia/77/1132 or Tahiti/79/S44754, were the closest relative to the Mexican viruses for the entire E gene. Representative plots of these analyses are presented in Figure 4-6 (lower half). Interestingly, the sylvatic Malaysia/75/P75514 strain that was the most distant in the similarity plot reached about 50% bootstrap support in some regions of the bootscan, illustrating the differences in the information provided by these two analyses.

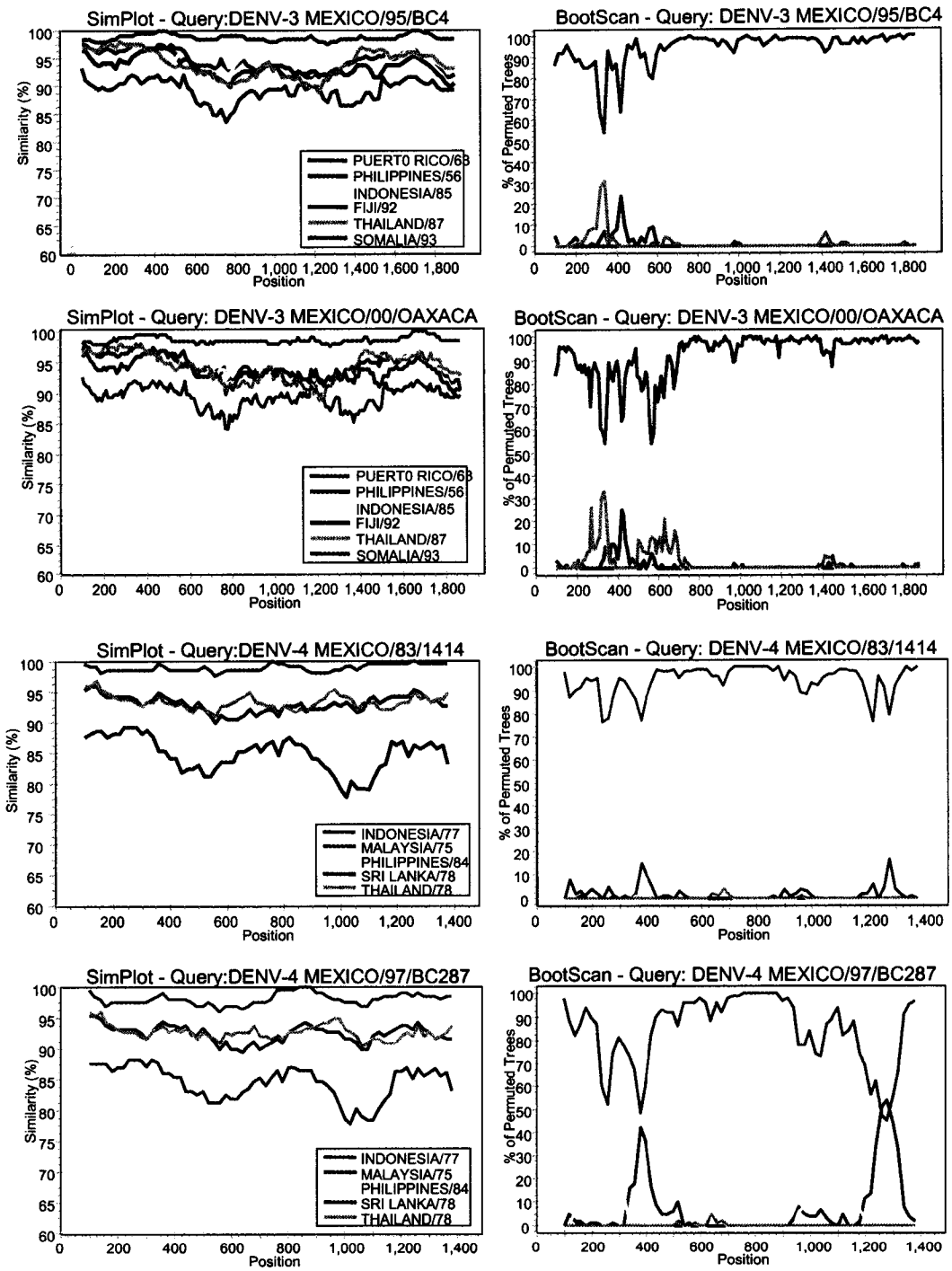


Figure 4-6: Similarity plots (left) and bootscans (right) of DENV-3 and DENV-4 isolates from Mexico. Analysis were performed sliding a window of 200 nucleotides in steps of 10 using the program Simplot. Color coding in the bootscanning is the same that for the similarity plot of the homologous isolate.

For all Mexican DENV-1 sequences used as queries, two genotype V sequences alternated as the most similar to the Mexican viruses in the similarity plots (Figure 4-7, left column). They correspond to isolates Burma/76/PRS228686 and Nigeria/68/IBH28328. Similar results where observed when either of two isolates from Ivory Coast was included instead of the Nigerian (data not shown). All these viruses belong to genotype V (see Figure 3-4). In the bootscans this alternation is even clearer (Figure 4-7, right column). The sequence preferentially supported as the closest to the Mexicans were Burma/76 for the first 560 nucleotides, none from positions 560 to 800, Nigeria/68 from 800 to 1240, and back to Burma/76 for the remainder of the sequence. The percent of permuted trees surpassed 90% for both Burma/76 and Nigeria/68 in the bootscan of Mexico/86/1756.

Since the results were similar with all DENV-1 isolates from Mexico and only one strain is known to have circulated in the Americas, it was hypothesized that the pattern would be the same for all DENV-1 isolates from the western hemisphere. To test this, three other DENV-1 sequences isolated in the Americas were analyzed by bootscanning: Jamaica/77/PRS288690, French Guiana/89/FGA, Brazil/90/BR90. The former is believed to represent the original introduction of DENV-1 in the Americas and the other two were previously identified as recombinants in a previous work (Holmes et al., 1999). The sequence of isolate DENV-1 Singapore/90/275, also reported as a recombinant (Tolou et al., 2001), was also analyzed in the same way. Results are shown in Figure 4-8.

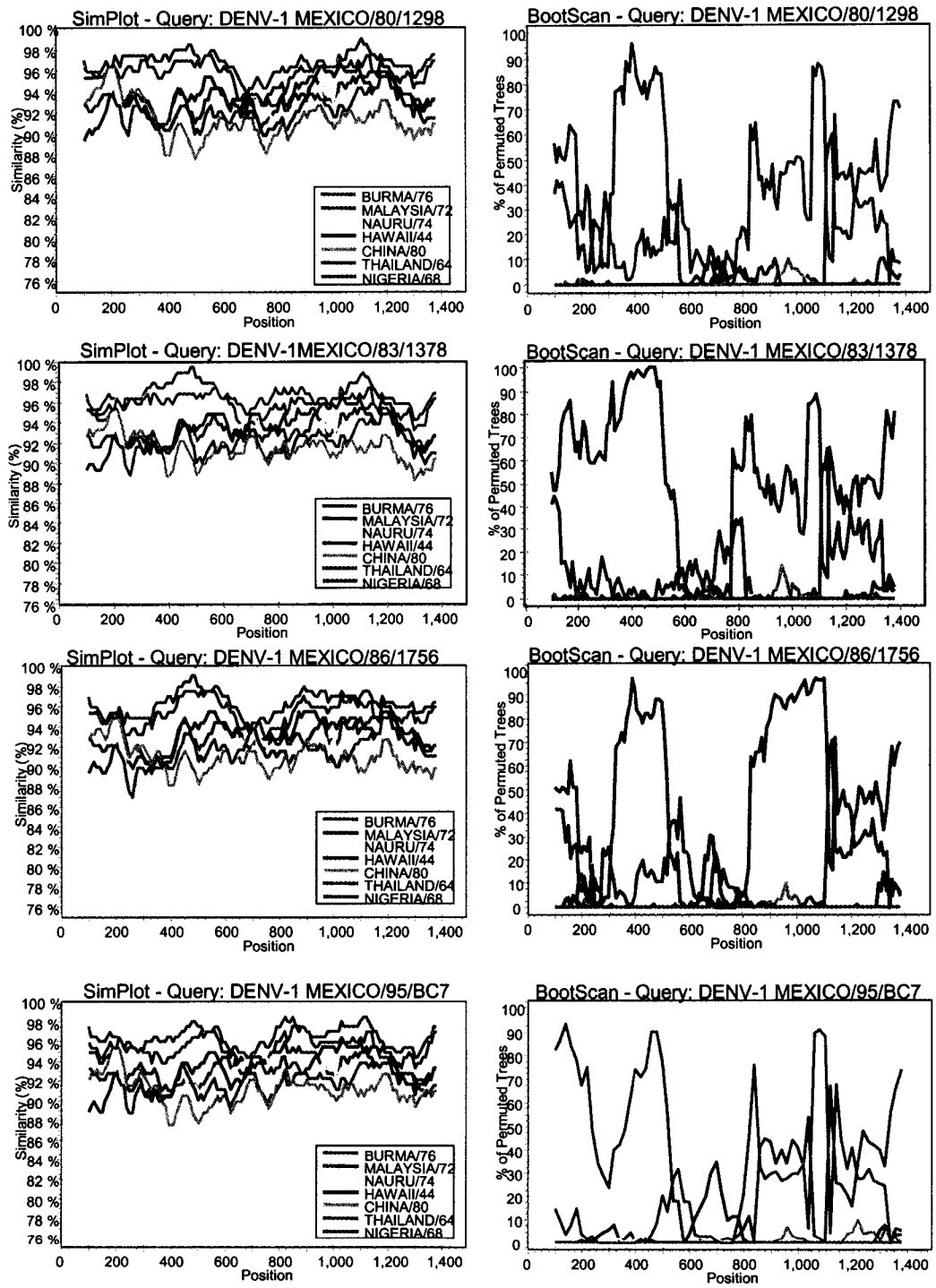


Figure 4-7: Similarity plots (left) and bootscans (right) of DENV-1 isolates from Mexico. Analysis were performed sliding a window of 200 nucleotides in steps of 10 using the program Simplot. Bootscaning was performed by the neighbor-joining method with 250 bootstrap replicates. Color coding for the bootscanning is the same that for all the plots.

Very similar patterns to those of the Mexican DENV-1 were observed in the bootscans of all these strains although the support was usually lower than for isolate Mexico/86/1756 (Figure 4-8). These results suggest that the same evolutionary event(s) may account for the mosaic pattern observed in all these DENV-1 sequences.

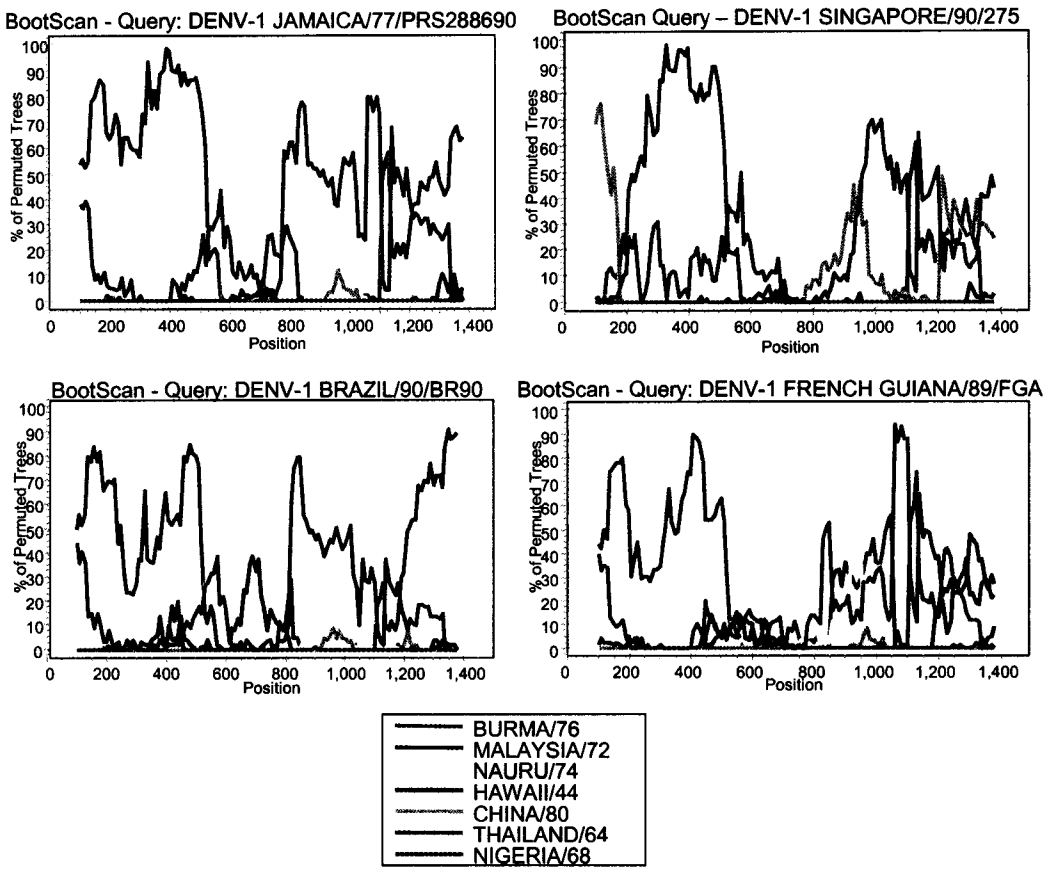


Figure 4-8: Bootscans of DENV-1 Jamaica/77/PRS288690 and three other DENV-1 isolates previously identified as recombinants (Holmes et al., 1999, Tolou et al., 2001) to demonstrate the similarity among patterns. Analysis were performed sliding a window of 200 nucleotides in steps of 10 using the program Simplot. Bootscanning was performed by the neighbor-joining method with 250 bootstrap replicates. Color coding for the bootscanning is the same that for all the plots.

To test the possibility of recombination even further, Chimaera analysis was performed and maximum likelihood trees were constructed for the sequences delimited by the putative breakpoints. The results are shown in Figure 4-9. As suggested by the bootscan, the Mexico/86/1756 isolate clusters with the African isolates in the intermediate segment (nts. 500-1240) but with the Burma isolate for the remainder of the E gene. Two peaks are observed in the plot of p-values generated with Chimaera, but their positions do not coincide with the breakpoints suggested in the other methods.

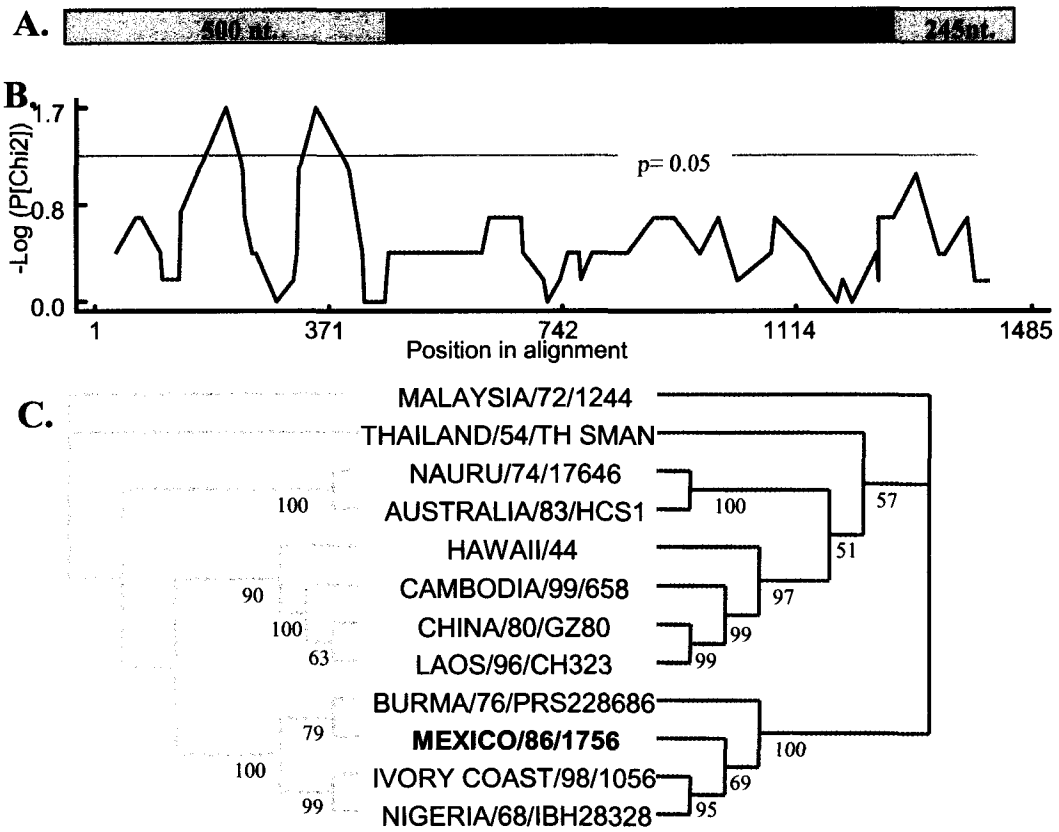


Figure 4-9. Additional test for recombination in DENV-1 Mexico/86/1756 isolate. **A.** Graphic depiction of DENV-2 E gene showing the segments of the putative recombination event. **B.** Plotting of the log p value of the χ^2 statistic computed by Chimaera, executed in the RDP program (see text for details). **C.** Cladograms showing the different clustering of Mexico/86/1756 depending on the analyzed segment. Color of the trees corresponds to the segments used in the maximum likelihood analysis as delimited in A. Numbers are 1,000 replicates bootstrap values. The trees were constructed in PAUP* 4.01.

Results for some DENV-2 strains analyzed for recombination in the C-prM-E-NS1 segment are presented in Figure 4-10. While the most recent isolates from 2001 and 2002 were more closely related to Jamaica/83/1409 and other isolates in the American-Asian genotype, those of 1983, 1984 and 1992 were more similar to other sequences in the American genotype, either Puerto Rico/69/159, Puerto Rico/77/1328 or Peru/96/IQT2913. However the Mexico/83/200787 sequence exhibited an abrupt inversion of the similarity pattern near the end of the analyzed segment, diverging from the American genotype and becoming almost identical to the Jamaica/83/1409 strain in the last 119 nucleotides of E gene (top of Figure 4-10). This result is corroborated in the bootscan where the switch between genotypes is supported by more than 95% of the permuted trees.

To further substantiate this potential recombination event, Chimaera analysis was applied and separate maximum likelihood phylogenetic trees were constructed for the last 119 nucleotides and for the remainder of the analyzed sequence. The results are shown in Figure 4-11. Chimaera detected a crossover that appears as a neat peak in the plot of p-values in the region where the other methods suggested recombination. Another peak is also observed near the center of the sequence, probably reflecting a region of low homology between Mexico/83/200787 and other sequences of American genotype of DENV-2, which is observed in the similarity plot and the bootscan (Figure 4-10, top). In the separate maximum likelihood analysis, Mexico/83/200787 isolate clusters within the American genotype in the larger segment of its sequence and this clustering is supported by a 100% bootstrap value. For the last 119 nucleotides this isolate clusters in the American-Asian genotype with bootstrap support of 73%. Lower bootstrap values for the 3' end of the E gene are probably due to the shortness of the segment.

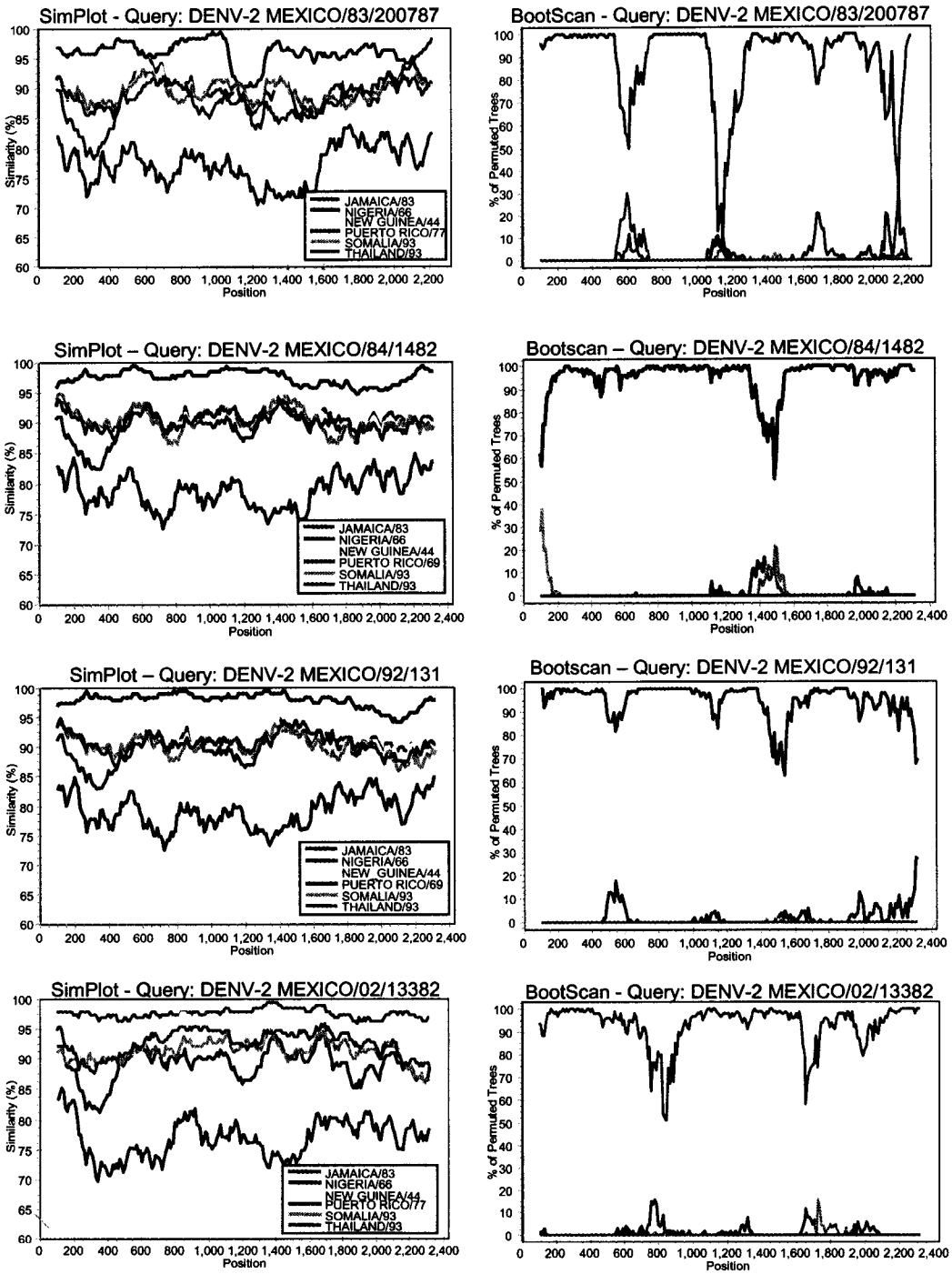


Figure 4-10. Similarity plots (left) and bootscans (right) of DENV-2 representatives of the American and American-Asian genotypes from Mexico. Analysis were performed sliding a window of 200 nucleotides in steps of 10 using the program Simplot. Bootscanning was performed by the neighbor-joining method with 250 bootstrap replicates. Color coding is the same that for all the plots.

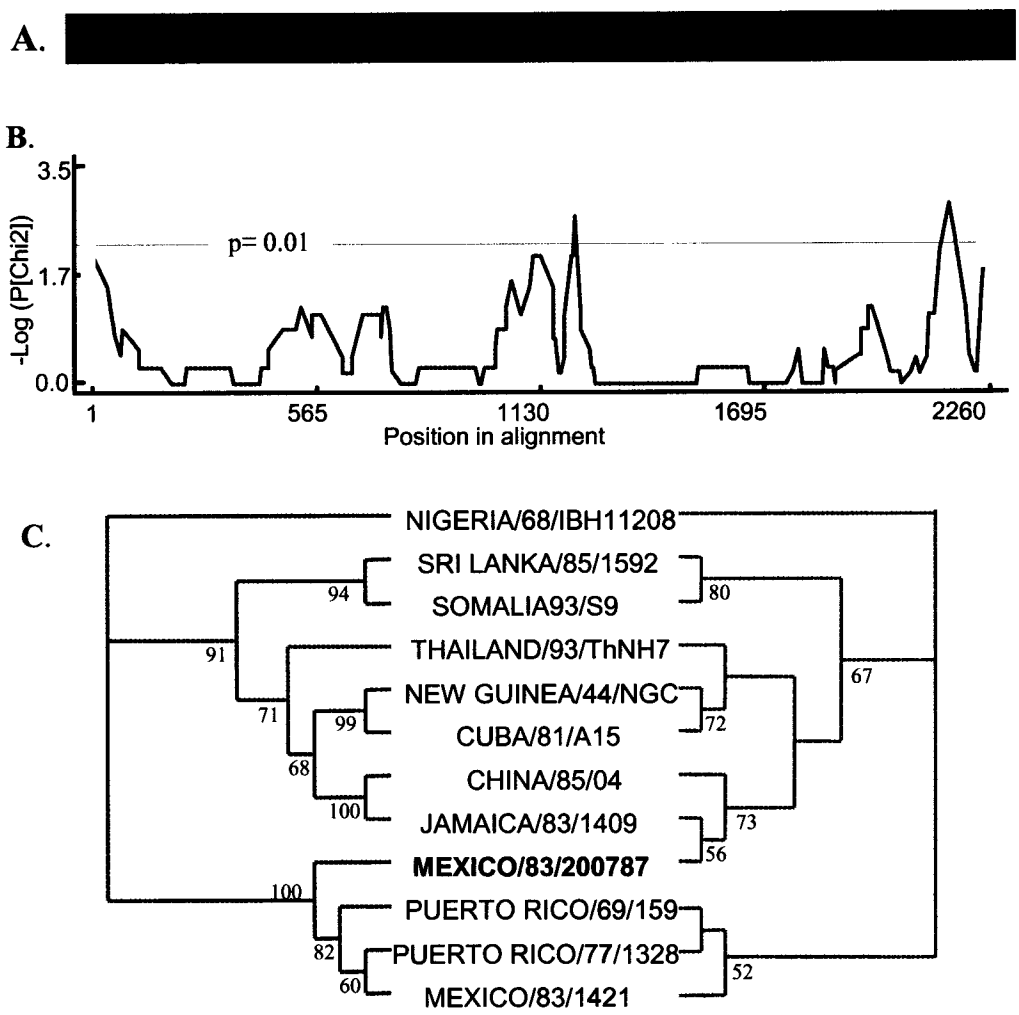


Figure 4-11. Additional test for recombination in DENV-2 Mexico/83/200787 isolate. **A**. Graphic depiction of DENV-2 E gene showing the two segments of the putative recombination event. **B**. Plotting of the log p-value of the χ^2 statistic produced with Chimaera and executed in the RDP program (see text for details). **C**. Cladograms showing the different clustering of Mexico/83/200787 depending on the analyzed segment. Color of the trees corresponds to the E gene segments delimited in A. Numbers in the trees are 1,000 replicates bootstrap values. The trees were inferred with the maximum likelihood algorithm in the PAUP* package and drawn with the TreeView program.

Four Mexican DENV-2 strains belonging to different genotypes were investigated for recombination in their complete genome sequence. Bootscans of these analyses are shown in Figure 4-12. Isolates MX/QuintanaRoo/94/BC139, MX/Guerrero/97/C-932 and MX/Yucatan/01/12914 clearly resemble viruses in the American, Asian-2 and American-Asian genotypes respectively in their entire genome. Results for isolate MX/Yucatan/96/BC17 were not so clear. This sequence did not closely resemble any other isolate in the similarity plot although in the bootscanning some support was found for its identity as a divergent isolate in the Cosmopolitan genotype. This result is in agreement with that in the phylogenetic analysis (Chapter 2). No evidence of recombination was detected in these full-genome analyses.

DISCUSSION

Currently available tools to search for recombination are not very sensitive. Most of them fail to detect recombination between sequences with less than 5% divergence (Posada, 2002, Posada & Crandall, 2001, Wiuf et al., 2001). The homoplasy test (Maynard Smith & Smith, 1998) is an exception but it produces an unacceptable rate of false positives (Posada & Crandall, 2001). Our results confirm the inadequacy of these methods for closely related sequences (Figure 4-5). Most procedures also fail to detect remote recombination events (Posada et al., 2002). Following the recommendations of these studies, various methods were used here to maximize the likelihood of detecting potential recombinants and to avoid conclusions based in a single method.

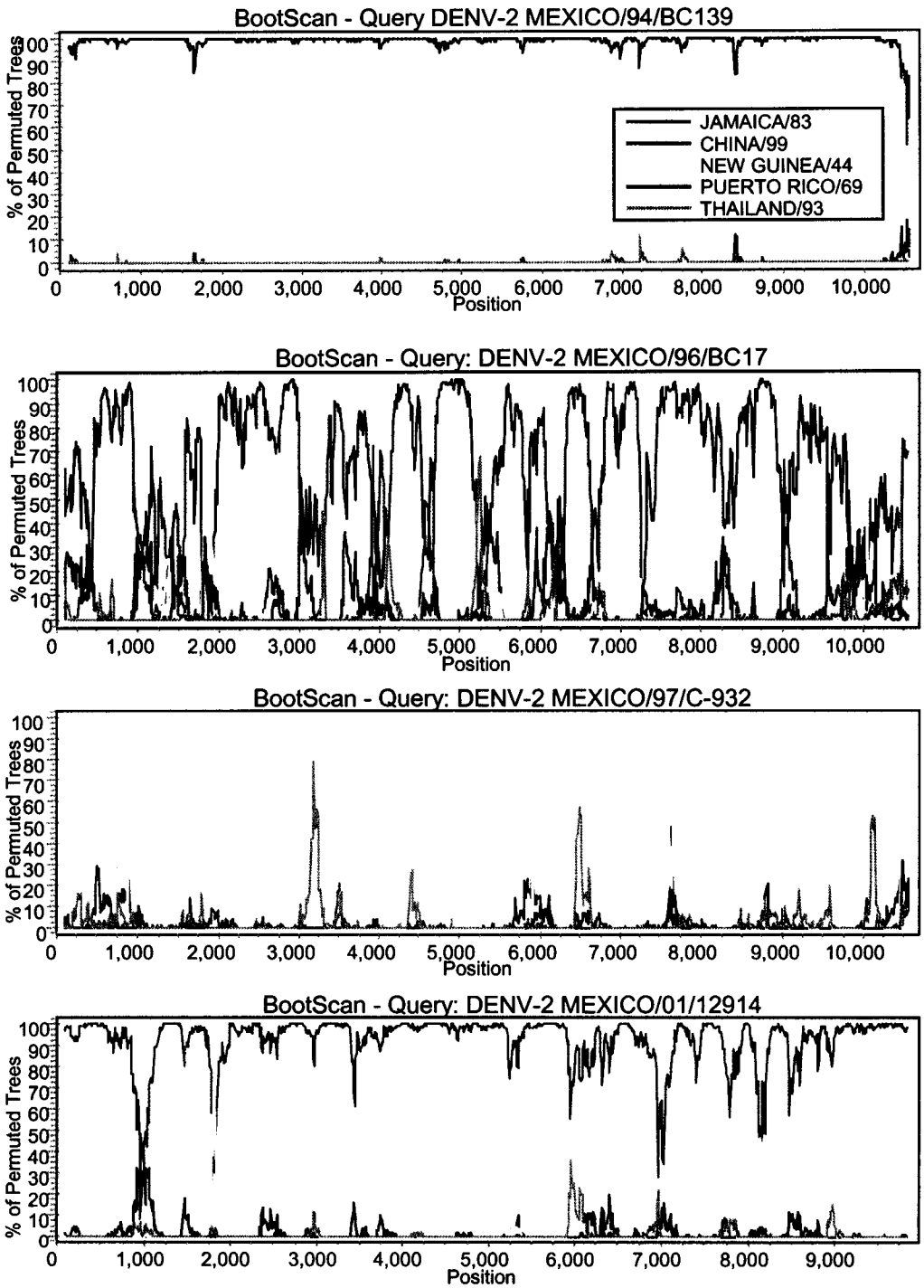


Figure 4-12: Bootscans of the complete genome of Mexican DENV-2 representatives of, from top to bottom, American, Cosmopolitan, Asian-2 and American-Asian genotypes. Analysis were performed sliding a window of 200 nucleotides in steps of 10 using the program Simplot. Bootscaning was performed by the neighbor-joining method with 250 bootstrap replicates. Color coding is the same that for all the plots.

Similarity plots and diversity plots are graphical methods extensively used to search for recombinants in different kinds of data sets. These methods were used to detect most of DENV recombinants reported so far (Worobey et al., 1999). However, these are exploratory tools that do not permit assessment of the significance of the results. Phylogenetic methods like bootscanning have also been frequently used. By performing the conventional permutations they allow some statistical testing of the reliability of their graphic output. Here we used bootscanning and defined a conservative 90% limit of significance. However some researchers think that comparing relative levels of bootstrap support is an inappropriate way to test alternative hypotheses (Posada & Crandall, 2001). Methods based on distribution of substitutions are not visually explicit and in some cases do not identify which sequences in the dataset are recombinants or parents but they do generate summary statistics providing for more conventional hypothesis testing. Here we used the maximum match chi-square test or Chimaera (Posada & Crandall, 2001) as a confirmatory tool in sequences that exhibited patterns suggestive of recombination in the previous tests.

Two probable cases of recombination were found in this study. The DENV-1 sequences of Mexico and Jamaica/77 oscillate between an Asian (Burma/76) and several African (Nigeria/68 and Ivory Coast/98) sequences as the more likely ancestors (Figure 4-7). This putative recombination event is reminiscent of the first reported DENV chimeric sequences (Holmes et al 1999). In that work, using diversity plots two isolates from Brazil (BR/90) and French Guiana (FGA/89) were identified as the product of recombination between Asian and Caribbean viruses (represented by the Singapore/90/S275 and Jamaica/77/CV1636 isolates, respectively). The more likely breakpoint was located between nucleotides E-306 and E-312, not very distant to

positions E-426 and E-506 of the first crossover estimated in this study with a different test. The Singapore/90/S275 isolate was, in turn, identified as a recombinant by Tolou et al (2001) with the more likely parents being a West African (Ivory Coast/99) and either an Asian (Cambodia/98) or an East African (Djibouti/98) viruses. The crossover was located in positions E-534 or E-164, depending on the second parent used for the estimation.

The similarities in geographic origin of the DENV-1 strains and in the location of the crossovers suggest that this and other studies (Holmes et al., 1999, Tolou et al., 2001) have identified the same recombination event, which probably happened well before the introduction of DENV-1 into the Americas and involved Asian and/or African strains. The original product of that event probably remained in Asia (as attested by the Singapore/90 isolate) and was introduced to the Americas in 1977.

For this possible recombination event, the boundaries of the genomic exchange are not very clear and the statistical support is not very strong. It could be that the sequences identified here as parental are only distantly related to the true parents. Two or more events could be superimposed in this genomic region. Another possibility is that the recombination is so remote that many substitutions have accumulated in the sequences blurring the identities of the parents and the positions of the breakpoints. All of these possibilities can confuse the results and decrease the statistical significance of the analyses (Posada & Crandall, 2001).

The second possible recombination reported here is more clear. The sequence of Mexico/83/200787 isolate is a mosaic between two strains previously detected in the western hemisphere: the American DENV-2 genotype, represented here by isolates

Puerto Rico/69/159 and Puerto Rico/77/1328, and the American-Asian genotype, represented by Jamaica/83/1409. Isolates related to the latter were not detected in Mexico before year 2,000, but it co-circulated with viruses of the American genotype in the Caribbean during the early 1980's (Vorndam et al., 1994). Thus, the recombinant event probably originated in that region and was introduced to Mexico later. Another DENV-2 isolated in Mexico in 1983-4 and 1992-4 did not have evidence of this event, so the recombinant strain was apparently not widely dispersed. Antigenic, phylogenetic and pathogenic studies of the Mexico/83/200787 have been conducted by others (Monath et al., 1986, Ruiz et al., 2000, Sanchez & Ruiz, 1996). This virus has an atypical profile of reactivity with a panel of monoclonal antibodies (Monath et al., 1986). In a phylogenetic study, this isolate did not cluster with any other isolate and branched before the splitting of DENV-2 into genotypes (Ruiz et al., 2000). This could have been caused by the mosaic structure of the sequence. Topological distortions in phylogenetic analyses can be produced by recombination (Schierup & Hein, 2000a).

The co-infection of humans with different serotypes has been documented many times (Gubler et al., 1985, Lorono-Pino et al., 1999, Wang et al., 2003). However we did not detect interserotype recombinants that would have been very evident in the similarity plots. Other studies have also failed to detect such recombinants (Worobey et al., 1999). Their absence could be related to incompatibility between different sequences at different loci in the genome of different serotypes. In contrast, recombinations between different genotypes of the same serotype have been detected several times, although the concurrent circulation of different genotypes is a more rare event. Taken together, these two facts suggest that the rate of recombination increase with sequence similarity, which has been experimentally observed in other viruses (Lai, 1992). As a corollary, it seems reasonable to expect that recombination between

strains of the same genotype or between viruses of the same strain may be frequent. The lack of resolution of detection methods in these situations could be due, in part, to the fact that a large number of crossovers would lead to excessive sequence scrambling and loss of the little signal contained in an alignment with a low number of informative sites (Holmes & Burch, 2000). Most detection methods are expected to perform poorly under such circumstances (Posada et al., 2002).

This study expanded our perception of the importance of recombination as a fundamental force in the evolution of DENV. However, the number of recombination events detected here could underestimate the real frequency of that kind of genetic exchange. First, the entire genome was studied in four DENV-2 strains only, and second, the methods available to search for recombination were inadequate to detect mosaics of closely related genomes. Without a good estimation of the frequency of such events, the relevance of recombination in natural populations will not be completely appreciated.

CHAPTER 5

DEVELOPMENT OF AN *IN-VITRO* MODEL FOR DENGUE VIRUS RECOMBINATION

Recombination, defined as the exchange of genetic material between different molecules of DNA or RNA, is one of the driving forces in virus evolution. Its double role in the generation of diversity and elimination of deleterious mutations has been extensively reviewed and was presented in the previous chapter of this dissertation (Domingo & Holland, 1997, Lai, 1992, Posada et al., 2002, Worobey & Holmes, 1999).

There is now an abundant literature documenting recombination in RNA viruses (summarized in Table 4-1). RNA virus recombination has been reported more frequently in retroviruses and single-stranded positive sense (+) RNA viruses (Lai, 1992). A few cases of recombination in double-strand (Suzuki et al., 1998) and single-stranded negative sense (-) RNA viruses (Chare et al., 2003, Sibold et al., 1999) have also been reported. The molecular basis for the rarity of recombination in (-) RNA viruses is not completely understood, but could be related to the close association between the genomic RNA and the ribonucleoprotein complex (RNP) in these viruses. The RNP never disassembles from the (-) RNA and may, therefore, protect the viral RNA from exchanges of genetic material, regardless of the mechanism used in recombination (Chare et al., 2003, Conzelmann, 1998).

In the previous chapter a summary review of the presumed evolutionary consequences of recombination in RNA viruses was presented, as well as information on the currently available methods used to detect recombination in genomic sequences. A discussion on the mechanism of recombination in RNA viruses is lacking and is pertinent to the studies that will be presented here. Therefore, a review of the current knowledge on the molecular processes involved in RNA recombination is presented in the following paragraphs.

As in DNA recombination, RNA recombination has been classified as homologous or nonhomologous (Lai, 1992). Homologous recombination happens when two similar (or identical) RNA molecules exchange comparable genomic regions, which implies that the junction (crossover or brakepoint) between these molecules takes place in the same (or very close) sites in both genomes. The resulting recombinant molecule will be another similar genome that could be recognized as such only if there are differences between the parental molecules at both sides of the crossover point (Posada et al., 2002). Sometimes deletions, insertions, mismatches or non-templated nucleotides are introduced at or close to the breakpoint of the resulting recombinant; these cases have been denominated "aberrant homologous" recombinants (Lai, 1992).

Nonhomologous recombination is the exchange of unrelated sequences between genomes. In RNA viruses, it can happen between two unrelated virus infecting the same organism, between cellular and viral RNAs or between two different segments in the genome of the same virus (Nagy & Simon, 1997). Some examples of nonhomologous recombination follows:

1. Mouse hepatitis virus, unlike most other coronaviruses, contains a hemagglutinin-esterase gene that seems to be derived from Influenza C virus (Lai, 1992).

2. Bovine viral diarrhea virus (BVDV), a pestivirus, occasionally incorporates cellular sequences, most often the ubiquitin gene, into its RNA genome. This events force the virus to switch from a noncytopathic to a highly cytopathic phenotype, and generates the transformation of a chronic, but otherwise harmless infection, in a fatal disease (Kummerer et al., 2000).

3. The genome of turnip crinkle virus (TCV), a carmovirus, often recombines with satellite RNAs and subviral RNAs that may accompany the genomic RNA in the virion (Cascone et al., 1990).

4. Brome mosaic virus (BMV), a plant virus, has a multipartite genome composed of three RNA segments that contain an almost identical 3' non-coding region. Both homologous and nonhomologous recombination has been observed between the different segments of this virus. A great portion of the experimental work in the mechanism of recombination has been done using BMV as a model of both types of recombination (Alejska et al., 2001, Figlerowicz & Bujarski, 1998, Nagy & Simon, 1997).

The two main models that have been postulated to describe the mechanism of viral RNA recombination are: 1) breakage and ligation and 2) template-switching models. In the breakage and ligation model, recombination occurs independently of the synthesis of RNA. This is the accepted mechanism of recombination in DNA genomes. There are specific enzymes in DNA organisms that cut and rejoin the double stranded DNA molecules during recombination, a mechanism intrinsic to the life cycle of many species. No such enzymes are encoded in RNA virus genomes, but the cell could

provide them. Furthermore, ribozyme-mediated RNA breakage and ligation mechanisms have been demonstrated for splicing of group II introns (Morl & Schmelzer, 1990) and for repairing bacterial mRNA (Sullenger & Cech, 1994), which indicates that a similar mechanism is theoretically possible in viral RNA recombination.

An end-to-end transesterification reaction was proposed to explain the generation of recombinants between bacteriophage Qβ RNAs (Chetverin et al., 1997). This proposal was based in the observation that altering the 3'-end OH group of one of the recombining templates interfered with the subsequent recovery of recombinants. However, this observation does not completely preclude other enzyme-driven mechanisms (Nagy and Simon, 1997). Although the breaking and ligation model for recombination in RNA virus remains largely unproven, a better explanation for the recently demonstrated generation of recombinant RNAs in the absence of viral replication is lacking (Gallei et al., 2004).

The template-switching model, also known as copy-choice, has been supported in most studies on recombination in RNA viruses. In this model, mosaic RNAs are formed when the synthesis of a nascent RNA molecule by the viral RNA-dependent RNA polymerase (RdRp) is halted. The nascent molecule dissociates from the original (donor) RNA template and bind a different (acceptor) RNA template. The synthesis of the new genome is continued on the acceptor RNA template. During this process the RdRp could remain attached to the 3'-end or completely detach from the nascent strand. In any case an RdRp molecule would be able to find the 3'-end of the incomplete strand and use it as a primer to continue the synthesis of the new (recombinant) RNA genome (Alejska et al., 2001, Nagy & Simon, 1997).

The template-switching model invokes RNA replication as an essential component of the recombination process, and thus it is referred as a replicase-driven mechanism. Indeed, it was selected as the more likely model because recombinant genomes were no longer obtained under conditions in which RNA synthesis was blocked (Kirkegaard & Baltimore, 1986). The capacity of polymerases to synthesize RNA or DNA without pausing is referred to as the processivity of the enzyme. However, there is evidence that RNA synthesis is intrinsically nonprocessive. Polymerase pauses have been demonstrated during both RNA-dependent and DNA-dependent RNA synthesis (Mills et al., 1978). These pauses are supposed to facilitate the jumping of the nascent strand (and the polymerase) from one template to another.

The frequency and location of template-switching events seems to be dependent on primary and secondary structures in the donor, acceptor and nascent RNA molecules (Kim & Kao, 2001, Nagy et al., 1998). Among the primary structures that favor the polymerase jumping from one strand to the other are short U-rich stretches in the 3'-end of the nascent strand and AU-rich regions in the donor or acceptor templates (Nagy & Bujarski, 1996, Nagy et al., 1998). Formation of heteroduplexes between acceptor and donor strands is another important structure promoting template-switching. The region of complementarity between templates could be partially penetrated by the RdRp, but it would eventually force a pause in the synthesis and detachment of the nascent strand and RdRp from the donor template. The nascent strand could then bind the acceptor strand and the RNA synthesis resumed (Nagy et al., 1998, Romanova et al., 1986). In these cases the breakpoints are randomly distributed along the region of duplexing (Nagy & Simon, 1997).

Secondary structures in RNA are also important. Stem and loop (hairpin) structures could act as RdRp pausing signals in the donor template or as RdRp entry sites in the acceptor strands, as demonstrated in TCV and BMV (Alejska et al., 2001, Cascone et al., 1993). Some of these secondary structures function as replication enhancers or subgenomic promotors during normal RNA replication (Nagy et al., 1999, Panaviene & Nagy, 2003, Wierzchoslawski et al., 2003). Short stretches of either sequence homology (Cascone et al., 1993) or sequence complementarity (Alejska et al., 2001) adjacent to the hairpin are also important. In these cases in which a hairpin is promoting template-switching, recombination is more frequently nonhomologous and site-specific. Thus, these sites behave as hot-spots for recombination.

The RdRp itself can also determine the frequency and nature of recombination. Experiments mutagenizing the RdRp of BMV demonstrated that some substitutions influence the frequency of both homologous and nonhomologous recombination as well as the location of homologous crossovers (Figlerowicz et al., 1998). Location of nonhomologous breakpoints was not affected and seems to depend only on the secondary structure of the templates. Interestingly, a single substitutions in the core polymerase domain of RdRp inhibited nonhomologous RNA recombination without affecting the frequency of homologous crossovers, suggesting that some differences exists in the mechanisms of this two kinds of recombination (Figlerowicz et al., 1997). Similar mutagenesis experiments have been reported in other RNA viruses (Panaviene & Nagy, 2003). These data demonstrate that recombination of RNA viruses is mostly a replicase-driven process and provide support for the template-switching model.

Although this study focused on dengue virus (DENV), a positive sense RNA virus, a brief discussion of recombination in retroviruses fellows for methodological reasons.

The (+) RNA retroviral genome is copied into a double stranded (ds) DNA by the viral reverse transcriptase (RT). The peculiarities of the retrovirus life cycle demand the RT exchange templates at least two times during the normal replicative cycle (Katz & Skalka, 1994). RT has also been observed pausing during DNA synthesis at specific sites, or randomly at sites of breakage of the RNA template (Negroni & Buc, 2001). These characteristics make RT very susceptible to jumping from one template to other. RT-driven recombination has been observed during the synthesis of both minus-strand (-) DNA and plus-strand (+) DNA. In the former, it can proceed during the normal template exchange near the genome ends (strong-stop recombination) or in internal regions by the copy-choice model (Negroni et al., 1995). During the (+) DNA synthesis recombination occurs by a different mechanism referred as strand displacement-assimilation (Katz & Skalka, 1994, Negroni & Buc, 2001). These recombination events can be observed in cell culture, or in-vitro during reverse transcription mediated by either isolated virus particles or purified RT molecules.

As in (+) RNA viruses, recombination in retroviruses seems to be influenced by the primary and secondary structures or the RNA templates (Moumen et al., 2001, Negroni & Buc, 2001). It is also a temperature dependent process: recombination frequency decreases when the temperature is increased (Li & Zhang, 2001). Degradation of the template RNA by RNase-H domain of RT seems to be essential to RT-mediated recombination since mutagenesis of the RNase-H domain reduce or completely suppresses template-switching (Hwang et al., 2001, Negroni et al., 1995).

DENV, a (+) RNA virus, was recently added to the list of RNA viruses in which recombination has been detected. In the original reports, DENV sequences that had been previously determined for phylogenetic purposes were reanalyzed using split

decomposition and sliding window approaches to search for conflicting phylogenetic signals in different parts of the genome (Holmes et al., 1999, Worobey et al., 1999). About 10% of the studied sequences contained conflicting signal suggestive of recombination. In most of the cases, the putative recombinants seemed to be the product of genomic exchanges between two strains belonging to different genotypes within the same DENV serotype.

In the aftermath of these pioneering works, other reports of recombination in DENV have been published (Tolou et al., 2001, Twiddy & Holmes, 2003, Uzcategui et al., 2001), and additional cases were detected and presented in the previous chapter. All these suggest that recombination is not a rare event in natural populations of DENV. The validity of these findings has been questioned (Rico-Hesse, 2003). Indeed, sequences of some isolates closely related to those that were originally reported as recombinants have been recently published. They exhibit significant differences with the previous sequences and no evidence of recombination is found on them (Goncalvez et al., 2002, Peyrefitte et al., 2003).

One of the difficulties in resolving these ambiguities is the absolute lack of experimental work on recombination with DENV (Holmes & Burch, 2000). No information about frequency, mechanisms and factors promoting or constraining recombination in DENV is available. Such information would provide support for or dispel previous claims of recombination in natural populations and provide insight into the consequences of such events in the evolutionary potential of DENV.

In this study, an experimental system to test and manipulate recombination in DENV was investigated. Since information is lacking about the extent of homology

required to produce viable recombinants, the system was designed to test for homologous recombination between different genotypes of the same serotype, which have been reported in natural DENV populations.

MATERIALS AND METHODS

Experimental Design. Figure 5-1 depicts the main characteristics of the system designed to test for recombination. Coinfections (or crosses) of cells with two DENV-2 strains at a high multiplicity of infection (moi) was performed to assure that most cells were infected with both viruses. Viruses were allowed to propagate in cells for seven days. At the end of this period, total RNA was extracted from the cells, reverse transcribed using a consensus primer and amplified by PCR utilizing a forward primer specific for one strain and a reverse primer specific for the other strain. Controls for recombination during reverse transcription were performed mixing RNAs extracted from single virus-infected cultures. Controls for recombination during the PCR used mixtures of two cDNAs generated from RNAs of individual strains (Figure 5-1).

This system relied on the PCR property of only exponentially amplifying DNA when both primers anneal to the same DNA template. When performed with heterologous primers in a mixture of recombinant and nonrecombinant templates, this system would only amplify exponentially recombinants that have the right mosaic structure, so that both forward and reverse primers recognize the corresponding annealing sequences. Under these conditions nonrecombinant templates were expected to anneal to only one of the primers and a linear (nonexponential) amplification of molecules of uneven length would result (Figure 5-2).

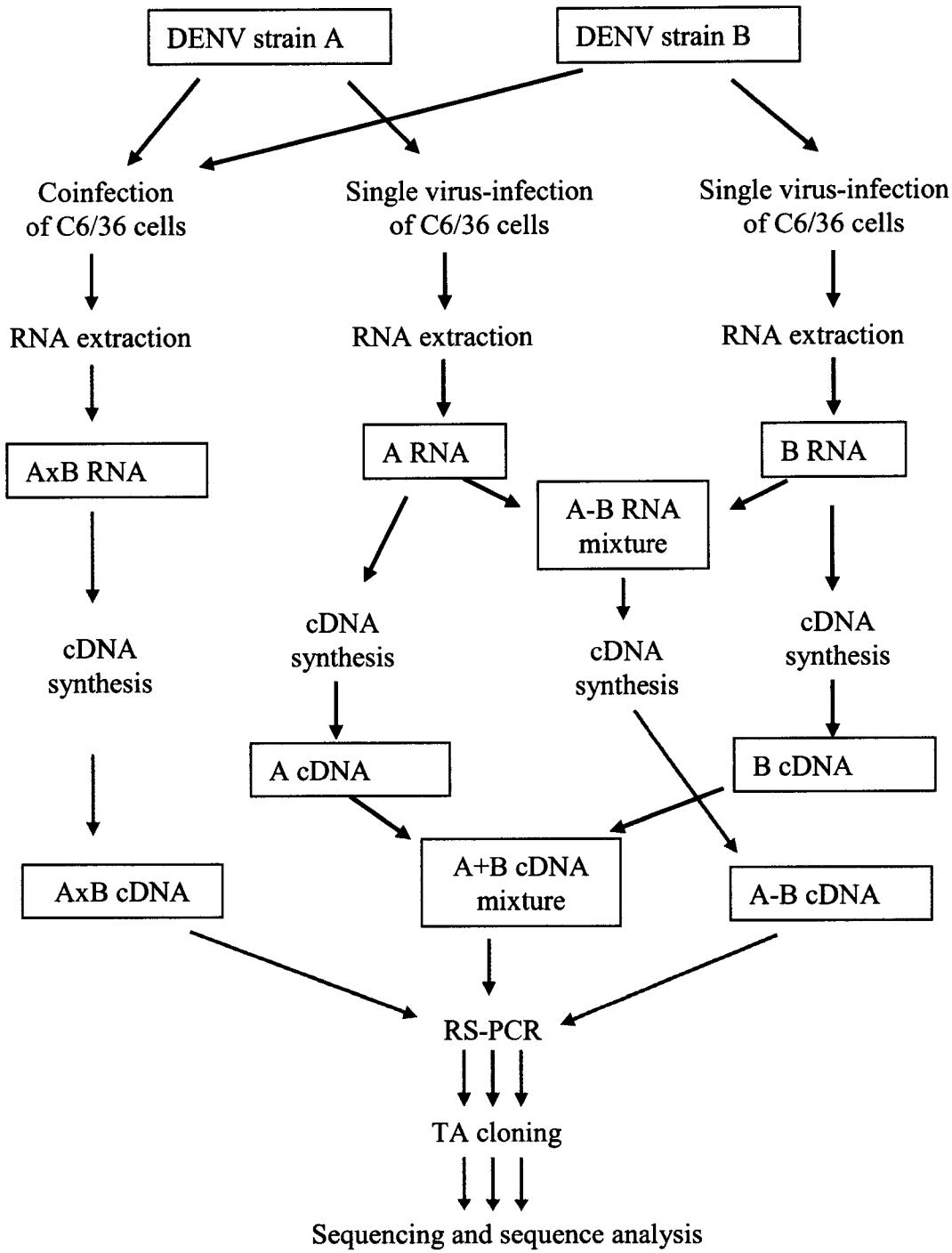


Figure 5-1. Schematic depiction of the strategy and procedures used in this study to search for recombination between two hypothetical DENV strains, A and B. Red, blue and green arrows indicate the order of procedures to generate AxB, A-B and A+B molecules, intended to test for recombination in cell culture, in reverse transcription and in recombinant specific PCR (RS-PCR), respectively. See text for more details.

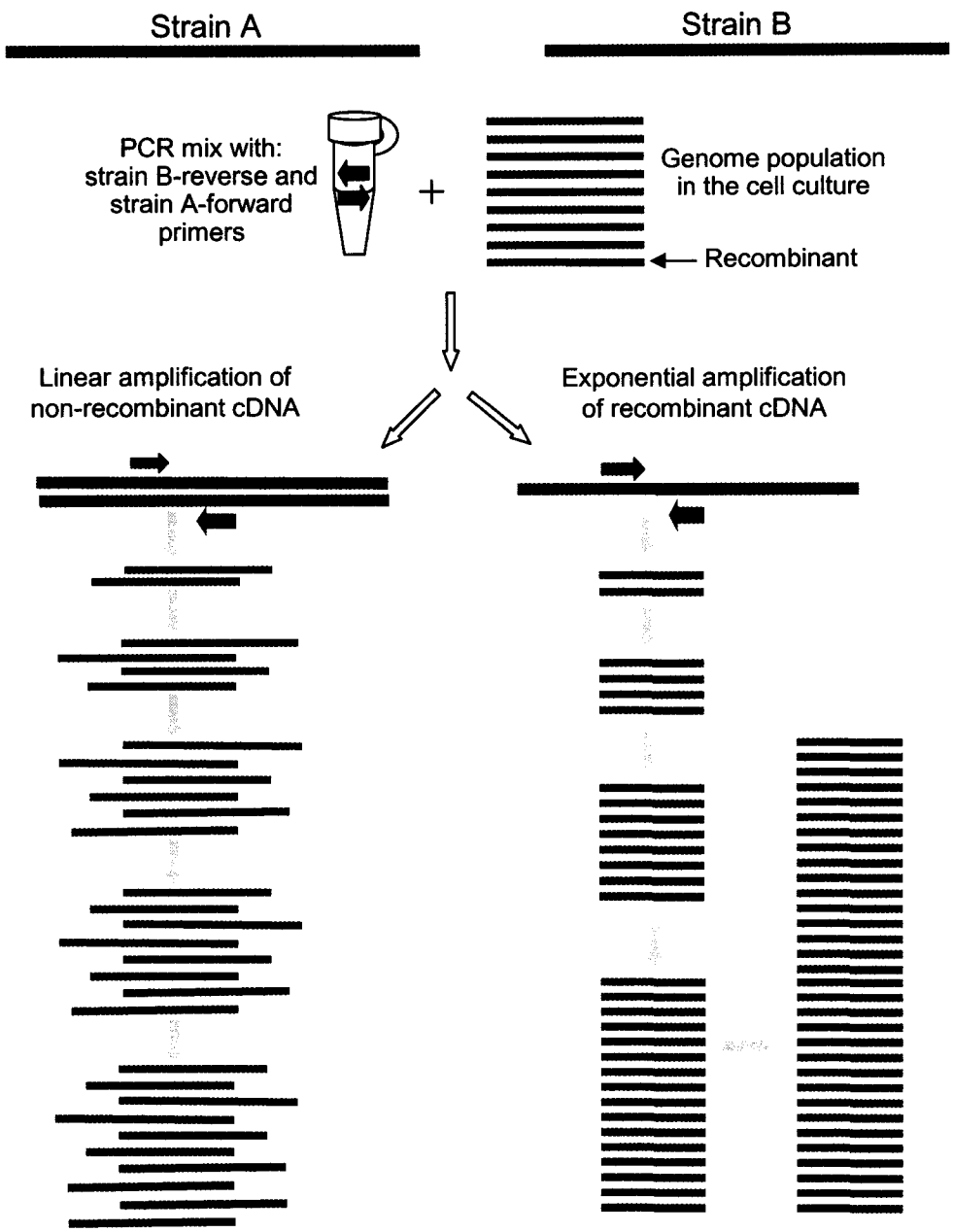


Figure 5-2. Representation of the selective exponential amplification of recombinant molecules during recombinant specific PCR (RS-PCR). Red and green bars represent sequences of two strains A and B, respectively. Two colored bars represent hypothetical recombinants between A and B. Primer specificity is represented by the same colors. Gray arrows are cycles in the PCR.

From now on, this system will be designated "recombinant-specific PCR" (RS-PCR). The products of this RS-PCR were purified and cloned in plasmids to select individual molecules. The segment of the plasmid containing the insert was sequenced and the resulting sequence was compared to those of the original (parental) strains. A more detailed description of these procedures follows.

Virus, Cells and Coinfections. Three DENV-2 strains belonging to different genotypes were selected for these experiments Jamaica/83/4109 (JAM), Sri Lanka/85/1592 (SL) and Puerto Rico/69/159 (PR). They were between the 4^{th} and 8^{th} passage. All the experiments were conducted in *Ae albopictus* C6/36 cells at 28 °C. The cells were grown in Leibowitz L-15 medium with 10% fetal bovine serum (FBS) and maintained in the same medium with 2% FBS. Virus stocks were titrated by end point dilution in C6/36 cells as previously described (Schoepp & Beaty, 1984).

To test for recombination, pairs of different DENV-2 strains were simultaneously inoculated in C6/36 monolayers, each virus at a moi of 3.0 $TCID_{50}$/cell. Experiments were performed in 75 cm^2 cell-culture flasks. Viruses were allowed to adsorb to cells for one hour at room temperature with gently rocking. The inocula were replaced by L-15 medium with 2% FBS and viruses were allowed to propagate for seven days. Aliquots of cell culture supernatant were then harvested and frozen at –70 °C. The cells were scraped, pelleted and frozen at the same temperature. C6/36 monolayers were also used to propagate individual strains to be used as controls in the experiments.

RNA extraction and reverse transcription. Cell pellets were lysed, homogenized and total RNA was extracted using the QIAshredder and Rneasy kits (Qiagen, Valencia, CA), a silica gel-based method. RNA was spectrophotometrically quantified. All cDNA syntheses in this study used reverse primer DV3-5368R (Table 4-2). This is a DENV-complex consensus primer annealing in the NS3 gene. The RT used in most experiments was SuperScript™ II Rnase H⁻ RT (Invitrogen-Life Technologies, Carlsbad, CA). Reverse transcriptions were performed in 0.6 ml tubes. The initial mixture contained 0.4 μL of dNTP mix (25mM each), 3.5 pmole of the reverse primer, 0.5-2.0 μg or total RNA and nuclease free (diethilpyrocarbonate treated) distilled water to 12 μL. This mixture was heated to 65°C for 5 minutes and quickly chilled on ice. Eight 8 μL of a second mixture containing 4 μL of 5X RT buffer, 0.1 M DTT, 20 units of ribonuclease inhibitor and 200 units of RT was added later, mixed by vortex and incubated at 42°C during 2 hours. The reaction was stopped by heating to 70 °C for 15 min., chilled to 4 °C and finally frozen at -20°C. In some experiments SuperScript™ II RT was replaced by ThermoScript™ RT (Invitrogen-Life Technologies, Carlsbad, CA), another RNase H⁻ reverse transcriptase, that is stable at higher temperatures. In these cases the temperature used during the cDNA synthesis varied between 42°C and 62°C (see results).

Primer design. The specificity of the RS-PCR required a careful design of the primers. To do this, sequences of the DENV-2 strains JAM and PR were downloaded from GenBank (accession numbers M20558 and M191970, respectively). The entire sequence of DENV-2 SL strain was not available at that time so the sequence of FJ-

10/China (accession number AF276619) was substituted. This strain belongs to the same "cosmopolitan" genotype (see chapter 3). The three sequences were aligned using the Clustal W program and the alignment was carefully screened in search of variable regions that permit designing strain-specific primers. The criteria used to assure strain-specificity was that each primer must be identical in sequence to the homologous strain but differ by at least three mismatches from the sequences of the other strains with one or more mismatches in or close to the 3'-end. The program Oligo 4.0 (National Biosciences, Plymouth, MN) was used to check candidate primers for appropriate thermodynamic stability and compatibility.

After an exhaustive search, three primer pairs were selected. They were designated by the strain they were specific for and with F or R (forward or reverse) depending on their orientation; for example JAM-F and JAM-R stand for forward and reverse primers, respectively, specific for the Jamaica/83/1409 strain. Primer sequences, aligned to those of the corresponding sequences in the genome of the three strains, are shown in Table 5-1. Most primers fulfilled the proposed criteria. PR-F, was an exception and had only one mismatch with SL; this happened because the sequence used for the design was not SL but FJ-10/China (see above), which had three mismatches with the designed primer. The sequences of the viruses used in these experiments were determined later, as part of the work described in the next chapter. The expected size of the DNA amplified in the RS-PCR varied between 1467 and 2113 base pairs (bp). The amplicons included sequences mostly of prM and E genes, and in some cases of the flanking C and NS1 genes.

Table 5-1. Primers used in recombinant-specific PCR (RS-PCR)

Primer name	Position in genome	Sequences 5′ → 3′ *	Product size**
JAM-F	393-414	Primer CAGGACAGCAGGCGTGATTATT JAM CAGGACAGCAGGCGTGATTATT SL CAGAACTGCAGGCATAATCATC PR TAGAACTGCAGGCATGATCATC	 2108 JAM-R 2113 SL-R 1733 PR-R
JAM-R	2478-2500	Primer GTGCACGTTATCTGTGATGAAG JAM GTGCACGTTATCTGTGATGAAG SL GTGTACATTGTCTGTGATAAAA PR ATGCACGTTATCTGTGACGAAT	 2108 JAM-F 1841 SL-F 2090 PR-F
SL-F	660-682	Primer GTAACTTATGGGACATGCACTG SL GTAACTTATGGGACATGCACTG JAM GTAACTTATGGGACATGTGCCA PR GTAACTTATGGGACATGTACCA	 1846 SL-R 1841 JAM-R 1467 PR-R
SL-R	2481-2505	Primer CCATGTGTGTACATTGTCTGTGATA SL CCATGTGTGTACATTGTCTGTGATA JAM CCATGTGTGCACGTTATCTGTGATG PR CCATGTATGCACGTTATCTGTGACG	 1846 SL-F 2108 JAM-F 2096 PR-F
PR-F	411-432	Primer CATCATGCTGATTCCAACAGTG PR CATCATGCTGATTCCAACAGTG JAM TATTATGTTGATTCCAACAGCG SL CATCATGCTGATTCCAACAGCG	 1715 PR-R 2090 JAM-R 2096 SL-R
PR-R	2103-2125	Primer AACTTCCTTTCTTGAACCAGTCC PR AACTTCCTTTCTTGAACCAGTCC JAM AACTTCCTTTCTTAAACCAGTTG SL AGCTTCCTTTCTTAAACCAGCTG	 1715 PR-F 1733 JAM-F 1467 SL-F

* Red letters correspond to mismatches between the primer and the indicated genome
** Expected product sizes (in base pairs) when used with the indicated primer

Recombinant-Specific PCR. RS-PCR assays were performed in volumes of 50 μL in thin-walled 0.6 μL tubes. The final concentrations of reagents in the PCR mixtures were as follows: 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.5 μM each primer, 0.05 units of *Taq* polymerase (Promega, Madison, WI) and 1X polymerase buffer. Two μL per reaction of template cDNA derived from the coinfected cultures were added. In order to control for recombination that could occur during the PCR (Bradley, 1997), reactions with one μL of cDNA from the two single-virus infected cultures were tested. In addition, control reactions with one μL of a single cDNA were also performed.

Unless otherwise specified, the thermal cycler program used for RS-PCR assays was: 90 °C during one min., 80 °C for no more than 2 min. to add the *Taq* polymerse (hot start), 35-45 cycles of denaturation at 94°C for 10 s., annealing at 55 °C for 30 s. and extension at 72 °C for 4 min., and a final extension of 7 min. at 72 °C. The product of the reaction was held at 4 °C until tested and then frozen at –20 °C. Five μL of the product were assayed by electrophoresis in 1% agarose with ethidium bromide.

In some assays, a mixture of *Taq* polymerase and *Tgo* polymerase (Expand Long Template PCR System, Roche Diagnostics, Manheim, Germany) was used instead of Promega's *Taq*. *Tgo* is a thermostable DNA polymerase from *Thermococcus gorgonarius* that has a 3'-5' exonuclease (proofreading) activity. Used in combination with *Taq* polymerase, it increases 3-fold the fidelity of DNA synthesis and considerable improves the yield of amplified product (Barnes, 1994, Cheng et al., 1994). In these cases the extension steps were conducted at 68 °C.

PCR product cloning and sequencing. Since a mixture of DNA molecules of diverse sequence was expected in mixed infections, the products of successful PCR reactions were cloned in a plasmid vector. In some cases, the product of the RS-PCR reaction was used in the cloning reaction, but in others, the product was first purified in 0.8% agarose gels stained with either ethidium bromide or 1.2 μg/ml of crystal violet (see results).

The cloning reaction was done with plasmid pCR 2.1-TOPO (TOPO TA cloning kit, Invitrogen, Carlsbad, CA) following manufacturer's instructions. Briefly, 2-4 μL of the PCR reaction were mixed with 1 μL of the activated vector and incubated for 5

minutes at room temperature. Two μL of the product were mixed with chemically competent cells, incubated on ice for 30 min., heat-shocked at 42°C for 30 s, iced, mixed with SOC medium and agitated at 37°C for 30 min. 20-100 μL of this transformation were spread in X-gal-treated LB plates containing 50 μg/ml of ampicillin and incubated overnight at 37°C. White colonies were picked and inoculated in tubes with liquid LB medium with ampicillin and incubated overnight.

Plasmid DNA was extracted from pelleted cells using conventional alkaline lysis minipreps (Sambrook, 1989). To verify the presence of an insert of the expected size, an aliquot of the extracted plasmid DNA was digested with EcoR1 endonuclease, and the product was purified by electrophoresis in 0.8% agarose. Selected plasmid DNAs were then purified in silica gel columns (QIAprep kit, QIAGEN, Valencia, CA) and quantified spectrophotometrically. Inserts were sequenced with internal primers specific for prM, E and NS1 genes of DENV-2 (Table 4-2). Additionally, M13 forward and reverse primers, which have annealing regions at both sides of the insertion site, were used to sequence the extremes of the insert to verify the sequence of the primers used in the RS-PCR. Sequences were determined using the Big Dye Terminator Cycle Sequencing System (ABI, Foster City, CA).

Sequence analysis. Sequences from the cloned RS-PCR products were aligned with those of the parental viruses and visually inspected in search of crossovers. Additionally, similarity plots of the involved sequences were produced with Stuart Ray's Simplot program (Lole et al., 1999) sliding a window of 100 nt. in steps of 10 nt. Breakpoints were described as the interval between the two informative (variable) positions that seems to be derived from different parents.

RESULTS

A test of the specificity of strain-specific primers was first conducted using templates derived from single virus-infected cultures. The results are shown in Figure 5-3. Each pair of primers amplified the cDNA of the homologous strain producing an amplicon of the expected size for all three sequences. No product was observed in the reactions with heterologous cDNAs and primers except for the reaction that contained PR-F/PR-R primers and SL cDNA, which yielded a band of the expected size for that primer pair. This could be due to the minimal difference (single mismatch) between the sequences of primer PR-F and SL genome (Table 5-1). Primers were specific enough for most strains, with the exception of the primer pair PR-F/PR-R for SL. For this reason, in the remainder of the study recombination between SL and PR was not investigated.

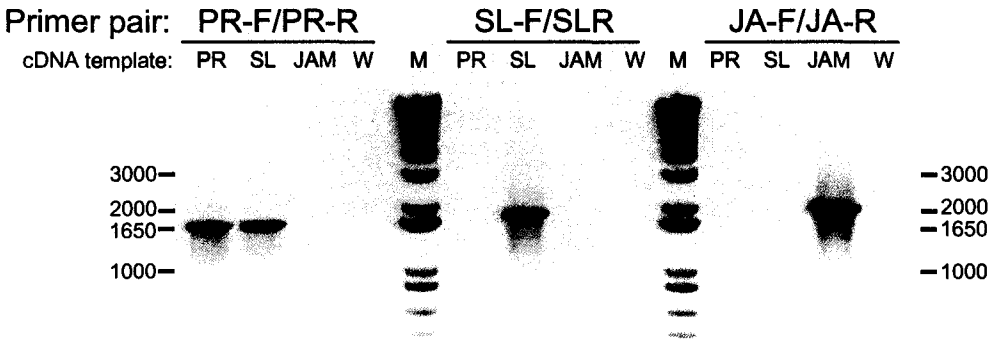


Figure 5-3. Testing of the specificity of the "strain-specific" primers. DENV-2 strains (and corresponding cDNAs) are designated as follows: JAM, Jamaica/83/1409; PR, Puerto Rico/69/159; SL, Sri Lanka/85/1592. JAM-F and JAM-R stand for the forward and reverse primers specific for JAM. Likewise, PR-F, PR-R, SL-F and SL-R are forward reverse primers for PR and SL, respectively; M, molecular weight marker (1kb plus DNA ladder, Invitrogen-Life technologies); W, water (negative control).

RS-PCR reactions were then set up for the cDNAs derived from the crosses of PR with JAM and SL with JAM, hereafter designated PxJ and SxJ, respectively. Reactions with mixtures of two cDNAs derived from different single virus-infected cultures were included to control for recombination during the RS-PCR. To distinguish these cDNA mixtures from reactions in which a single cDNA from a coinfection was used, they were designated P+J and S+J, stranding for the mixtures of PR with JAM cDNAs and the mixture of SL with JAM cDNAs, respectively. Control reactions in which a single cDNA derived from single virus-infected cultures or in which cDNA was replaced by water were also included. All possible combinations of forward and reverse primers for the two viruses in each cross were tested.

The results of the RS-PCR for the coinfections of PR with JAM and SL with JAM are shown in Figures 5-4 A and 5-4 B, respectively. For the first coinfections a faint band of the expected size is visible in the reactions with PxJ cDNA for both primer combinations, PR-F/JAM-R and JAM-F/PR-R. Strong bands of the right size are observed in reactions with combinations of homologous primers (PR-F/PR-R and JAM-F/JAM-R). No bands were observed in the control reactions for which an amplified product was not expected. No product was observed when mixtures of different cDNAs and heterologous primers were included, suggesting that during RS-PCR recombination did not happen. For the coinfections of SL and JAM similar results were observed except that no band was present for the reaction of SxJ cDNA with JAM-F/SL-R primers (Figure 5-4B).

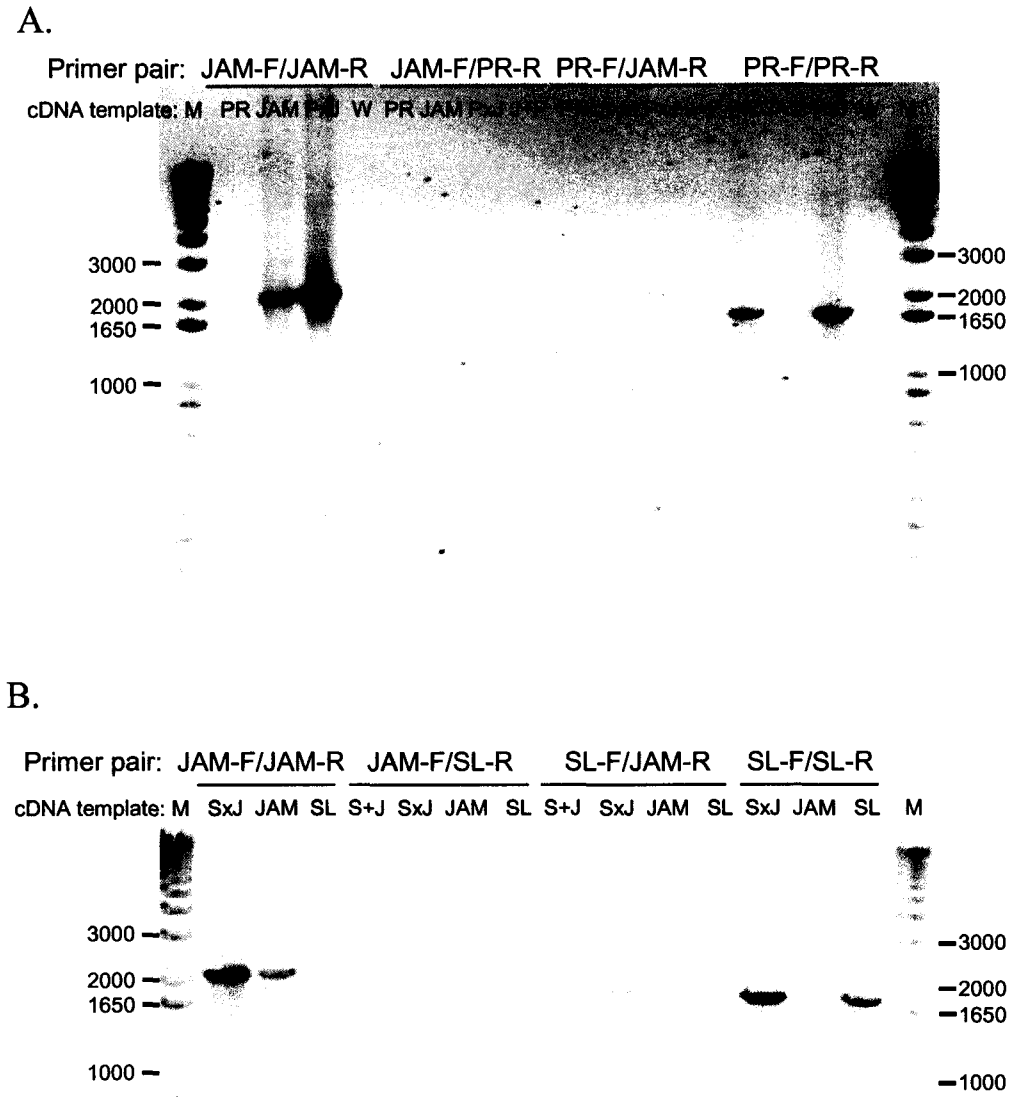


Figure 5-4. Results of the RS-PCR in the first round of experiments. DENV-2 strains (and corresponding cDNAs) are designated as follows: JAM, Jamaica/83/1409; PR, Puerto Rico/69/159, SL, Sri Lanka/85/1592. PxJ and SxJ are cDNA from crosses of PR with JAM (**A.**) and SL with JAM (**B.**), respectively. P+J and S+J are mixtures of cDNAs from PR with JAM and SL with JAM single virus-infected cultures, respectively; M, molecular weight marker (1kb plus DNA ladder, Invitrogen-Life technologies); W, water (negative control).

These results were regarded as satisfactory in spite of the small concentration of recombinant amplicons present in the reactions. The assays were repeated several times with variations in the RS-PCR protocol. Increasing the number of cycles from 35 to 40 or 45 increased the amount of PCR product, even in the reaction of SxJ cDNA with JAM-F/SL-R primers, for which no band was observed in the first try (data not shown). Thus, we proceeded with the cloning and sequencing steps of the experimental design.

Few colonies were recovered in the first tries of cloning the products of fresh RS-PCR reactions. Many of the clones contained shorter than expected inserts (data not shown). Purification of the PCR products in ethidium bromide stained agarose before cloning was then attempted. This was done to reduce the cloning of short molecules are insert into the plasmid more efficiently than larger ones (Zhou & Gomez-Sanchez, 2000). The cloning efficiency remained low. Only ten and seven clones containing the correct sized insert from PxJ and SxJ crosses, respectively, were obtained and sequenced. Two and one of these inserts, respectively, exhibited a sequence almost identical to the parent virus; in each of these one end was identical to the sequence of the heterologous primer. These cases were considered to be probable products of mispriming. The remainder inserts were indeed recombinants. The corresponding similarity plots of these sequences are shown in Figures 5-5 and 5-6. In each case, it can be seen that the sequence of the cloned RS-PCR product is 100% identical to one of the parents in one of its ends, and to the other parent in the other end.

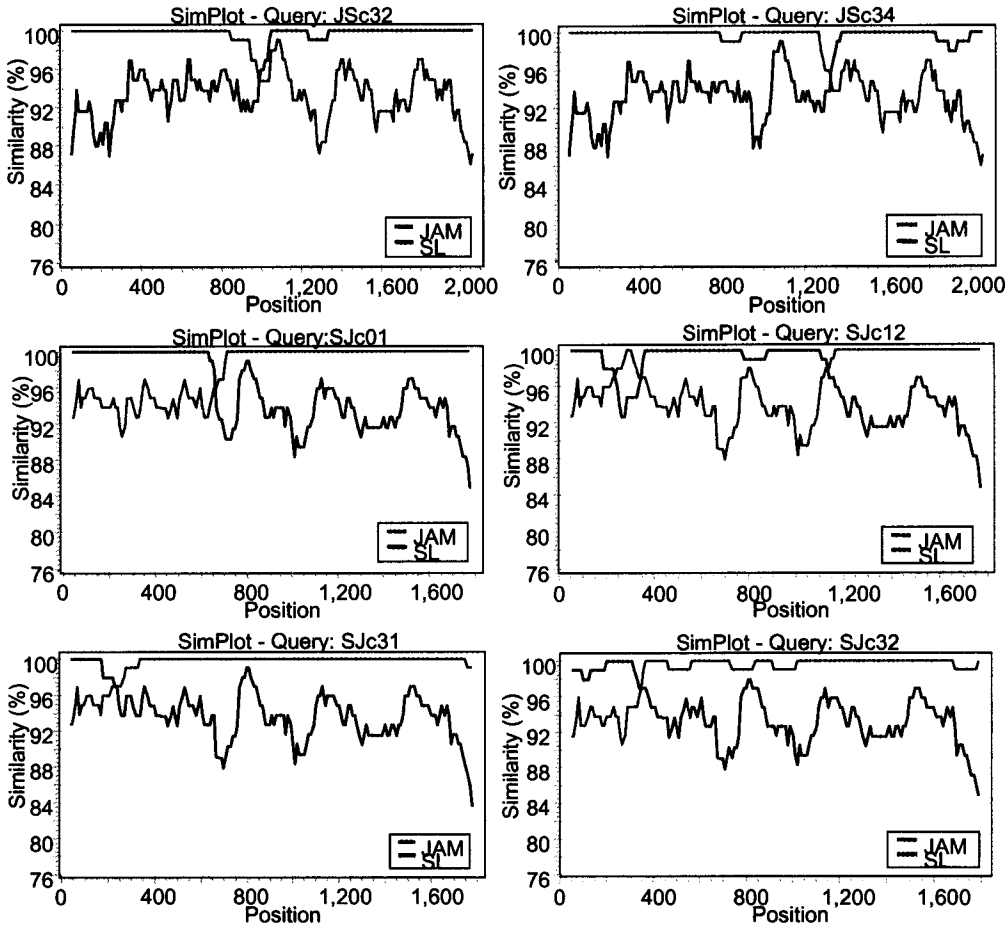


Figure 5-5. Similarity plots of the cloned sequences of RS-PCR products of the cross between DENV-2s strains Sri Lanka/85/1592 (SL) and Jamaica/83/1409 (JAM). Red and green lines represent the % similarity between the sequences of the insert and the parental JAM and SL viruses, respectively, measured in a window of 100 nucleotides (nt) moved in steps of 10 nt. The analysis was performed using SimPlot 3.0.

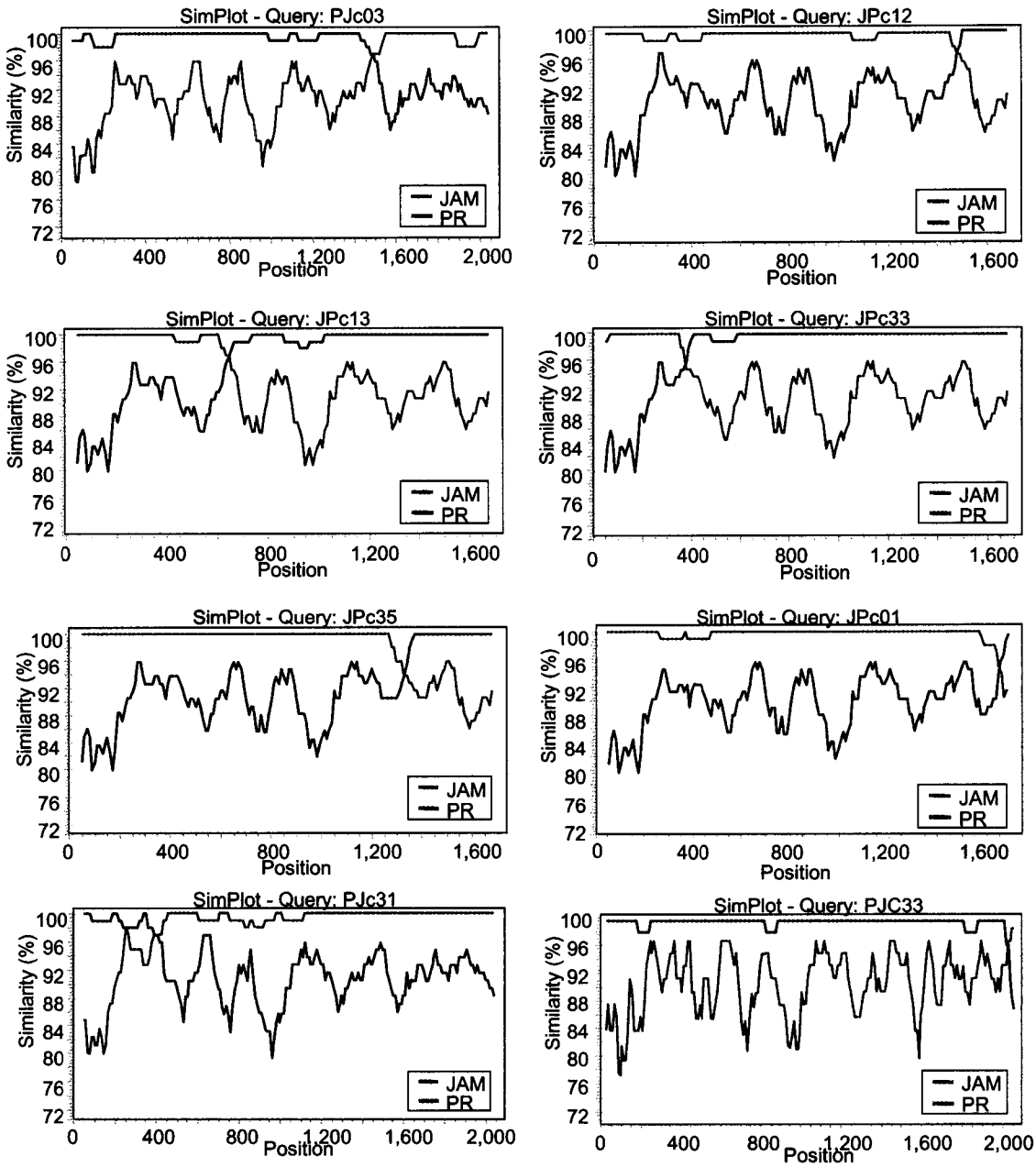


Figure 5-6. Similarity plots of the cloned sequences of RS-PCR products of the cross between DENV-2s strains Puerto Rico/69/1592 (PR) and Jamaica/83/1409 (JAM). Blue and red lines represent the % similarity between the sequences of the insert and the parental PR and JAM viruses, respectively, measured in a window of 100 nucleotides (nt) moved in steps of 10 nt. The analysis was performed using SimPlot 3.0 (Ray, 2003).

Several other findings are noteworthy. The similarity plots are different in all the cases, which indicates that no single recombinant sequence was cloned more than one time. Two sequences, SJc12 (Figure 5-5) and PJc31 (Figure 5-6), exhibit not one, but three breakpoints. The first of these crossovers in SJc12 coincides in location with the only breakpoint in clone SJc31. No other breakpoints are identically located and their positions are widespread along the entire sequence, suggesting that no hotspots for recombination exist in this system. For a more precise location of the crossovers, the sequences of the recombinant molecules are presented schematically Figure 5-7. The presence of a several point substitutions was observed in most inserts. They could have appeared either during the propagation of the viruses in culture or, more likely, could have been introduced by the *Taq* polymerase during the RS-PCR.

It was subsequently realized that the way in which the controls were set up in the previous experiment was incorrect. They controlled for recombination during the RS-PCR but not during the reverse transcription. As explained in the introduction, RT can switch templates during cDNA synthesis (Hwang et al., 2001, Negroni & Buc, 2001). It was conceivable that the RT could have jumped from one to the other genomic RNAs of the different virus strains. Since the controls were set up with mixtures of two cDNAs that were synthesized independently from different RNAs, they would not have detected possible template-switching events occurring during the reverse transcription. Thus, the experiments were started over.

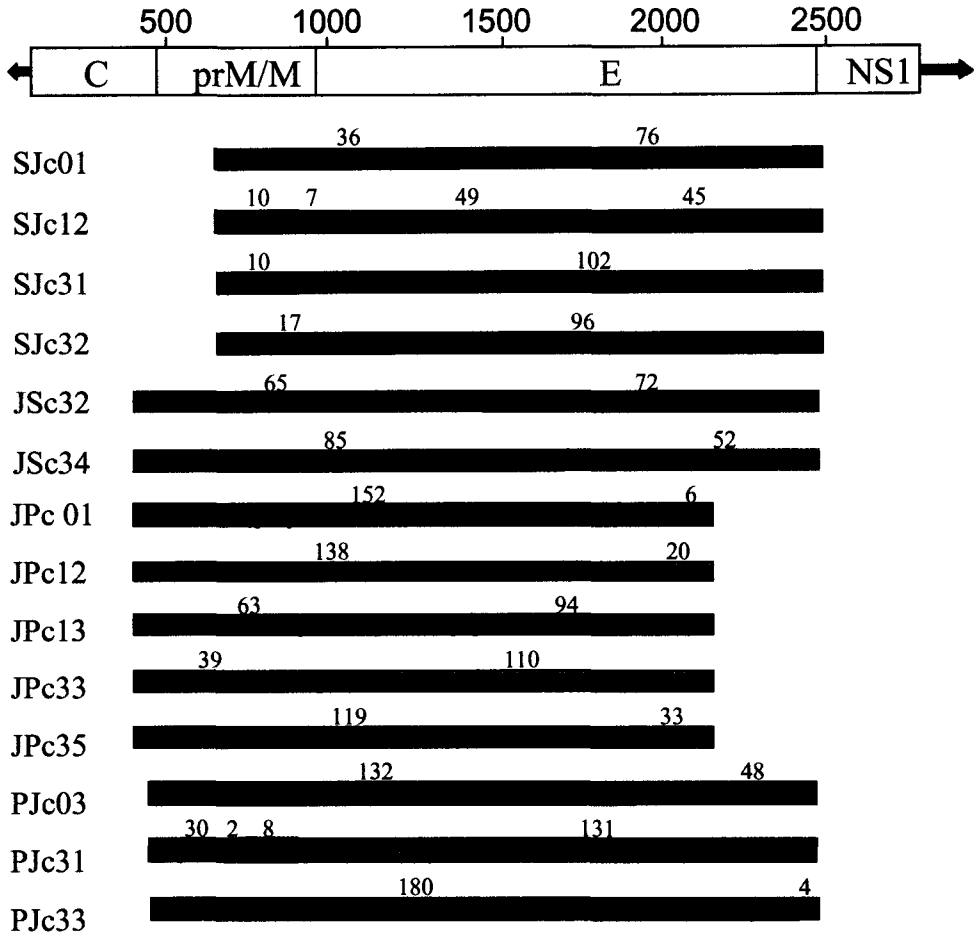


Figure 5-7. Schematic representation of the breakpoint location and parental origin of segments in the sequence of cloned RS-PCR products. Red, blue and green represent segments derived from JAM, PR and SL, respectively. Numbers above the bars indicate the number of informative sites supporting that parental origin of each segment. Vertical lines indicate the presence of point substitutions that were not regarded as additional crossovers.

A second round of experiments was initiated with new double and single infections in C6/36 cells and synthesis of fresh cDNA using RNA from the crosses PxJ, SxJ and, in parallel, with mixtures of RNAs from single virus-infected cultures. The latter were designated P-J or S-J after the virus RNAs used in their synthesis, PR and JAM or SL and JAM, respectively.

These cDNAs were used in new RS-PCR reactions that were similar to the previous ones with some modifications. The most important was the substitution of Promega's *Taq* for the mixture of *Taq* and *Tgo* DNA polymerases (Roche Diagnostic's Expand Long Template System). This modification was done to increase the concentration of amplified DNA was necessary to improve the cloning efficiency and to reduce the number of point mutations introduced by Taq polymerase. These point substitutions could force the polymerase to pause and eventually detach from the template during the DNA extension (Barnes, 1994). This could generate an incomplete strand that could eventually anneal to a different (but similar) template in the following PCR cycle, generating an artifactual recombinant.

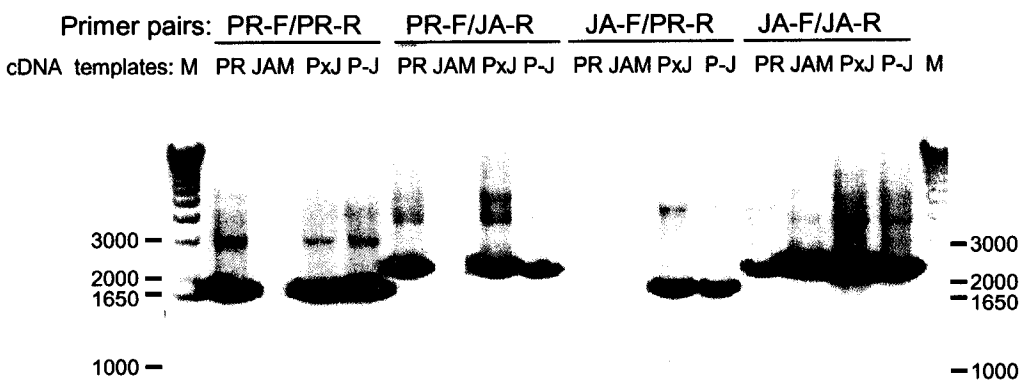
Another modification was that the products of the RS-PCR were purified by electrophoresis in agarose gels stained with crystal violet rather than with ethidium bromide. This was done to avoid damage in the PCR product during the visualization of the bands. UV light introduces nicks in DNA molecules that considerable reduce the cloning efficiency (Zhou & Gomez-Sanchez, 2000).

In this second round of assays with modifications, it was evident that some of the previous difficulties were solved but other problems arose. The RS-PCR yielded now strong bands in the PxJ and SxJ reactions. At the same time, strong bands were observed in control reactions, particularly in those with P-J cDNA (Figure 5-8A). The products of control reactions in which PR cDNA was combined with PR-F/JA-R and JA-F/JA-R primers, exhibited clear bands indicating that the reaction had lost its specificity. The RS-PCR for the JAM and SL coinfection also exhibited some bands in control reactions (data not shown).

A new test of specificity of the primers for the three DENV-2 strains was done using the *Taq*/*Tgo* mixture. This time nonspecific bands were observed demonstrating amplification of cDNAs of single virus-infected cultures with primer pairs supposed to be specific for different strains (data not shown).

To investigate these unexpected results, the RS-PCR products of P-J and PxR cDNAs amplified with primers PR-F/JAM-R were cloned and sequenced. This time high cloning efficiency was observed. The inserts in three clones of each (P-J or PxJ) were sequenced. All of these inserts were identical to the sequence of PR in all its length, except in the nucleotides corresponding to the primer JAM-R. However, the nucleotide corresponding to the 3'-end of that primer had been change from G to T, reducing to two the number of mismatches between the primer and the sequence of PR. It was later realized that polymerases with 3'-5' exonuclease activity can edit mismatches at the 3'-end of the primer if the remainder of its sequence exhibits considerable complementarity to the template. Since the specificity of the primers utilized in this study relied on differences at the 3'-end, it was concluded that proofreading polymerases could not be used in the RS-PCR any longer.

A.



B.

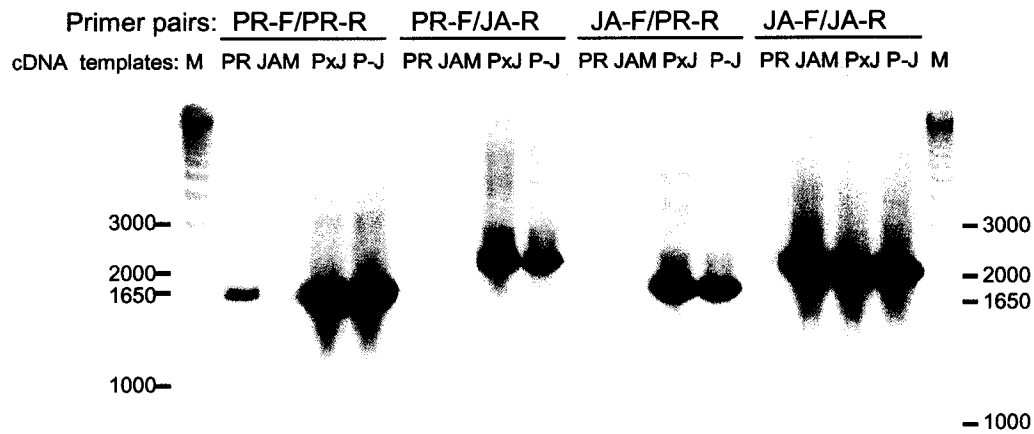


Figure 5-8. Results of RS-PCRS in the second round of experiments using (**A.**) *Taq/Tgo* mixture (Roche, see text for details) or (**B.**) *Taq* (Promega). DENV-2 strains (and corresponding cDNAs) are designated as follows: JAM, Jamaica/83/1409; PR, Puerto Rico/69/159. PxJ is the cDNA from the cross of PR with JAM. P-J is the cDNAs synthesized with mixtures of RNAs from PR and JAM single virus-infected cultures. M, molecular weight marker.

The experiments were resumed using the original protocol with *Taq* Promega. The RS-PCR using this enzyme yielded fewer nonspecific bands in comparison to that produced by the *Taq*/*Tgo* mixture in the same templates, but reactions of P-J cDNA with both, PR-F/JA-R and JA-F/PR-R primers, still produced strong bands (Figure 5-8 B). Sequencing of uncloned RS-PCR products demonstrated that their sequences were indeed recombinants between JAM and PR sequences (data not shown). Since the cDNA of these reactions was synthesized with RNAs extracted from single virus-infected cultures, recombination probably happened during either the reverse transcription, RS-PCR or both. None of the controls performed in the last experiments permitted differentiation between these three options, but absence of bands in the RS-PCR in the first round of experiments suggested that the recombination might have been generated during cDNA synthesis.

At this point new experiments were designed to test for and to prevent recombination during the reverse transcription. Since it have been demonstrated that recombination mediated by RT is temperature dependent (Li & Zhang, 2001), a new cDNA syntheses were performed with SuperScript™ II RT at temperatures ranging from 42 to 52°C. The new RS-PCR reactions were set up for combinations of heterologous primers. This time many cDNAs (PxJ, P-J and P+J), were included in order to determine in which step the observed recombination was happening. The results are in Figure 5-9.

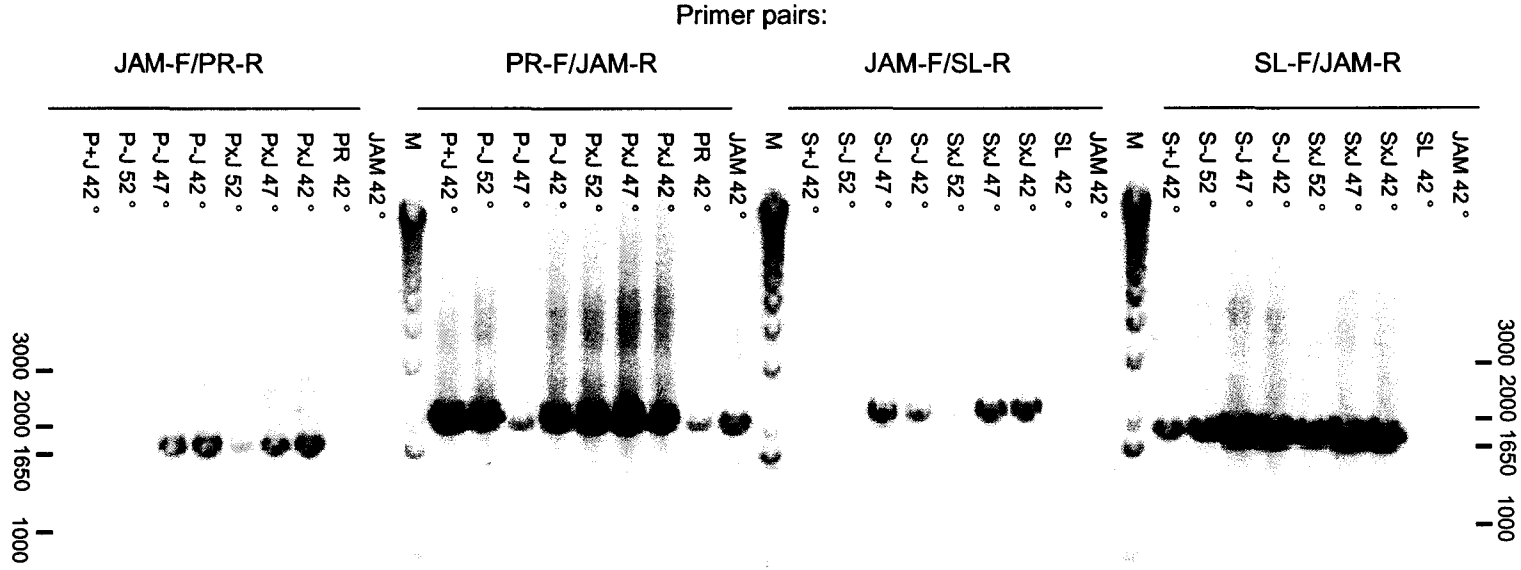


Figure 5-9. Results of RS-PCRs using cDNAs synthesized with SuperScript II RT (Invitrogen, Life technoogies) at different temperatures. DENV-2 strains (and corresponding cDNAs) are designated as follows: JAM, Jamaica/83/1409; PR, Puerto Rico/69/159; SL, Sri Lanka/85/1592; PxJ and SxJ are cDNA from coinfections of PR with JAM and SL with JAM respectively. P-J and S-J are cDNAs synthesized with mixtures of RNAs from PR with JAM and SL with JAM single virus-infected cultures; P+J and S+J are mixtures of cDNAs from PR with JAM and SL with JAM single virus-infected cultures, respectively; M, molecular-weight marker

Reactions with primer mixture PR-F/JAM-R exhibited bands in all the reactions including controls with PR and JAM cDNAs, suggesting nonspecificity of the primers. The other three primer mixtures (JAM-F/PR-R, JAM-F/SL-R and SL-F/JAM-R) did not produce any band in these single strain controls and could be analyzed. Reactions with primer pairs JAM-F/PR-R and JAM-F/SL-R yielded no nonspecific bands in controls with PR, SL, JAM, P+J and S+J cDNAs, but bands of similar sizes were observed in reactions using PxJ and P-J, and also SxJ and S-J cDNAs synthesized at 42 and 47°C. However, with cDNAs produced at 52°C faint bands were observed for PxJ and SxJ but neither for P-J nor S-J. These results strongly suggest that recombination was indeed taking place during cDNA synthesis at 42 and 47 °C. The results with cDNAs produced at 52°C could be interpreted in two ways: either reverse transcription was very inefficient or recombination was inhibited at that temperature. The latter interpretation let open the possibility that the faint RS-PCR bands observed with cDNAs PxJ and SxJ at 52°C were the results of recombination produced during virus replication in cell culture.

To test this possibility a final experiment was set up. This time new cDNAs were synthesized at 42, 52 and 62°C, a wider range of temperatures. Theoretically this would reveal whether or not the increase in temperature could be used as a method of inhibiting recombination during reverse transcription. Since SuperScript™ II RT had reduced activity at temperatures greater than 50°C, a new thermostable RT, ThermoScript™ RNase H^- RT, was used. This latter enzyme is active up to 70 °C (Invitrogen-Life Technologies Catalog 2004). The results of this last experiment for the cross between SL and JAM strains are shown in Figure 5-10.

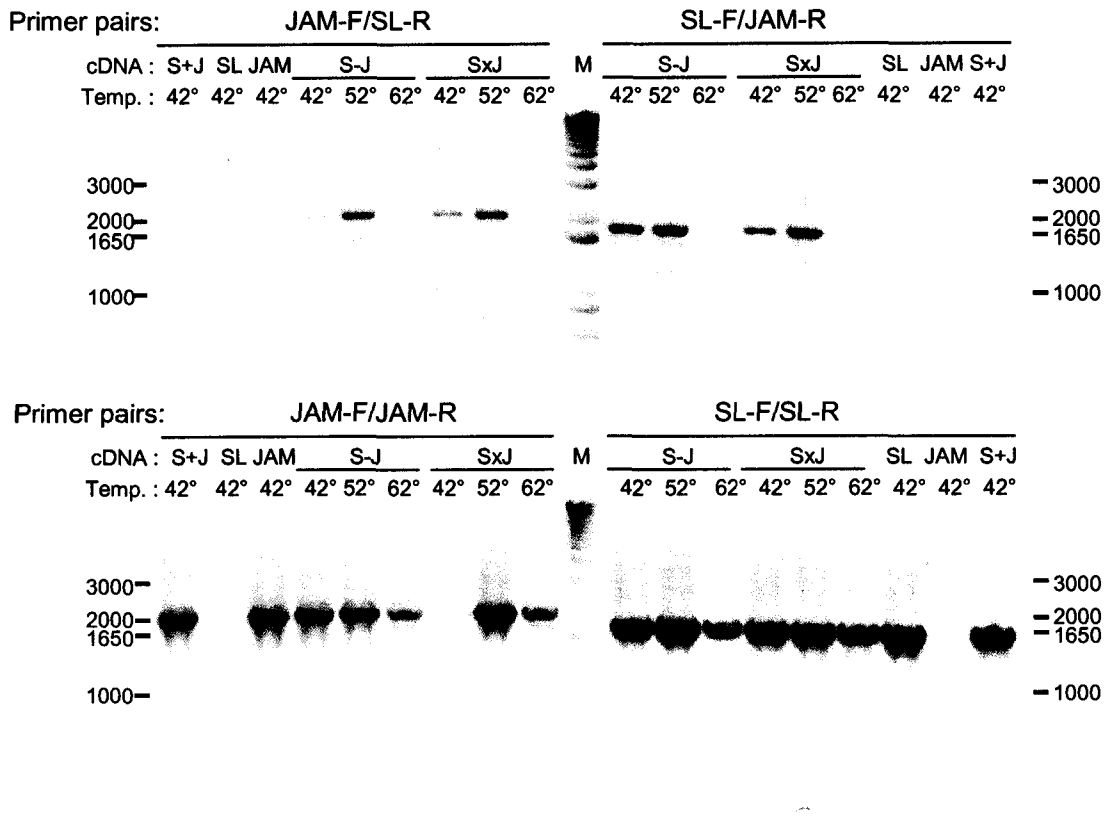


Figure 5-10. Result of one RS-PCR of the last round of experiments using cDNAs synthesized with ThermoScript RT (Invitrogen-Life technologies) at different temperatures. DENV-2 strains (and corresponding cDNAs) are denominated as follows: JAM, Jamaica/83/1409; SL, Sri Lanka/85/1592. SxJ is the cDNA from the coinfection SL with JAM; S-J is the cDNA synthsized with mixtures of RNAs from SL and JAM single virus-infected cultures; S+J are mixtures of cDNAs from SL and JAM single virus-infected cultures; temp., is the temperature of synthesis of each cDNA in centigrade; M, molecular weight marker.

The results for the coinfection between PR and JAM were similar and are not shown. Control reactions using homologous primers pairs (JAM-F/JAM-R and SL-F/SL-R) are shown in the lower panel of the figure. They demonstrate that cDNA synthesis could be achieved at all the used temperatures although some reduction in the product was observed at 62°C in most cases. Test reactions using heterologous primers (JAM-F/SL-R and SL-F/JAM-R) showed the presence of recombinant sequences in the reactions performed at 42 and 52 °C but not at 62 °C. (Figure 5-10, upper panel). The reaction was maximal at 52°C and no difference between S-J and SxJ products was observed. The absence of recombinant products for SxJ cDNA (at 62 °C) and for S+J cDNA (at 42 °C) suggested that no or very little recombination was happening during virus replication in cell culture and RS-PCR, respectively, and that most (if not all) of the recombinant products must have arisen during the reverse transcription.

DISCUSSION

Homologous recombination have been experimentally demonstrated for several (+) RNA viruses and retroviruses (Lai, 1992, Worobey & Holmes, 1999). In many experiments, mutant viruses with selective traits have been used. These include neutralization escape, antiviral resistant, host range and temperature-sensitive mutants. In most cases, these phenotypic traits can be traced to single amino acid substitutions. Using two mutant strains with, for example, conditional lethal mutations that mapped to separate regions in the genome it is possible to test for recombination. Viruses are propagate first at permissive conditions, allowing the polymerase to switch templates, and later at conditions that are restrictive for both parental strains selects for recombinants with crossovers between the locations of the substitutions that eliminated

these mutations. This has been the preferred method used with coronavirus, picornavirus and togavirus (Baric et al., 1990, Kirkegaard & Baltimore, 1986, Weiss & Schlesinger, 1991). For some other viruses that recombine at a high rate, no selective step is required and direct testing of plaque purified clones or cDNA clones could be used (Nagy & Bujarski, 1995).

The polymerase chain reaction (PCR) provides the possibility of detecting recombinant molecules in a mixture of recombinant and nonrecombinant DNAs (Figure 5-2). With mixtures of primers specific for sequences present in different genomes, PCR can be used to detect both, homologous or heterologous recombinants between those genomes. That capability has been extensively used to detect insertions of specific DNA sequences in plasmids, as part of conventional molecular biology technique (Sambrook, 1989). The use of PCR to detect recombination between genomes of related RNA viruses coinfecting a single experimental host has been reported in a few cases (Li et al., 1999, van Vugt et al., 2001).

In this chapter, the steps performed in the design, construction and testing of a PCR-based system to detect homologous recombination between divergent strains of a single virus, DENV-2, was presented. The PCR, here designated recombinant specific PCR (RS-PCR) was able to preferentially amplify recombinant molecules in a mixture of predominantly nonrecombinant genomes. However this RS-PCR was not problem-free. Some of the individual amplified molecules seemed to be produced by mispriming rather than by actual recombination; this was a major problem when the mixture of *Taq* and *Tgo* DNA polymerases was used (Figure 5-8 A). This polymerase-mixture approach has proven to be a powerful means of improving the yield and fidelity of PCR, especially when long products are expected (Cheng et al., 1994). The

possibility of false positive results induced by the 3'-end editing of the primer by proofreading polymerases must be kept in mind whenever a mixture of related templates are present in the sample.

A number of studies have demonstrated that recombination can happen during PCR (Bradley, 1997, Judo et al., 1998, Odelberg et al., 1995, Shammas et al., 2001). It seems to be due mostly to incomplete extension of strands that can later prime on a different template. However, template-switching has also been reported (Odelberg et al., 1995, Shammas et al., 2001). Some simple measures like increased extension time, reduced number of cycles and addition of some substances that suppress pauses of polymerase have been advocated to reduce this problem (Judo et al., 1998, Shammas et al., 2001).

However, most recombinant sequences detected here seem to have arisen during reverse transcription and not in the PCR. No bands where observed whenever a mixture of two separately synthesized cDNAs were included in a RS-PCR reaction; the only exception was a reaction S+J with primers SL-F/JAM-R (Figure 5-9). Conversely, bands regularly resulted after amplification of a single cDNAs produced with mixed RNAs (lanes P-J and S-J in Figures 5-8B, 5-9 and 5-10).

Recombination during reverse transcription has been reported many times and seems to be a normal step in retrovirus life cycle (Katz & Skalka, 1994). Both, cis-acting elements and RT processivity seems to be major determinants of the frequency of recombination (Negroni & Buc, 2001). A necessary requirement is the degradation of the RNA template by RNase-H domain of RT (Hwang et al., 2001, Negroni et al., 1995). The two RTs used in this study, SuperScript™ II RT and Thermoscript™ RT,

have a mutated RNase domain (RNase$^-$), which is supposed to greatly reduce RNA degradation (Invitrogen-Life Technologies, Catalogue 2004). However, RNA degradation seems to occur in enough extent to allow template-switching in the system used here, as demonstrated by regular production of recombinants.

Another important factor in recombination during reverse transcription is temperature. Li and Zhang (2001) demonstrated that recombination frequency is steadily reduced when the temperature is increased from 31 °C to 43 °C. This factor was exploited here to suppress recombination during cDNA synthesis using temperatures greater than 42°C. A reduction in the intensity of RS-PCR band was generally observed when cDNAs produced with SuperScript^{TM} II RT at progressively higher temperatures were used (lanes P-J and S-J, Figure 5-9). Interestingly, when Thermoscript^{TM} RT was used, accumulation of recombinant products was greater at 52°C, moderate at 42°C and low or absent at 62°C (Figure 5-10). This could be a particular effect of temperature in the frequency of recombination or, alternatively, an artifact dependent on different polymerizing efficiencies of the enzyme at different temperatures. A time-resolved quantitative RS-PCR system could clarify this issue.

Overall, no clear evidence of recombination during the replication of DENV-2 in cell culture was obtained in this study. The similar band intensity between PxJ and P-J lanes (Figures 5-8B. and 5-9) and between SxJ and S-J lanes (Figures 5-9 and 5-10), suggests that all the recombination observed was produced during cDNA synthesis. This conclusion is supported by the absence of amplification of SxJ cDNAs generated at 62 °C, a temperature at which reverse transcription still happened (Figure 5-10).

This is not to deny that recombination between different DENV strains could actually occur. Most of the work performed here was devoted to improve the system to detect recombinants, and not to test for recombination under different conditions. It is not know, for example, if recombination is equally likely (or unlikely) in vertebrate and insect hosts or what degree of homology is required for DENV recombination to occur. Thus, more experimental work must be done before conclusions can be drawn about the feasibility, frequency and factors conditioning recombination in DENV.

CHAPTER 6

EVOLUTIONARY FORCES SHAPING THE ADAPTATION OF AN ARBOVIRUS TO VERTEBRATE AND/OR MOSQUITO CELLS

RNA viruses possess enormous evolutionary potential and have the fastest evolving genomes in nature with rates of about 10^{-3} nucleotide substitutions per site per year (s/s/y), about six orders of magnitude greater than eukaryotic organisms (Holmes, 2003, Steinhauer & Holland, 1987). This rapid evolutionary pace is seen in viruses like influenza A that cause acute infections and pass rapidly from one individual to the other (Buonagurio et al., 1986, Gorman et al., 1992), as well as in viruses that establish chronic infections like HIV during the course of a single-host infection (Wain-Hobson, 1993). These two viruses can escape herd or individual immunity, respectively, because of the rapid substitutions of amino acids in neutralizing epitopes on the virion surface proteins (Gorman et al., 1992, Seibert et al., 1995). Rapid evolution has been documented in many other RNA viruses (Buonagurio et al., 1986, Sala & Wain-Hobson, 2000, Steinhauer & Holland, 1987). This unusual evolutionary potential is thought to be due to the lack of fidelity and proofreading activity of viral RNA polymerases (and reverse transcriptases in the case of retroviruses), which allow rates of about one mutation per genome per replication cycle (Drake & Holland, 1999). Two other factors contribute to the rapid evolution of these agents: large population sizes (in the order of 10^{12} organisms per infected host) and the short generation time (usually a few hours) (Holmes, 2003, Novella et al., 1999).

Not all RNA viruses evolve at a rapid rate: Human T-cell lymphotropic viruses 1 and 2 (HTLV-1/2) evolve at ~10^{-6} s/s/y apparently because of longer generation times characteristic of their transmission cycle (Van Dooren et al., 2004). In its natural reservoir, aquatic birds, Influenza A virus evolves at a much slower rate than in humans and other vertebrates, especially at non-synonymous sites, suggesting that this virus has reached an adaptive optimum in the avian host that is maintained by stabilizing selection (Gorman et al., 1992). Among arboviruses, alphaviruses evolve at a rate between 2-7 x 10^{-4} s/s/y, almost a ten-fold lower rate than other RNA viruses (Weaver et al., 1994). La Crosse, a bunyavirus, exhibits a high degree of genetic stability during experimental transmission cycles (Baldridge et al., 1989). Tick-borne flaviviruses evolve at about 10^{-4} s/s/y (Zanotto et al., 1996). Dengue virus (DENV) evolution rate has been estimated between 10^{-4} to 10^{-3} s/s/y, about the same estimated for alphaviruses (Twiddy et al., 2003, Zanotto et al., 1996).

The slower evolutionary rate in arboviruses has been attributed to the necessity of alternation between two disparate hosts, one vertebrate and other arthropod. This double adaptive landscape is thought to impose a high degree of stabilizing selection that constrains the exploration of the sequence space, thereby limiting the fixation rate of new mutations (Weaver, 1995, Weaver et al., 1999). This would force the virus to inhabit the reduced space in which the fitness peaks of the two adaptive landscapes overlap. Two corollaries of this model are: 1) natural populations of arboviruses must have a lower diversity than other RNA viruses, and 2) they should maintain a lower reproductive efficiency (fitness) than single-host viruses due to the double selective constraints. This is a form of balancing selection. The model also predicts that if one of

the hosts is serially bypassed, the substitution rate will be augmented and the virus will gain fitness in the host that is being used and will lose fitness in the bypassed host.

However, a different model could also be proposed. The alternation of hosts would periodically remove the evolutionary constraints particular to each host. This would allow the virus to explore portions of the sequence space that are deleterious in one of the hosts but are neutral or even favorable in the other. This fluctuation in the adaptive landscape would generate a disrupting selection that would increase the diversity of the viral population and improve the chances of finding new fitness peaks, which in turn could accelerate the substitution rate. Indeed, the fact that almost all arboviruses have an RNA genome has been attributed to the necessity of genetic plasticity that allows the rapid adaptation to the changing environment imposed by arbovirus cycles (Beaty et al., 1997). If this is true, another explanation must be sought to understand the lower evolutionary rate of arboviruses.

In 1999 two studies designed to test the first of these two contradictory evolutionary scenarios were published (Novella et al., 1999, Weaver et al., 1999). Either eastern equine encephalitis virus (EEEV) or vesicular stomatitis virus (VSV) was subjected to serial passages in insect cells, in vertebrate cells and in the two cell-systems, alternately. Phenotypic and genotypic adaptations to the different cell types were observed by fitness assessment and partial genome sequencing, respectively. In agreement with the predictions of the first model, the study using EEEV found that fitness declined in the bypassed host, increased in the host used for the adaptation and that the number of substitutions in the alternately passaged virus was reduced. The only finding that did not match the predictions was that alternating passages generated fitness increases similar to those found in single-host passages in both cell-types. In

the study by Novella *et al* (1999), however, very different results were observed: VSV increased its fitness in both the used and the bypassed hosts cells and the alternately passaged viruses accumulated more mutations than those passaged only in the single-host-cell system. The conclusions of the two studies were contradictory, the first supporting the double-constraint hypothesis and the other dismissing it. Thus, the reasons for the lower evolutionary rate in arboviruses and the evolutionary consequences of host alternation are not resolved.

Here, the results of a similar experiment using an unrelated agent, Dengue virus type 2 (DENV-2) are presented. By serially adapting different DENV-2 strains to insect, vertebrate and alternating cell environments, the hypotheses that phenotypic and genotypic changes are host-cell type specific and that host alternation restricts virus evolution were tested. Additionally, amino acid substitutions accumulated in each environment allowed to gain insight into some differences in the life cycle depending on the host in which the virus propagates.

MATERIALS AND METHODS

Viruses, Cells and Serial Passages. Four DENV-2 strains were selected for serial passaging in cell culture: New Guinea C/44 (NGC), Jamaica/83/1409 (JAM), Puerto Rico/69/159 (PR) and Sri Lanka/85/1592 (SL). They have been classified in the Asian-2, American-Asian, American and Cosmopolitan genotypes, respectively (Chapter 3). Most of them had been passed 4-6 times in C6/36 cells. However, NGC, the prototype DENV-2, had been passed at least 35 times in different animals and cells. All of the strains were obtained from the reference collection of the Division of Vector-Borne Infectious Disease, Center for Disease Control and Prevention (CDC), Fort Collins, CO.

Figure 6-1 summarizes the adaptive process and the tests performed. Each strain was subjected to three different adaptive series of twenty-four passages in cell-culture: one in *Ae albopictus* C6/36 cells (C6 series), one in LLC-MK_2 monkey kidney cells (LLC series) and a third one alternating between C6/36 and LLC-MK_2 systems (ALT series). Cells were grown in 25 cm^2 cell culture flasks either at 28°C (C6/36) or at 37°C in 5% CO_2 atmosphere (LLC-MK_2). Media used were either Leibowitz L-15 (C6/36) or 199 (LLC-MK_2), supplemented with 10% fetal bovine serum (FBS). When the monolayers were confluent cells were inoculated with one ml of a 1/10 dilution of the supernatant of the previous passage, and the flask was gently rocked for one hour. The inoculum was removed and replaced with fresh medium with 2% FBS. For the initial passages, the viruses were allowed to replicate for seven days. With the progressive adaptation to the cell culture, the inocula had to be further diluted to 1/100 or 1/1,000 and the duration of the virus replication shortened (see results). The virus suspension from the last passage was used to feed mosquitoes and to prepare viral stocks, which were stored in aliquots at –70°C. Each viral stock obtained during the process was labelled according to strain, passage and treatment. For example, PR24LLC stands for the Puerto Rico/69/159 stock obtained after the 24^{th} passage from LLC-MK_2 cells.

Detection and quantitation of virus. At selected passages, viral replication was verified by an indirect immunofluorescence assay (IFA), using either DENV-2 specific (3H5) or flavivirus group reactive (4G2) monoclonal antibodies, followed by biotinylated anti-mouse antibody and fluorescein-conjugated streptavidin (Amersham, Arlington Heights, IL) as previously described (Bennett et al., 2002). For each passage in each series, presence or absence of cytopathic effect was recorded and an aliquot of the supernatant was frozen at –70°C.

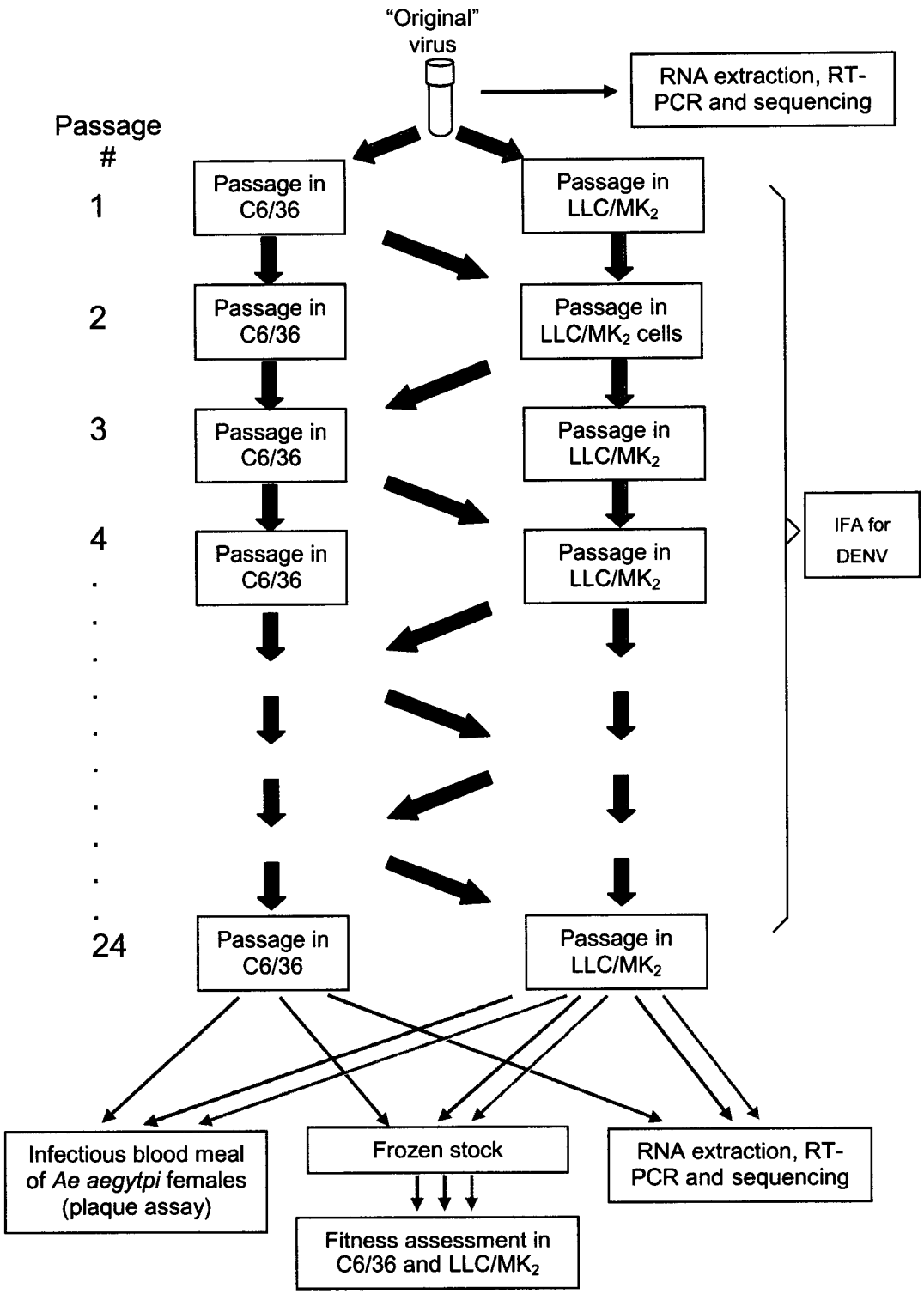


Figure 6-1. Schematic view of the adaptive process and laboratory tests used in this study. Blue, red and green arrows represent the C6, LLC and ALT series, respectively. The four DENV-2 strains were subjected of the same process.

Quantification of viruses in cell-culture supernatants or frozen stocks was performed by conventional plaque assay in monolayers of LLC-MK_2 cells grown in 6 well plates. Samples were serially diluted and 0.1 ml of each 10-fold dilution was allowed to adsorb to cells during one hour. The inoculum was removed and four ml of nutrient agar (0.066% yeast extract, 0.33% lactalbumin hydrolysate, 4% fetal bovine serum, 0.1 mg/ml gentamycin, 0.004 mg/ml amphotericin B and 0.7% purified agar in Earl's balanced salt solution) were added to each well. After seven days of incubation at 37°C with 5% CO_2, two ml of a second overlay of nutrient agar with 0.016% neutral red was added. Viral titers in plaque forming units per ml (pfu/ml) were calculated and recorded two days later along with plaque size.

Fitness assessment. Fitness of passaged and unpassaged viruses was assessed by their relative abilities to replicate in mosquito and vertebrate cells and to orally infect and disseminate in mosquitoes. Passage 24 viral stocks were titrated by plaque assay and used to infect both C6/36 and LLC-MK_2 monolayers at a multiplicity of infection (moi) of 0.005. After two, four and six days of incubation viral titers were determined by plaque assay.

Viral infectivity in mosquitoes was measured as described (Bennett *et al* 2002). Briefly, about 100 fasting *Ae aegypti* females (Rex D strain) were provided with an infectious blood meal composed of equal parts of defibrinated sheep blood and fresh viral culture product (supernatant and cells) of the last passage of each series.

Infectivity of viruses from ALT series was tested at passage 24 (LLC-MK2) and 25 (C6/36). Viruses from LLC series were given an additional passage in C6/36 cells to increase the titer and were tested before and after this additional passage. Viral titer in the blood meal was determined by plaque assay as previously described. After fourteen days of extrinsic incubation mosquito heads were severed, squashed onto glass slides and fixed with acetone. Disseminated DENV-2 infection was determined by IFA as described above.

Molecular analysis of serially passaged viruses. At the beginning and at the end (1^{ST} and 24^{th} passages) of the cell adaptation process viral RNAs were extracted from the cell culture supernatant using a silica gel-based method (QIAamp viral RNA extraction kit, Qiagen, Valencia, CA). The RNAs were reverse transcribed using the D2-3512R anti-sense primer (Table 4-2) and RAV-2 reverse transcriptase (Amersham, Piscataway, NJ). Then, a 3,398 nt long viral genomic sequence encompassing most of C gene and complete prM, E, NS1 genes was amplified by PCR using primers D1 and D2-3512 (Table 4-2) and Platinum® *Taq* DNApol High Fidelity (Invitrogen, Carlsbad, CA). The thermal cycling protocol consisted of ten cycles of 94°C for 10 sec., 55°C for 30 sec. and 72°C for 3 minutes, followed by 30 additional cycles at the same temperatures and times except that the extension was increased 10 seconds per cycle to improve the yield. Product size and uniqueness was verified by 0.8% agarose gel electrophoresis. When additional DNA bands were present the right-sized DNA product was purified in crystal violet-stained 0.8% agarose gels. The PCR products were then

cleaned using the either the QIAquick PCR purification kit or the QIAquick gel extraction kit (Qiagen).

Both strands of the PCR products were sequenced with the primers listed in Table 4-2 using the Big Dye Terminator Cycle Sequencing System (ABI, Foster City, CA). RNA and deduced protein sequences of the four DENV-2 strains obtained after each of the three series of passages were aligned upon codons and compared with their respective sequences at the beginning of the experiment to identify substitutions occurring during the process.

Several methods were used to test for evidence of positive selection during the adaptive process the: Tajima's D and Fu & Li F* are well known statistics that test the null hypothesis of neutral evolution, based in the distribution of polymorphisms in the sequences of an alignment. Yang's nonsynonymous/synonymous rate ratio ($\omega = d_n/d_s$) tests specifically for the presence of directional selection (Yang et al., 2000). These analyses were performed using DnaSP computer software (Rozas et al., 2003).

RESULTS

Phenotypic Changes.

Viral replication and cytopathogenicity. During the initial passages cytopathic effects (CPE) were not detected in any of the series during the seven days of observation. Viral presence was verified by IFA. All DENV-2 strains gradually induced CPE first in C6/36 and the alternating series and later in LLC-MK_2. Since CPE was

appearing earlier in each successive passage, it was necessary to reduce the duration of each passage to three to four days in C6/36 cells and to four to five days in LLC-MK_2 cells. In addition, the virus inocula had to be diluted 1/100 for most series or 1/1,000 for NGC series to avoid the rapid destruction of the monolayers. IFA was not necessary beyond passage twelve.

CPE differed depending on the cell line infected. In C6/36 cells, CPE consisted mainly of syncytia that evolved from focal to widespread. In LLC-MK_2 cells, CPE was characterized by cell pycnosis and clumping, but never progressed to involve all the monolayer. This was in agreement with the IFA analysis, which always demonstrated virus in less than 50% of LLC-MK_2 cells but usually in 100% of the C6/36 cells.

CPE also varied with the viral strain. NGC, the prototype DENV-2, generated CPE earlier than the other viruses in all series and the morphological changes were more severe. In LLC-MK_2 cells, NGC produced considerable detachment of cells, an effect not observed with the other strains. In C6/36 cells, NGC produced rapidly-growing poorly-delimited syncytia. In other strains syncytia progressed more slowly and were better defined. The JAM strain did not replicate well in LLC-MK_2 cells, and the virus was lost two times during the initial passages. It was necessary to return to an earlier passage to continue the series. The JAM strain produced very little or no CPE at all in this cell line. However in mosquito cells it was as cytopathic as the other strains.

Plaque morphology. Marked differences were observed in the size of the plaques depending on the series (Figure 6-2 and Table 6-1). Plaques produced by viruses passaged in C6/36 cells were smaller in size than those of the original viral stocks (Figure 1b). Viruses serially passaged in LLC-MK_2 cells usually generated large plaques; however JAM produced medium size ones (Figure 1c). Viruses resulting from the alternating passage series produced variably sized plaques (Figure 1d).

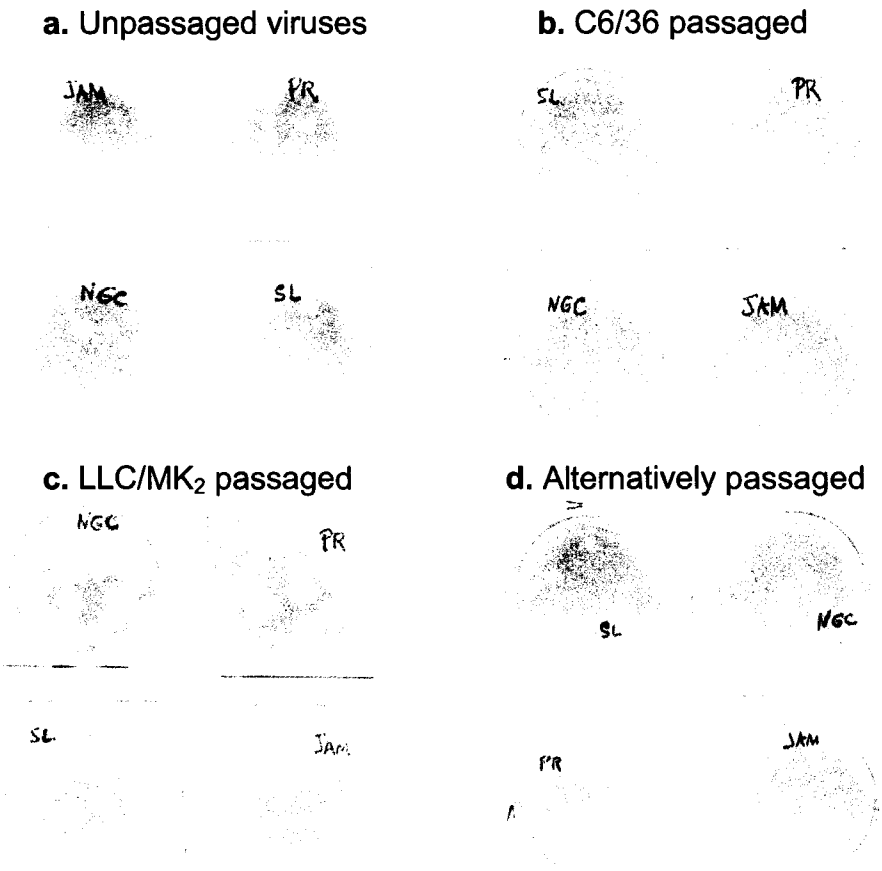


Figure 6-2. Effect of serial passage in plaque size and morphology in viruses of (a.) original "unpassaged" stock or after 24 passages in (b.) C6/36 cells, (c.) LLC-MK_2 cells and (d.) alternatively in both cell lines. DENV-2 strains are Jamaica/83/1409 (JAM), Puerto Rico/69/159 (PR), New Guinea/44/NGC (NGC) and Sri Lanka/85/1592 (SL).

Table 6-1. Plaque assay results and mosquito infectivity at the beginning and at the end of the adaptive process

Series and Passage	Plaque size***	Virus titer (pfu/ml) in the infectious blood meal	Infectivity in Mosquitoes: infected/total [%]
JAM1C6	Medium	$2.4x10^6$	21/60 [35.0]
SL1C6	Medium	$1.3 x10^7$	39/60 [65.0]
PR1C6	Medium	$3.3 x10^6$	26/60 [43.3]
NGC1C6	Small/Large	$1.0 x10^6$	7/60 [11.7]
JAM24LLC	Medium	$1.2x10^5$	0/70 [0.0]
SL24LLC	Large	$1.8x10^6$	0/60 [0.0]
PR24LLC	Large	$8.1x10^5$	0/60 [0.0]
NGC24LLC	Large	$2.9x10^6$	0/60 [0.0]
JAM25*LLC	Small/Medium	$5.7x10^7$	0/60 [0.0]
SL25*LLC	Large	$4.1x10^7$	0/60 [0.0]
PR25*LLC	Medium/Large	$1.1x10^8$	1/60 [1.7]
NGC25*LLC	Medium/Large	$1.2x10^9$	0/60 [0.0]
JAM24C6	Small	$5.0x10^6$	4/32 [12.5]
SL24C6	Medium	$2.9x10^7$	28/60 [46.6]
PR24C6	Small	$1.4x10^6$	7/47 [14.9]
NGC24C6	Very small	N.R.**	13/57 [22.8]
JAM24ALT	Small	$1.2x10^3$	0/18 [0.0]
SL24ALT	Medium	$4.8x10^4$	0/32 [0.0]
PR24ALT	Large	$2.5x10^6$	0/30 [0.0]
NGC24ALT	Medium	$9.0x10^5$	1/25 [4.0]
JAM25ALT	Medium	$5.0x10^7$	3/24 [12.5]
SL25ALT	Large	$7.5x10^7$	5/19 [26.3]
PR25ALT	Large	$6.5x10^6$	0/28 [0.0]
NGC25ALT	Large	$8.5x10^7$	7/41 [17.1]

* 24 passages in LLC/MK_2 plus one additional passage in C6/36
** N.R.: non-readable
*** Plaque sizes: small < 2mm, medium 2-5 mm, large > 5mm

Viral infectivity in mosquitoes. Titers found in the viral suspension used in the infectious blood meals are shown in Table 6-1. Series ending in C6/36 cells had titers between $1.4x10^6$ and $1.2x10^9$. Series ending in LLC-MK_2 cells had titers lower than $3.0x10^6$. An additional passage in C6/36 cells was made to increase the titer. Infection rates in mosquitoes are shown in Figure 6-3 and Table 6-1. No viral antigen was observed in the heads of mosquitoes challenged with LLC series at 24th passage and only one mosquito was infected after the additional (25th) passage in C6/36 cells. Passage 24 viral suspensions of ALT series produced almost no infection but after the additional (25th) passage in C6/36 disseminated infection rates between 12.5% and 26.3% were observed for all but one series. This exception was probably due to the one log lower titer in the PR25ALT suspension. Viruses passaged in C6/36 cells only yielded infectivity rates between 12.5% and 46.6% (Table 6-1).

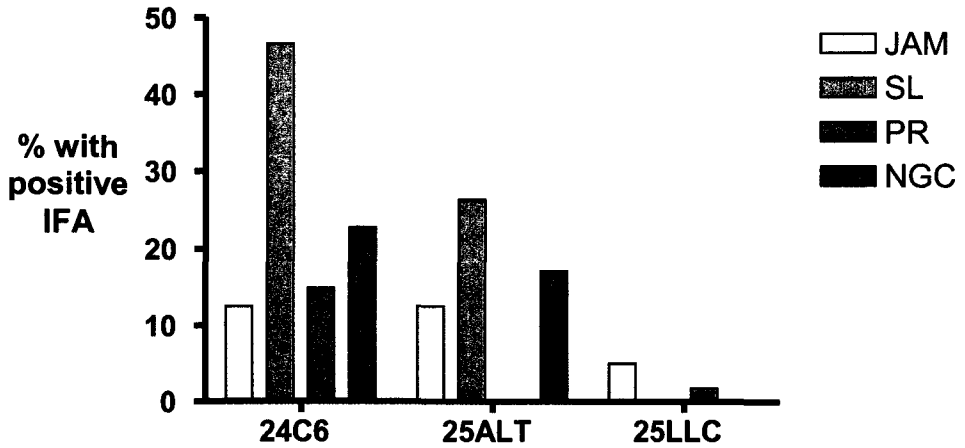


Figure 6-3. Infectivity in mosquitoes of the cell culture-adapted viruses after *per os* challenge with the product of the last passage. This includes the additional passage in C6/36 cells for LLC and ALT series. Infection was determined by IFA in head squashes after 14 days of extrinsic incubation period.

Virus fitness in vertebrate and mosquito cells. The replicative capacity of the cell culture-adapted viruses was assessed for the Puerto Rico/69/159 strain only (Figure 6-4). In LLC-MK_2 cells the virus resulting from LLC series yielded a higher titer early in the incubation period. By day four, the alternately adapted virus reached similar titer as the vertebrate-only adapted virus. The viruses in the original stock and in the C6 series replicated at a slower rate and reached a lower final titer. In C6/36 cells, most viruses initially replicated at a similar rate except the LLC-adapted viruses. However, this virus recovered later and exhibited a similar yield at the end of the experiment. In both cell systems, the viruses produced in C6 series stopped or greatly reduced its replication by the fourth day. The ALT series viruses yielded higher or equal titers to viruses in other series by day four.

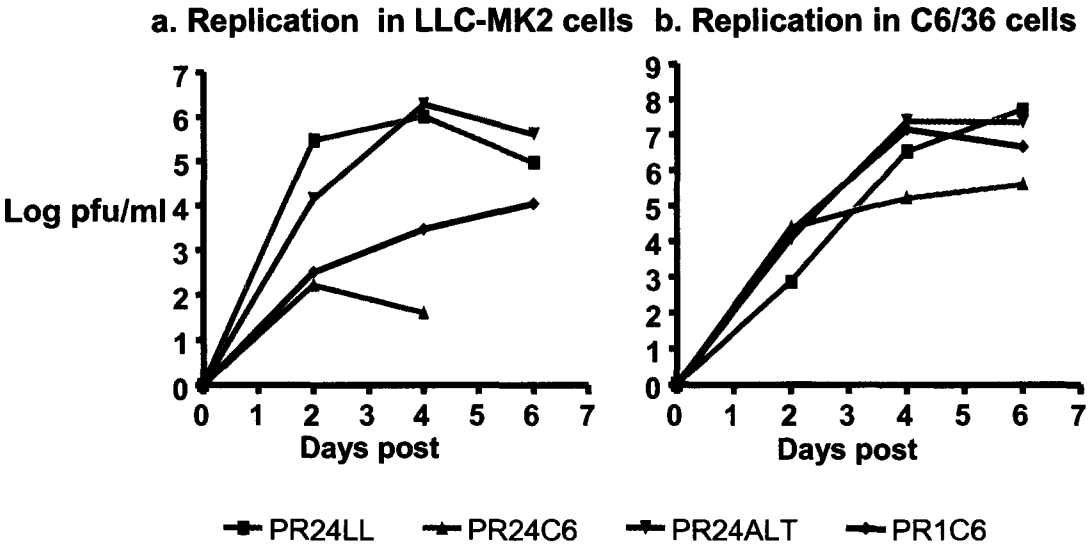


Figure 6-4. Fitness assessment in (a.) LLC-MK2 cells and (b.) C6/36 cells of the cell culture-adapted PR strain in comparison to the "non-adapted" virus (PR1C6). Cells were inoculated at a multiplicity of infection (moi) of 0.005 and titers were determined by plaque assay.

Genotypic Changes.

Nucleotide and amino acid substitutions in the C-prM-E-NS1 genes observed after the serial growth in different cell-systems are listed in Table 6-2. Overall there were thirty-seven nucleotide substitutions, thirty-one (83.8%) of them were non-synonymous. Of these non-synonymous changes, twenty (64.5%) were "non-conservative", defined as substitutions that produce a major change in the charge, polarity or molecular weight of the amino acid residue. The six synonymous substitutions were not repeated among strains or treatments and were not clustered by genome region or specific series. There was at least one deduced amino acid substitution in each series; NGC had the most (fourteen) and SL the fewest (four). There were more non-synonymous changes during passages in vertebrate cells (fifteen) and less during mosquito cell passages (seven). Alternating series accumulated an intermediate number (nine). Six amino acid replacements were observed in four different positions in the pr portion of prM gene, one in M, seventeen in ten positions of E, and eight in seven positions in NS1. No substitutions were observed in the C gene. Most (80.4%) amino acid changes occurred in sites that are highly conserved across strains and genotypes in nature (Gritsun et al., 1995).

Sixteen (51.6%) of the non-synonymous changes were unique to a single series. Fifteen (48.4%) were parallel mutations repeated either in two series of a single strain or in different strains with the same treatment. No substitution was in common between LLC and C6 series but some of the changes in alternately passaged viruses were shared with either one or the other cell line series.

Table 6-2. Nucleotide and amino acid substitutions observed in the cell culture adapted viruses

Position in genome	Nucleotide substitution	Position in protein (domain)	Aminoacid substitution	Series with the substitution
521	A → G	pr-28	Glu → Gly	NGC-LLC, NGC-ALT
601	T → C	pr-55	Phe → Leu	NGC-LLC, NGC-ALT
610	C → G	pr-58	Gln → Glu	PR-LLC
674	C → A	pr-79	Thr → Lys	NGC-LLC
868	A → G	M-53	Ile → Val	SL-LLC
954	G → A	E-6 (I)	Met → Ile	JAM-LL
1063	T → C	E-43 (I)	Phe → Leu	JAM-LL, SL-LL
1093	C → T	E-53 (I-II)	Pro → Ser	NG-LL
1149	T → A	E-71 (II)	Asp → Glu	NGC-LL, NGC-ALT
1176	C → G		none	PR-LL
1393	A → G	E-153 (I)	Asn → Asp	NGC-ALT
1399	A → T	E-155 (I)	Thr → Ser	PR-C6
1400	C → T	E-155 (I)	Thr → Ile	SL-C6, SL-ALT, NGC-C6, JAM-C6, JAM-ALT
1514	T → A	E-193 (I-II)	Phe → Tyr	PR-LL
1540	G → A	E-202 (II)	Glu → Lys	JAM-LL
2140	T → A	E-402 (stalk)	Phe → Ile	NGC-C6
2253	T → C		none	PR-C6
2551	T → A	NS1-44	Ser → Thr	PR-C6
2563	A → C	NS1-48	Lys → Gln	JAM-ALT
2682	T → C		None	NGC-LL
2722	A → G	NS1-101	Lys → Glu	PR-ALT
2735	G → A	NS1-105	Arg → Gln	NGC-C6
2842	G → A	NS1-141	Ala → Thr	JAM-LL
2926	C → T		none	NGC-ALT
3120	T → C		none	NGC-C6
3121	A → G	NS1-234	Asn → Asp	NGC-ALT
3125	G → A	NS1-235	Gly → Glu	NGC-LL, PR-LL
3130	C → T		none	NGC-C6

Figure 6-5 shows the topographic location of the amino acid mutations in the crystal structure of the soluble fragment of DENV-2 envelope (E) glycoprotein dimer (Modis et al., 2004). Most of the substitutions are in domains I and II and none in domain III. They cluster mainly in two regions: in (or near) the interphase or hinge between domains I and II (positions 53, 193 and 202) or near the glycan at Asn 153 in domain I (positions 6, 43, 153 and 155).

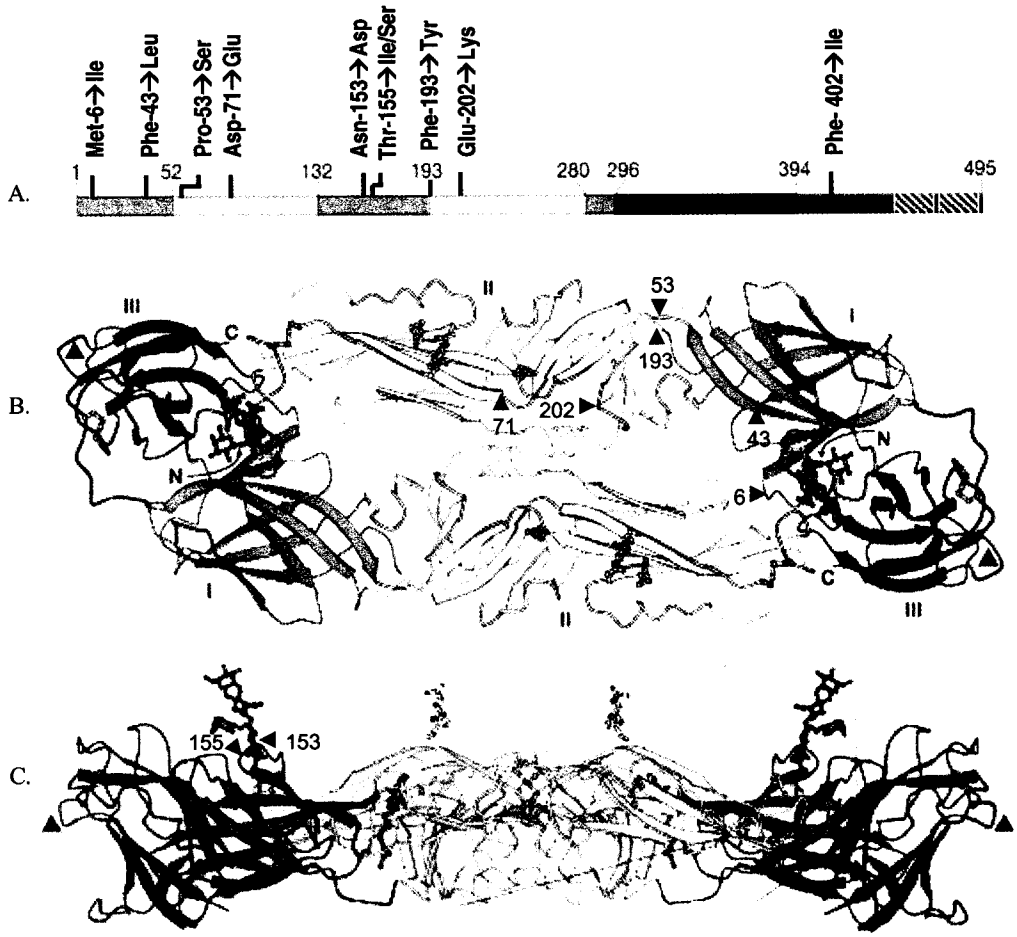


Figure 6-5. Mapping of the substitutions in the 3D-structure of E-protein of DENV. **A**. Amino acid replacements and their positions in the extended sequence of the polipeptide. **B.** Mapping of the substitutions selected upon adaptation to vertebrate (LLC-MK2) cells (top view). **C**. Mapping of the substitutions selected in insect (C6/36) cells and in alternatively passaged series (lateral view). Structural domains are colored in red (domain I), yellow (domain II) and dark blue (domain III). This 3-D structures were determined by Modis et al., (2004).

Mutations in genome positions 1393, 1399 and 1400 of DENV-2 genome affect the glycosylation motif (Asn-X-Thr) by substituting either Asn or Thr at positions 153 and 155, respectively, in the amino acid sequence of the E glycoprotein (Figure 6-5C.). These glycosylation-disrupting substitutions were observed in C6/36 and alternatively passaged viruses only. Other substitutions appearing in the E gene were observed mainly in LLC or ALT series (Figure 6-5 B.).

Table 6-3 shows the results of the tests performed in search of statistical support for positive selection happening during the adaptation process. The non-synonymous substitution rate was more than twice higher than synonsymous substitution rate in LLC and ALT series but lower in C6 series. Fu & Li's F* and Tajima's D test exhibited weakly negative results that are very similar for all the treatments. However, when the sequences of all the series were pooled and analyzed togheter, both tests produced a strong positive value that was statistically significant in the Fu & Li's F* test ($p < 0.02$).

Table 6-3. Result of the tests for selection in the adapted strains.

Statistic	Passage series			All 12 passage series
	C6	LLC	ALT	
	7/2582	15/2581	9/2581	31/2581
(d_n**)	(0.00271)	(0.00581)	(0.00349)	(0.01201)
	3/766	2/767	1/767	6/767
(d_s***)	(0.00392)	(0.00261)	(0.00130)	(0.00783)
$\omega = d_n/d_s$	0.692	2.226	2.674	1.534
Fu & Li's F*	-0.339 (n.s.)	-0.343 (n.s.)	-0.336 (n.s.)	2.088 ($p<0.02$)
Tajima's D	-0.373 (n.s.)	-0.370 (n.s.)	-0.364 (n.s.)	1.256 (n.s.****)

** d_n: non-synonymous substitutions per non-synonymous site;
*** d_s: synonymous substitutions per synonymous site;
**** n.s.: non-significant

DISCUSSION

An adaptive experiment was conducted to explore the evolutionary consequences of the obligatory alternation of DENV-2 between insect and vertebrate host-cells similar to previous studies using VSV or EEEV (Chen et al., 2003, Cooper & Scott, 2001, Novella et al., 1999, Weaver et al., 1999). In those studies one or several replicates of a single strain of an arbovirus were subjected to different adaptive regimens consisting of serial passages in mosquito cells, vertebrate cells and alternating between the two systems.

This study is unique in that four different strains of DENV-2 were subject of the same treatments and in that it is the only one that addressed mosquito infectivity after the adaptive process. Replications of the adaptive process in the same strain were not performed, which limits the strength of the inferences about the evolutionary forces that generated the changes. However, the consistency of some of the phenotypic and genotypic changes observed among the four different strains, and the uniqueness of others, suggested deterministic or stochastic effects, respectively. It also allowed discriminating between strain-specific or serotype-wide effects, an inference that could not be made if replicates of a single strain were used.

The six synonymous nucleotide substitutions observed (Table 6-2) were more likely fixed by random genetic drift. They were not repeated in different series and did not exhibit any particular regularity that suggested a selective trend. They provide an estimate of the background neutral substitution rate in the experiment and will not be

discussed any further. Conversely, non-synonymous substitutions were more likely fixed by different selective forces because approximately two thirds of them were non-conservative and about half were replicated in more than one series in this study or were previously observed in other studies (Guirakhoo et al., 1993, Johnson et al., 1994, Lee et al., 1997).

Several substitutions deserve special consideration. The Met→Ile mutation at E-6 position in the glycoprotein observed in JAM-LLC series only had been previously seen in the same strain but in the opposite direction after propagating it in C6/36 cells at a low pH (Guirakhoo et al., 1993). All the other strains had isoleucine in that position, which suggests that it was the original residue and that the methionine in the initial JAM stock must have been introduced in passages previous to this experiment. What was observed is more likely a reversion to the ancestral state when the virus was subjected to a vertebrate-only adaptive regimen. This change could be regarded as a strain-specific effect of directional selection acting in opposite directions depending on the host-cell used.

Similar events happened with substitutions in positions 521, 601 and 1149 all observed in NGC-LLC and NGC-ALT series. These three mutations preexisted as minor species in the original NGC stock as revealed by double peaks in the electropherogram (data not showed) and their positions are also polymorphic in natural populations. After the adaptive process, they were fixed in one amino acid upon adaptation in mosquito cells or in the other in the vertebrate and alternating series. This is an example of lineage sorting due to differential selection, although fixation by random genetic drift cannot be conclusively excluded. The absence of substitutions in

these sites in other DENV-2 strains suggests that selection is not only dependent on the environment but also on the genetic background of the virus.

Three non-conservative substitutions at positions 53, 193 and 202 occurred in three different DENV-2 strains in the LLC series and mapped in or close to the boundaries of domains I and II (Figure 6-5 B.). This region is regarded as the hinge in the dramatic conformational change that E glycoprotein undergoes during the fusion of viral and cell membranes (Modis et al 2004). These changes could be modifying the flexibility of the hinge and thereby facilitating the fusion process in LLC-MK_2 cells. These data suggest that the fusion differs between vertebrate and mosquito cells, which may be related to the differences between the composition of insect and vertebrate cell membranes.

The most compelling evidence of directional selection acting in these adaptive processes is the presence of parallel mutations in different strains undergoing the same treatment. Several of these amino acid changes occurred. For instance, in position E-43 the Phe→Leu substitution was observed in JAM-LLC and SL-LLC series. Leucine is smaller than phenylalanine but both amino acids have neutral charge and similar hydrophobicity, so the significance of this change is uncertain. An interesting adaptation was observed at adjacent amino acid positions 234 and 235 in the NS1 protein. The NGC-LLC, NGC-ALT and PR-LLC series exhibited substitutions directed toward an acidic amino acid, either Asn234→Asp or Gly235→Glu. The JAM-ALT, SL-ALT and PR-ALT series did not yield a well-fixed substitution in these positions, but the presence of double peaks in the electropherogram revealed that the viral populations were polymorphic in one or the other site, suggesting that a replacement process was underway. This may be due to directional selective pressure for the acquisition of an

acidic residue in any of two contiguous positions. Since neither the spatial structure nor the function of NS1 protein are known, the significance of these changes is not clear at this point.

The clearest example of selective evolution was the disruption of the gycosylation site at Asn153 by different nucleotide substitutions at positions 1393, 1399 and 1400, which in turn resulted in replacement of either Asn153 or Thr155. These changes were observed in the consensus sequences of all series involving virus passage in mosquito cells, except for PR-ALT. In this latter series double peaks in the electropherogram at positions 1393 and 1400 suggested that the viral population was subdivided into two groups with substitutions in one or the other amino acid. Double peaks in the nucleotides coding for these Asn153 and Thr155 were also observed in the series that did show substitutions in the consensus sequence. These multiple double peaks as well as unique changes in either the first or the third position of the glycosylation motif (Asn-X-Thr) indicate that different solutions were used to solve the same adaptive problem of how to remove the glycan at Asn153. The replacement of the threonine by serine in PR-C6 series would still retain the glycosylation motif but the consistency of the changes in the other strains suggests that serine is not glycosylated in this case. The Asn-X-Ser motif is a less efficient substrate than Asn-X-Thr. The disruption of this glycosylation site could be interpreted as an example of directional selection or, alternatively, of strong purifying selection acting only in mosquito cells. Interestingly these changes also appeared in all alternating but in none of the LLC series, which suggests that the presence or absence of the glycan is neutral for replication in LLC-MK_2 cells. The loss of this glycosylation motif had been observed before in DENV-2 Jamaica (Guirakhoo et al., 1993) and in DENV-3 (Lee et al., 1997) when these viruses were serially propagated in mosquito cells.

This glysosylation site is highly conserved in natural populations of all DENV serotypes. The constancy of this motif, in spite of its deleterious effect in mosquito cells, implies that the presence of an oligosaccharide at that site is required in the natural cycle of DENV. The elimination of the glycosylation did not abrogate the infection in the mosquito by the oral route (Figure 6-3). Thus, it must be required in the vertebrate part of the cycle. The glycan at this position has been implicated in DENV entry to the cell (Hung et al., 1999) and, recently, a lectin-like molecule, dendritic-cell-specific ICAM3-grabbing non-integrin (DC-SIGN) has been proposed as the receptor of DENV in dendritic cells (Navarro-Sanchez et al., 2003, Tassaneetrithep et al., 2003). These are probably the first cells infected with the virus after the mosquito bite. DC-SIGN binds mannose residues within high-mannose oligosaccharides, the kind of glycans utilized by insect cells (Feinberg et al., 2001). It seems reasonable to postulate that the glycosylation site at Asn153 persists for the sake of the early steps of the infection in primates, although its presence is otherwise deleterious or dispensable. The basis for the negative effect of that glycosylation in C6/36 cells is not clear but it could impose some kind of spatial hindrance since it projects outward from the E glycoprotein dimer (Figure 6-5 C.).

Despite the consistency of these mutations in viruses serially adapted to C6/36 cells no statistical support for positive selection was found in any test performed for these series (Table 6-3). It is likely that background purifying selection in most other sites of the analyzed genomic region could have reduced the non-synonymous to synonymous substitution rate ratio (ω). Conversely, this indicator was positive for viruses in ALT and LLC series, a result that suggests that directional selection took place in several sites of the sequence of viruses undergoing adaptation to vertebrate

cells (Yang et al., 2003). This suggestion was not supported by Tajima's D and Fu & Li's F* tests but this could have been due to the small sample size (four sequences) included in the analysis of individual series. When all the series were analyzed together these statistics were strongly positive, and in one case (F*) the result was statistically significant.

Comparing the indicators and statistics in Tables 3-3 and 6-3 it is clear that evolution in natural populations and in the cell-culture microenvironment develops in very different ways. Most variability in the sequence of DENV-2 isolates was in synonymous sites indicating strong purifying selection while overall predominance of non-synonymous substitutions in the laboratory-adapted strains suggests directional selection. Another finding supporting this difference was that most amino acid replacements occurring in cell-culture were located at residues that are highly conserved in natural populations. This suggests that the *in vitro* adaptive landscape does not resemble that of the natural cycle. For example, no substitutions were observed in structural domain III of E glycoprotein, the most variable domain in nature (Gritsun et al., 1995). Domain III contains the putative ligand for the viral receptor and it is the most important antigenic region for the induction of protective immunity (Crill & Roehrig, 2001, Hung et al., 2004, Roehrig, 2003, Roehrig et al., 1998). The presence of herd immunity may have stimulated diversification in this region in the natural population over time by immune evasion, an adaptation not necessary in cell-culture. The lack of substitutions in this domain observed in these experiments also suggests that some viral receptors are conserved between insect and vertebrate cells. However, in nature, different receptors and co-receptors could be acting, depending on the host and tissue being infected (Bielefeldt-Ohmann et al., 2001). The findings presented in the glycosylation motif described in the former paragraphs support this hypothesis.

Significant modifications in the phenotype were also observed after serial passage in the simplified microenvironment of cell cultures. At the end of the adaptive process plaque size decreased in all C6 series and increased in most LLC adapted viruses (Figure 6-2). Failure to produce bigger plaques by JAM24C6 is in agreement with the aforementioned difficulty of this virus to serially propagate and induce CPE in vertebrate cells. ALT series yielded viruses that produced different plaque sizes that could be as big as those of the LLC series (PR24ALT) or almost as small as those in the C6 series (JAM24ALT). In general, it can be said that this phenotypic character evolved in a cell-type specific manner, as expected.

When the PR cell culture-adapted viruses were compared with the virus in the original stock significant improvements in their fitness in vertebrate cells were observed in LLC and ALT but not in C6 series (Figure 6-4). In C6/36 there was neither an increase nor a reduction on their replicative capacity in PR24C6 and PR24ALT series. This could be due to the fact that the original stock already had five passages in mosquito cells that may have pre-adapted the viral population to an optimal fitness in C6/36 that could not be improved any further. PR24LLC virus exhibited a delay in its replication in this cell line early in the incubation period but it rapidly regained fitness at a level similar to viruses in other series. In general, different series demonstrated a tendency to replicate better in the cell system in which they were adapted, but they exhibited few or no fitness losses in the bypassed cell type. These results are intermediate to those reported for VSV (Novella *et al* 1999) and for EEEV (Weaver *et al* 1999). As in all the other studies (Cooper & Scott, 2001, Novella et al., 1999, Weaver et al., 1999), virus undergoing an alternating host-cell treatment did not exhibit a significant reduction in its fitness in comparison to either single-host series.

Virus adaptation to the cell culture environment seems to have decreased the viral capacity to infect or to disseminate in mosquitoes to different extents depending on the cell type used (Table 6-1 and Figure 6-3). Disseminated infection was detected in most strains after serial passage in C6/36 cells but with lower efficiency than was observed with the initial stocks of low passaged viruses. Failure to detect infection after passage 24 in ALT series could be due to the lower viral titer obtained in the LLC-MK_2 cell line used in that 24^{th} passage, because after an additional passage in C6/36 cells, these viruses recovered their ability to infect mosquitoes. However, an additional passage in C6/36 did not significantly recover infectivity in viruses propagated exclusively in vertebrate cells in spite of the very high titer obtained after the additional passage (Table 6-1). It seems that some changes present in all viruses serially passaged in LLC-MK2 cells interfere either with infection of the mosquito midgut or with its dissemination to other tissues without a significant impairment of its capacity to replicate in C6/36 cells (Figure 6-4B.). The nature of such genetic modifications is not clear at this moment. The substitutions near the hinge of the E glycoprotein could be part of the changes reducing infectivity in mosquitoes but there were no substitutions in this region in the SL24LLC virus, which also failed to infect mosquitoes. Study of the entire genome sequence of these adapted viruses could shed some light on the interaction of the virus with different mosquito tissues.

We did not test the infectivity of the adapted virus in primates or any other vertebrate animal. It seems reasonable to predict that the lack of glycosylation at position E-153 will prevent C6/36 and alternately passaged virus from infecting vertebrates by the natural route. Testing this hypothesis may provide further evidence

of the role of dendritic cells and DC-SIGN receptor in the pathogenesis of DENV infection.

It is not possible at this point to attribute specific changes in the phenotype to specific substitutions in the genome. However, these results could be used to formulate hypotheses about the function of some amino acid residues in the virus-cell interaction and about their value in the adaptive process in different hosts. These hypotheses can be tested using site directed mutagenesis in DENV-2 infectious clones that are now available.

The reduced number of nucleotide substitutions observed in viruses passaged in alternating series in a different study were presented in support of the prevailing idea that arboviral evolution is constrained by the double selective environment imposed by host alternation (Weaver et al., 1999). However, another study of similar design failed to demonstrate that tendency (Novella et al., 1999). In this study the number of mutations observed in the ALT series (9) was intermediate to the 15 found in LLC series and the 7 in the C6 series. These results did not support or dispel any preconceived idea about the implications of the alternation in host in the evolutionary rates observed in arboviruses. However the similar fitness gains observed in ALT series in comparison to homologous cell series, which is consistent with what was found in all the other studies (Cooper & Scott, 2001, Novella et al., 1999, Weaver et al., 1999), undermines the notion of a compromised fitness of arboviruses in nature.

In summary, this study showed that different selective pressures were acting on DENV when it was serially propagated in cells of different lineage. The replication of some phenotypic and genotypic changes across different strains undergoing similar

adaptive regimes showed that directional selection was occurring during the process. On the other hand, this study suggested that cultured vertebrate and insect cells are not such disparate adaptive systems. There seems to be considerable overlap of the fitness landscapes in both host-cells that enable the viruses adapted to a host cell type to thrive in the other. However, no inferences about the implications of the alternation between hosts could be derived from the data. Some of the changes in gene E that were consistently observed in different adaptive series provided insight into the virus-cell interaction, mainly in the adsorption and penetration steps. Sequencing of the whole genome of the strains adapted in this study could provide similar insight in other steps of the viral cycle.

CHAPTER 7

SUMMARY AND CONCLUSIONS

Several observational and experimental approaches were used to investigate DENV evolution, the main objective of the studies presented in this dissertation. The phylogenetic analyses presented in chapters 2 and 3 demonstrated that, for most DENV serotypes, a single strain has been circulating in the Mexico during the last two decades. These strains exhibited considerable mobility among countries and frequently went to extinction in one region but were reintroduced a few years later, apparently from neighboring countries. The DENV-1 strain circulating in Mexico during the 1990's was not derived from that isolated in the 1980's in the same country although both strains belong to the same genotype. DENV-3 was introduced more recently, which accounts for the lower diversity observed among isolates. No clear pattern of evolution for this serotype has emerged yet. DENV-4 was unique in that it evolved in Mexico in a linear fashion suggesting that it did not stop circulating since its introduction in 1983.

DENV-2 was atypical in that multiple genotypes were detected in the virus collection studied here. The strain circulating during the 1980's and early 1990's belonged to the American genotype. As in other countries of the Western Hemisphere this genotype has been largely replaced in Mexico by DENV-2 strains classified in other genotypes. Strains belonging to the Cosmopolitan and Asian-2 genotypes were identified in Mexican DENV-2 isolates from the mid 1990's. This was the first, and so far the only report of the Cosmopolitan genotype in the Americas (Diaz et al., 2002).

The isolates of the Asian-2 genotype added evidence for the controversial presence of a strain of this genotype in this continent (Rico Hesse, 2003).

By the year 2000, a new DENV-2 strain from the American-Asian genotype was isolated in Mexico. This was the same strain detected almost twenty years before in Jamaica, which has been implicated as responsible for the emergence of DHF in the Americas (Rico-Hesse, 1997). However, this study showed that the appearance of DHF in Mexico during the mid 1990's was temporally associated with the introduction of other genotypes of DENV-2 as well as DENV-3 genotype III. This work also provides evidence that the DENV-3 introduced in Central America in 1994 was more likely imported from Africa.

The occurrence of recombination in DENV in nature and in the laboratory was also investigated. It was found that the genomes of all DENV-1 isolated to date in the Americas exhibited a mosaic structure suggestive of a remote recombination event between strains circulating in the Eastern Hemisphere. A single DENV-2 isolate obtained from a 1983 Mexican sample showed a sequence suggestive of a recombination event between viruses of the American and the American-Asian genotypes. The search for DENV recombinants in the laboratory was not so productive and clear evidence of this phenomenon in a cell culture system was not obtained.

In the final part of this work, the evolution of DENV in different cell culture environments was investigated. Serial passage of four DENV-2 strains in a mammalian cell line generated multiple nonsynonymous substitutions that were predicted to significantly alter the three-dimensional structure and mobility of domains I and II of E glycoprotein on the virion surface. This treatment abolished the oral infectivity of the

virus in mosquitoes but only slightly reduced its infectivity for insect cells. Serial passage of the same strains in mosquito cells generated fewer substitutions but a single major change, the disruption of a glycosylation site, was regularly observed in all strains. This treatment produced a modest reduction in the oral infectivity in mosquitoes and in vertebrate cells. Alternate passage of viruses in vertebrate and mosquito cells induced changes intermediate to those observed in each single-cell type passage.

Comparison of the genetic changes observed in nature and *in vitro* allowed inferring the main evolutionary forces involved in DENV evolution. It was found that in nature, DENV continuously accumulate synonymous mutations that are not reproduced among lineages. The few amino acid substitutions observed were not associated with epidemiological events. Major changes in the incidence and severity of DENV infections were more consistently related to introduction of strains that had not been circulating before. Genetic changes observed during serial propagation of DENV in cell culture exhibited a different pattern. Most substitutions were nonsynonymous and occurred in sites that are highly conserved in nature. About half of these changes were reproduced in more than one strain undergoing the same treatment.

The reproducibility and nonsynonymous nature of substitutions observed in viruses serially passaged in cell culture clearly indicate that directional selection is the predominant force in such circumstances. This selection is dependent on the environment since no substitution was repeated in viruses adapted to different cell culture types. In nature, the paucity of nonsynonymous mutations indicates that such changes are strongly constrained, which indicates that purifying selection is a dominant force. Most substitutions were silent suggesting that they were fixed by random genetic

drift. More dramatic changes were produced by migratory events represented by introductions of strains imported from overseas. Thus, different evolutionary forces determine DENV evolution *in vivo* and *in vitro*.

REFERENCES

Agol, V. I., Drozdov, S. G., Frolova, M. P., Grachev, V. P., Kolesnikova, M. S., Kozlov, V. G., Ralph, N. M., Romanova, L. I., Tolskaya, E. A. & Viktorova, E. G. (1985). Neurovirulence of the intertypic poliovirus recombinant v3/a1-25: characterization of strains isolated from the spinal cord of diseased monkeys and evaluation of the contribution of the 3' half of the genome. *J Gen Virol* **66 (Pt 2)**, 309-16.

Alejska, M., Kurzyniska-Kokorniak, A., Broda, M., Kierzek, R. & Figlerowicz, M. (2001). How RNA viruses exchange their genetic material. *Acta Biochim Pol* **48**, 391-407.

Anderson, R., Wang, S., Osiowy, C. & Issekutz, A. C. (1997). Activation of endothelial cells via antibody-enhanced dengue virus infection of peripheral blood monocytes. *J Virol* **71**, 4226-32.

Anonymous (1996). Dengue fever at the U.S.-Mexico border, 1995-1996. *MMWR Morb Mortal Wkly Rep* **45**, 841-4.

Armstrong, P. M. & 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**, 159-68.

Armstrong, P. M. & Rico-Hesse, R. (2003). Efficiency of dengue serotype 2 virus strains to infect and disseminate in Aedes aegypti. *Am J Trop Med Hyg* **68**, 539-44.

Baldridge, G. D., Beaty, B. J. & Hewlett, M. J. (1989). Genomic stability of La Crosse virus during vertical and horizontal transmission. *Arch Virol* **108**, 89-99.

Baric, R. S., Fu, K., Schaad, M. C. & Stohlman, S. A. (1990). Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. *Virology* **177**, 646-56.

Barnes, W. J. & Rosen, L. (1974). Fatal hemorrhagic disease and shock associated with primary dengue infection on a Pacific island. *Am J Trop Med Hyg* **23**, 495-506.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A* **91**, 2216-20.

Beaty, B. J., Borucki, M., Farfan, J. & White, D. (1997). Arbovirus-vector interactions: determinants of arbovirus evolution. *Factors in the emergence of disease, Elsevier*, 23-35.

Bennett, K. E., Olson, K. E., Munoz Mde, L., Fernandez-Salas, I., Farfan-Ale, J. A., Higgs, S., Black, W. C. t. & Beaty, B. J. (2002). Variation in vector competence for dengue 2 virus among 24 collections of Aedes aegypti from Mexico and the United States. *Am J Trop Med Hyg* **67**, 85-92.

Bielefeldt-Ohmann, H., Meyer, M., Fitzpatrick, D. R. & Mackenzie, J. S. (2001). Dengue virus binding to human leukocyte cell lines: receptor usage differs between cell types and virus strains. *Virus Res* **73**, 81-9.

Blok, J., Samuel, S., Gibbs, A. J. & Vitarana, U. T. (1989). Variation of the nucleotide and encoded amino acid sequences of the envelope gene from eight dengue-2 viruses. *Arch Virol* **105**, 39-53.

Bonner, S. M. & O'Sullivan, M. A. (1998). Endothelial cell monolayers as a model system to investigate dengue shock syndrome. *J Virol Methods* **71**, 159-67.

Bradley, R., D., and Hillis, D., M., (1997). Recombinant DNA sequences generated by PCR amplification. *Mol Biol Evol* **14**, 592-593.

Brandt, W. E., McCown, J. M., Gentry, M. K. & Russell, P. K. (1982). Infection enhancement of dengue type 2 virus in the U-937 human monocyte cell line by antibodies to flavivirus cross-reactive determinants. *Infect Immun* **36**, 1036-41.

Bravo, J. R., Guzman, M. G. & Kouri, G. P. (1987). Why dengue haemorrhagic fever in Cuba? 1. Individual risk factors for dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS). *Trans R Soc Trop Med Hyg* **81**, 816-20.

Bray, M., Men, R. & Lai, C. J. (1996). Monkeys immunized with intertypic chimeric dengue viruses are protected against wild-type virus challenge. *J Virol* **70**, 4162-6.

Briseno-Garcia, B., Gomez-Dantes, H., Argott-Ramirez, E., Montesano, R., Vazquez-Martinez, A. L., Ibanez-Bernal, S., Madrigal-Ayala, G., Ruiz-Matus, C., Flisser, A. & Tapia-Conyer, R. (1996). Potential risk for dengue hemorrhagic fever: the isolation of serotype dengue-3 in Mexico. *Emerg Infect Dis* **2**, 133-5.

Buonagurio, D. A., Nakada, S., Parvin, J. D., Krystal, M., Palese, P. & Fitch, W. M. (1986). Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. *Science* **232**, 980-2.

Burke, D. S., Nisalak, A., Johnson, D. E. & Scott, R. M. (1988). A prospective study of dengue infections in Bangkok. *Am J Trop Med Hyg* **38**, 172-80.

Butrapet, S., Rabablert, J., Angsubhakorn, S., Wiriyarat, W., Huang, C., Kinney, R., Punyim, S. & Bhamarapravati, N. (2002). Chimeric dengue type 2/type 1 viruses

induce immune responses in cynomolgus monkeys. *Southeast Asian J Trop Med Public Health* **33**, 589-99.

Cascone, P. J., Carpenter, C. D., Li, X. H. & Simon, A. E. (1990). Recombination between satellite RNAs of turnip crinkle virus. *Embo J* **9**, 1709-15.

Cascone, P. J., Haydar, T. F. & Simon, A. E. (1993). Sequences and structures required for recombination between virus-associated RNAs. *Science* **260**, 801-5.

CDC (1995). Dengue type 3 infection--Nicaragua and Panama, October-November 1994. *MMWR Morb Mortal Wkly Rep* **44**, 21-4.

Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. (1990). Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol* **44**, 649-88.

Chang, D. M. & Shaio, M. F. (1994). Production of interleukin-1 (IL-1) and IL-1 inhibitor by human monocytes exposed to dengue virus. *J Infect Dis* **170**, 811-7.

Chang, G. J. (1997). Molecular biology of dengue virus. *in Gubler DJ, Kuno G, Dengue and Dengue Hemorrhagic Fever* Cab International, Wallingford UK.

Chang, G. J., Davis, B. S., Hunt, A. R., Holmes, D. A. & Kuno, G. (2001). Flavivirus DNA vaccines: current status and potential. *Ann N Y Acad Sci* **951**, 272-85.

Chang, G. J., Kuno, G., Purdy, D. E. & Davis, B. S. (2004). Recent advancement in flavivirus vaccine development. *Expert Rev Vaccines* **3**, 199-220.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454-5.

Chare, E. R., Gould, E. A. & Holmes, E. C. (2003). Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J Gen Virol* **84**, 2691-703.

Chen, W. J., Wu, H. R. & Chiou, S. S. (2003). E/NS1 modifications of dengue 2 virus after serial passages in mammalian and/or mosquito cells. *Intervirology* **46**, 289-95.

Chen, Y., Maguire, T., Hileman, R. E., Fromm, J. R., Esko, J. D., Linhardt, R. J. & Marks, R. M. (1997). Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* **3**, 866-71.

Cheng, S., Fockler, C., Barnes, W. M. & Higuchi, R. (1994). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc Natl Acad Sci U S A* **91**, 5695-9.

Chetverin, A. B., Chetverina, H. V., Demidenko, A. A. & Ugarov, V. I. (1997). Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell* **88**, 503-13.

Chu, M. C., O'Rourke, E. J. & Trent, D. W. (1989). Genetic relatedness among structural protein genes of dengue 1 virus strains. *J Gen Virol* **70 (Pt 7)**, 1701-12.

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 (Pt 7)**, 1877-84.

Chungue, E., Deubel, V., Cassar, O., Laille, M. & Martin, P. M. (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 (Pt 12)**, 2765-70.

Clark, T. (2002). Break-bone fever. *Nature* **416**, 672-674.

Cologna, R. & Rico-Hesse, R. (2003). American genotype structures decrease dengue virus output from human monocytes and dendritic cells. *J Virol* **77**, 3929-38.

Conzelmann, K. K. (1998). Nonsegmented negative-strand RNA viruses: genetics and manipulation of viral genomes. *Annu Rev Genet* **32**, 123-62.

Cooper, L. A. & Scott, T. W. (2001). Differential evolution of eastern equine encephalitis virus populations in response to host cell type. *Genetics* **157**, 1403-12.

Crill, W. D. & Roehrig, J. T. (2001). Monoclonal antibodies that bind to domain III of dengue virus E glycoprotein are the most efficient blockers of virus adsorption to Vero cells. *J Virol* **75**, 7769-73.

Diaz, F. J., Farfan-Ale, J. A., Olson, K. E., Lorono-Pino, M. A., Gubler, D. J., Blair, C. D., Black, W. C. t. & Beaty, B. J. (2002). Genetic variation within the premembrane coding region of dengue viruses from the Yucatan peninsula of Mexico. *Am J Trop Med Hyg* **67**, 93-101.

Domingo, E. & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu Rev Microbiol* **51**, 151-78.

Drake, J. W. & Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A* **96**, 13910-3.

Eckels, K. H., Dubois, D. R., Putnak, R., Vaughn, D. W., Innis, B. L., Henchal, E. A. & Hoke, C. H., Jr. (2003). Modification of dengue virus strains by passage in

primary dog kidney cells: preparation of candidate vaccines and immunization of monkeys. *Am J Trop Med Hyg* **69**, 12-6.

Eckels, K. H. & Putnak, R. (2003). Formalin-inactivated whole virus and recombinant subunit flavivirus vaccines. *Adv Virus Res* **61**, 395-418.

Espinoza-Gomez, F., Hernandez-Suarez, C. M., Rendon-Ramirez, R., Carrillo-Alvarez, M. L. & Flores-Gonzalez, J. C. (2003). [Interepidemic transmission of dengue in the city of Colima, Mexico]. *Salud Publica Mex* **45**, 365-70.

Farfan, J. A., Olson, K. E., Black, W. C. t., Gubler, D. J. & Beaty, B. J. (1997). Rapid characterization of genetic diversity among twelve dengue-2 virus isolates by single-strand conformation polymorphism analysis. *Am J Trop Med Hyg* **57**, 416-22.

Feinberg, H., Mitchell, D. A., Drickamer, K. & Weis, W. I. (2001). Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **294**, 2163-6.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368-76.

Felsenstein, J. & Yokoyama, S. (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**, 845-59.

Figlerowicz, M. & Bujarski, J. J. (1998). RNA recombination in brome mosaic virus, a model plus strand RNA virus. *Acta Biochim Pol* **45**, 847-68.

Figlerowicz, M., Nagy, P. D. & Bujarski, J. J. (1997). A mutation in the putative RNA polymerase gene inhibits nonhomologous, but not homologous, genetic recombination in an RNA virus. *Proc Natl Acad Sci U S A* **94**, 2073-8.

Figlerowicz, M., Nagy, P. D., Tang, N., Kao, C. C. & Bujarski, J. J. (1998). Mutations in the N terminus of the brome mosaic virus polymerase affect genetic RNA-RNA recombination. *J Virol* **72**, 9192-200.

Foster, J. E., Bennett, S. N., Carrington, C. V., Vaughan, H. & McMillan, W. O. (2004). Phylogeography and molecular evolution of dengue 2 in the Caribbean basin, 1981-2000. *Virology* **324**, 48-59.

Foster, J. E., Bennett, S. N., Vaughan, H., Vorndam, V., McMillan, W. O. & Carrington, C. V. (2003). Molecular evolution and phylogeny of dengue type 4 virus in the Caribbean. *Virology* **306**, 126-34.

Gagnon, S. J., Mori, M., Kurane, I., Green, S., Vaughn, D. W., Kalayanarooj, S., Suntayakorn, S., Ennis, F. A. & Rothman, A. L. (2002). Cytokine gene

expression and protein production in peripheral blood mononuclear cells of children with acute dengue virus infections. *J Med Virol* **67**, 41-6.

Gallei, A., Pankraz, A., Thiel, H. J. & Becher, P. (2004). RNA recombination in vivo in the absence of viral replication. *J Virol* **78**, 6271-81.

Georgescu, M. M., Delpeyroux, F., Tardy-Panit, M., Balanant, J., Combiescu, M., Combiescu, A. A., Guillot, S. & Crainic, R. (1994). High diversity of poliovirus strains isolated from the central nervous system from patients with vaccine-associated paralytic poliomyelitis. *J Virol* **68**, 8089-101.

Gomez-Dantes, H. (1992). Monografia sobre el dengue en Mexico. *Secretaria de Salud, Mexico*.

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**, 110-9.

Gorman, O. T., Bean, W. J. & Webster, R. G. (1992). Evolutionary processes in influenza viruses: divergence, rapid evolution, and stasis. *Curr Top Microbiol Immunol* **176**, 75-97.

Green, S., Vaughn, D. W., Kalayanarooj, S., Nimmannitya, S., Suntayakorn, S., Nisalak, A., Lew, R., Innis, B. L., Kurane, I., Rothman, A. L. & Ennis, F. A. (1999). Early immune activation in acute dengue illness is related to development of plasma leakage and disease severity. *J Infect Dis* **179**, 755-62.

Gritsun, T. S., Holmes, E. C. & Gould, E. A. (1995). Analysis of flavivirus envelope proteins reveals variable domains that reflect their antigenicity and may determine their pathogenesis. *Virus Res* **35**, 307-21.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev* **11**, 480-96.

Gubler, D. J. & Clark, G. G. (1995). Dengue/dengue hemorrhagic fever: the emergence of a global health problem. *Emerg Infect Dis* **1**, 55-7.

Gubler, D. J., Kuno, G., Sather, G. E. & Waterman, S. H. (1985). A case of natural concurrent human infection with two dengue viruses. *Am J Trop Med Hyg* **34**, 170-3.

Gubler, D. J., Sather, G. E., Kuno, G. & Cabral, J. R. (1986). Dengue 3 virus transmission in Africa. *Am J Trop Med Hyg* **35**, 1280-4.

Guirakhoo, F., Arroyo, J., Pugachev, K. V., Miller, C., Zhang, Z. X., Weltzin, R., Georgakopoulos, K., Catalan, J., Ocran, S., Soike, K., Ratterree, M. & Monath,

T. P. (2001). Construction, safety, and immunogenicity in nonhuman primates of a chimeric yellow fever-dengue virus tetravalent vaccine. *J Virol* **75**, 7290-304.

Guirakhoo, F., Hunt, A. R., Lewis, J. G. & Roehrig, J. T. (1993). Selection and partial characterization of dengue 2 virus mutants that induce fusion at elevated pH. *Virology* **194**, 219-23.

Guzman, M. G., Deubel, V., Pelegrino, J. L., Rosario, D., Marrero, M., Sariol, C. & 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**, 241-6.

Guzman, M. G. & Kouri, G. (2002). Dengue: an update. *Lancet Infect Dis* **2**, 33-42.

Guzman, M. G. & Kouri, G. (2003). Dengue and dengue hemorrhagic fever in the Americas: lessons and challenges. *J Clin Virol* **27**, 1-13.

Guzman, M. G., Kouri, G. P., Bravo, J., Soler, M., Vazquez, S. & Morier, L. (1990). Dengue hemorrhagic fever in Cuba, 1981: a retrospective seroepidemiologic study. *Am J Trop Med Hyg* **42**, 179-84.

Guzman, M. G., Kouri, G. P., Bravo, J., Soler, M., Vazquez, S., Santos, M., Villaescusa, R., Basanta, P., Indan, G. & Ballester, J. M. (1984). Dengue haemorrhagic fever in Cuba. II. Clinical investigations. *Trans R Soc Trop Med Hyg* **78**, 239-41.

Guzman, M. G., Rosario, D., Mune, M., Alvarez, M., Rodriguez, R. & Kouri, G. (1996a). [Genetic relatedness of the dengue 3 virus isolated in the outbreak of dengue hemorrhagic fever in Nicaragua, 1994]. *Rev Cubana Med Trop* **48**, 114-7.

Guzman, M. G., Vazquez, S., Martinez, E., Alvarez, M., Rodriguez, R., Kouri, G., de los Reyes, J. & Acevedo, F. (1996b). [Dengue in Nicaragua, 1994: reintroduction of serotype 3 in the Americas]. *Bol Oficina Sanit Panam* **121**, 102-10.

Hahn, C. S., Lustig, S., Strauss, E. G. & Strauss, J. H. (1988). Western equine encephalitis virus is a recombinant virus. *Proc Natl Acad Sci U S A* **85**, 5997-6001.

Hales, S., de Wet, N., Maindonald, J., Woodward, A. (2002). Potential effect of climate change on global distribution of dengue fever: an empirical model. *Lancet* **360**, 830-834.

Hall, B. (2001). Phylogenetic trees made easy. *Sinauer Associates*.

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

Halstead, S. B. (1997). Epidemiology of dengue and dengue hemorrhagic fever. *in Gubler DJ, Kuno G, (Eds.) Dengue and Dengue Hemorrhagic Fever.* Cab International, Wallingford UK.

Halstead, S. B. (2002). Dengue. *Curr Opin Infect Dis* **15**, 471-6.

Halstead, S. B. & Deen, J. (2002). The future of dengue vaccines. *Lancet* **360**, 1243-5.

Halstead, S. B., Larsen, K., Kliks, S., Peiris, J. S., Cardosa, J. & Porterfield, J. S. (1983). Comparison of P388D1 mouse macrophage cell line and human monocytes for assay of dengue-2 infection-enhancing antibodies. *Am J Trop Med Hyg* **32**, 157-63.

Halstead, S. B., O'Rourke, E. J. & Allison, A. C. (1977). Dengue viruses and mononuclear phagocytes. II. Identity of blood and tissue leukocytes supporting in vitro infection. *J Exp Med* **146**, 218-29.

Halstead, S. B., Shotwell, H. & Casals, J. (1973). Studies on the pathogenesis of dengue infection in monkeys. I. Clinical laboratory responses to primary infection. *J Infect Dis* **128**, 7-14.

Halstead, S. B., Streit, T. G., Lafontant, J. G., Putvatana, R., Russell, K., Sun, W., Kanesa-Thasan, N., Hayes, C. G. & Watts, D. M. (2001). Haiti: absence of dengue hemorrhagic fever despite hyperendemic dengue virus transmission. *Am J Trop Med Hyg* **65**, 180-3.

Halstead, S. B., Venkateshan, C. N., Gentry, M. K. & Larsen, L. K. (1984). Heterogeneity of infection enhancement of dengue 2 strains by monoclonal antibodies. *J Immunol* **132**, 1529-32.

Halstead, S. B. & Yamarat, C. (1965). Recent Epidemics of Hemorrhagic Fever in Thailand. Observations Related to Pathogenesis of a "New" Dengue Disease. *Am J Public Health Nations Health* **55**, 1386-95.

Hammon, W. M., Rudnick, A., Sather, G., Rogers, K. D. & Morse, L. J. (1960). New hemorrhagic fevers of children in the Philippines and Thailand. *Trans Assoc Am Physicians* **73**, 140-55.

Harris, E., Sandoval, E., Xet-Mull, A. M., Johnson, M. & Riley, L. W. (1999). Rapid subtyping of dengue viruses by restriction site-specific (RSS)-PCR. *Virology* **253**, 86-95.

Harris, E., Videa, E., Perez, L., Sandoval, E., Tellez, Y., Perez, M. L., Cuadra, R., Rocha, J., Idiaquez, W., Alonso, R. E., Delgado, M. A., Campo, L. A., Acevedo, F., Gonzalez, A., Amador, J. J. & Balmaseda, A. (2000). Clinical, epidemiologic, and virologic features of dengue in the 1998 epidemic in Nicaragua. *Am J Trop Med Hyg* **63**, 5-11.

Hartl, D. L., Clark, A.G. (1997). Presumed evolutionary benefit of recombination. *Principles of Population Genetics, 3rd Ed.* Sinauer (ed.) Sutherland, MA, 181.

Hawkes, R. A. (1964). Enhancement of the Infectivity of Arboviruses by Specific Antisera Produced in Domestic Fowls. *Aust J Exp Biol Med Sci* **42**, 465-82.

Henchal, E. A., Henchal, L. S. & Schlesinger, J. J. (1988). Synergistic interactions of anti-NS1 monoclonal antibodies protect passively immunized mice from lethal challenge with dengue 2 virus. *J Gen Virol* **69 (Pt 8)**, 2101-7.

Henchal, E. A., McCown, J. M., Seguin, M. C., Gentry, M. K. & Brandt, W. E. (1983). Rapid identification of dengue virus isolates by using monoclonal antibodies in an indirect immunofluorescence assay. *Am J Trop Med Hyg* **32**, 164-9.

Henchal, E. A., Repik, P. M., McCown, J. M. & Brandt, W. E. (1986). Identification of an antigenic and genetic variant of dengue-4 virus from the Caribbean. *Am J Trop Med Hyg* **35**, 393-400.

Hey, J. (1998). Selfish genes, pleiotropy and the origin of recombination. *Genetics* **149**, 2089-97.

Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* **266**, 383-402.

Hirst, G. K. (1962). Genetic recombiantion with Newcastle disease virus, poliovirus and influenza virus. *Cold Spring Harbor Symposiums in Quantitative Biology* **27**, 303-309.

Hober, D., Nguyen, T. L., Shen, L., Ha, D. Q., Huong, V. T., Benyoucef, S., Nguyen, T. H., Bui, T. M., Loan, H. K., Le, B. L., Bouzidi, A., De Groote, D., Drouet, M. T., Deubel, V. & Wattre, P. (1998). Tumor necrosis factor alpha levels in plasma and whole-blood culture in dengue-infected patients: relationship between virus detection and pre-existing specific antibodies. *J Med Virol* **54**, 210-8.

Hober, D., Poli, L., Roblin, B., Gestas, P., Chungue, E., Granic, G., Imbert, P., Pecarere, J. L., Vergez-Pascal, R., Wattre, P. & et al. (1993). Serum levels of tumor necrosis factor-alpha (TNF-alpha), interleukin-6 (IL-6), and interleukin-1 beta (IL-1 beta) in dengue-infected patients. *Am J Trop Med Hyg* **48**, 324-31.

Holder, M. & Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* **4**, 275-84.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982). Rapid evolution of RNA genomes. *Science* **215**, 1577-85.

Holmes, E. C. (2003). Molecular clocks and the puzzle of RNA virus origins. *J Virology* **77**, 3893-3897.

Holmes, E. C. & Burch, S. S. (2000). The causes and consequences of genetic variation in dengue virus. *Trends Microbiol* **8**, 74-7.

Holmes, E. C., Worobey, M. & Rambaut, A. (1999). Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol* **16**, 405-9.

Huang, Y. H., Lei, H. Y., Liu, H. S., Lin, Y. S., Chen, S. H., Liu, C. C. & Yeh, T. M. (2003). Tissue plasminogen activator induced by dengue virus infection of human endothelial cells. *J Med Virol* **70**, 610-6.

Huang, Y. H., Lei, H. Y., Liu, H. S., Lin, Y. S., Liu, C. C. & Yeh, T. M. (2000). Dengue virus infects human endothelial cells and induces IL-6 and IL-8 production. *Am J Trop Med Hyg* **63**, 71-5.

Hudson, R. R. & Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147-64.

Huerre, M. R., Lan, N. T., Marianneau, P., Hue, N. B., Khun, H., Hung, N. T., Khen, N. T., Drouet, M. T., Huong, V. T., Ha, D. Q., Buisson, Y. & Deubel, V. (2001). Liver histopathology and biological correlates in five cases of fatal dengue fever in Vietnamese children. *Virchows Arch* **438**, 107-15.

Hung, J. J., Hsieh, M. T., Young, M. J., Kao, C. L., King, C. C. & Chang, W. (2004). An external loop region of domain III of dengue virus type 2 envelope protein is involved in serotype-specific binding to mosquito but not mammalian cells. *J Virol* **78**, 378-88.

Hung, S. L., Lee, P. L., Chen, H. W., Chen, L. K., Kao, C. L. & King, C. C. (1999). Analysis of the steps involved in Dengue virus entry into host cells. *Virology* **257**, 156-67.

Hwang, C. K., Svarovskaia, E. S. & Pathak, V. K. (2001). Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching. *Proc Natl Acad Sci U S A* **98**, 12209-14.

Innis, B. L. & Eckels, K. H. (2003). Progress in development of a live-attenuated, tetravalent dengue virus vaccine by the United States Army Medical Research and Materiel Command. *Am J Trop Med Hyg* **69**, 1-4.

Innis, B. L., Nisalak, A., Nimmannitya, S., Kusalerdchariya, S., Chongswasdi, V., Suntayakorn, S., Puttisri, P. & Hoke, C. H. (1989). An enzyme-linked immunosorbent assay to characterize dengue infections where dengue and Japanese encephalitis co-circulate. *Am J Trop Med Hyg* **40**, 418-27.

Jacobs, M. & Levin, M. (2002). An improved endothelial barrier model to investigate dengue haemorrhagic fever. *J Virol Methods* **104**, 173-85.

Johnson, A. J., Guirakhoo, F. & Roehrig, J. T. (1994). The envelope glycoproteins of dengue 1 and dengue 2 viruses grown in mosquito cells differ in their utilization of potential glycosylation sites. *Virology* **203**, 241-9.

Judo, M. S., Wedel, A. B. & Wilson, C. (1998). Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res* **26**, 1819-25.

Katz, R. A. & Skalka, A. M. (1994). The retroviral enzymes. *Annu Rev Biochem* **63**, 133-73.

Kew, O., Morris-Glasgow, V., Landaverde, M., Burns, C., Shaw, J., Garib, Z., Andre, J., Blackman, E., Freeman, C. J., Jorba, J., Sutter, R., Tambini, G., Venczel, L., Pedreira, C., Laender, F., Shimizu, H., Yoneyama, T., Miyamura, T., van Der Avoort, H., Oberste, M. S., Kilpatrick, D., Cochi, S., Pallansch, M. & de Quadros, C. (2002). Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. *Science* **296**, 356-9.

Kim, M. J. & Kao, C. (2001). Factors regulating template switch in vitro by viral RNA-dependent RNA polymerases: implications for RNA-RNA recombination. *Proc Natl Acad Sci U S A* **98**, 4972-7.

King, A. D., Nisalak, A., Kalayanrooj, S., Myint, K. S., Pattanapanyasat, K., Nimmannitya, S. & Innis, B. L. (1999). B cells are the principal circulating mononuclear cells infected by dengue virus. *Southeast Asian J Trop Med Public Health* **30**, 718-28.

Kirkegaard, K. & Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. *Cell* **47**, 433-43.

Kliks, S. C., Nimmanitya, S., Nisalak, A. & Burke, D. S. (1988). Evidence that maternal dengue antibodies are important in the development of dengue hemorrhagic fever in infants. *Am J Trop Med Hyg* **38**, 411-9.

Kliks, S. C., Nisalak, A., Brandt, W. E., Wahl, L. & Burke, D. S. (1989). Antibody-dependent enhancement of dengue virus growth in human monocytes as a risk factor for dengue hemorrhagic fever. *Am J Trop Med Hyg* **40**, 444-51.

Kofler, R. M., Heinz, F. X. & Mandl, C. W. (2004). A novel principle of attenuation for the development of new generation live flavivirus vaccines. *Arch Virol Suppl*, 191-200.

Kouri, G., Guzman, M. G. & Bravo, J. (1986). Hemorrhagic dengue in Cuba: history of an epidemic. *Bull Pan Am Health Organ* **20**, 24-30.

Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G., Baker, T. S. & Strauss, J. H. (2002). Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717-25.

Kummerer, B. M., Tautz, N., Becher, P., Thiel, H. & Meyers, G. (2000). The genetic basis for cytopathogenicity of pestiviruses. *Vet Microbiol* **77**, 117-28.

Kuno, G. (1995). Review of the factors modulating dengue transmission. *Epidemiol Rev* **17**, 321-35.

Kuno, G. & Bailey, R. E. (1994). Cytokine responses to dengue infection among Puerto Rican patients. *Mem Inst Oswaldo Cruz* **89**, 179-82.

Kuno, G., Moore, C. G. & Sather, G. E. (1981). Factors affecting syncytial development in Aedes pseudoscutellaris cells by dengue viruses. *Am J Trop Med Hyg* **30**, 870-5.

LaFleur, C., Granados, J., Vargas-Alarcon, G., Ruiz-Morales, J., Villarreal-Garza, C., Higuera, L., Hernandez-Pacheco, G., Cutino-Moguel, T., Rangel, H., Figueroa, R., Acosta, M., Lazcano, E. & Ramos, C. (2002). HLA-DR antigen frequencies in Mexican patients with dengue virus infection: HLA-DR4 as a possible genetic resistance factor for dengue hemorrhagic fever. *Hum Immunol* **63**, 1039-44.

Lai, C. J. & Monath, T. P. (2003). Chimeric flaviviruses: novel vaccines against dengue fever, tick-borne encephalitis, and Japanese encephalitis. *Adv Virus Res* **61**, 469-509.

Lai, M. M. (1992). Genetic recombination in RNA viruses. *Curr Top Microbiol Immunol* **176**, 21-32.

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**, 545-51.

Lanciotti, R. S., Gubler, D. J. & Trent, D. W. (1997). Molecular evolution and phylogeny of dengue-4 viruses. *J Gen Virol* **78 (Pt 9)**, 2279-84.

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 (Pt 1)**, 65-75.

Lanciotti, R. S., Roehrig, J. T., Deubel, V., Smith, J., Parker, M., Steele, K., Crise, B., Volpe, K. E., Crabtree, M. B., Scherret, J. H., Hall, R. A., MacKenzie, J. S., Cropp, C. B., Panigrahy, B., Ostlund, E., Schmitt, B., Malkinson, M., Banet, C., Weissman, J., Komar, N., Savage, H. M., Stone, W., McNamara, T. & Gubler, D. J. (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* **286**, 2333-7.

Lee, E., Weir, R. C. & Dalgarno, L. (1997). Changes in the dengue virus major envelope protein on passaging and their localization on the three-dimensional structure of the protein. *Virology* **232**, 281-90.

Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., Villalobos, I., de, C., Ramos, C. & Rico-Hesse, R. (1999). Dengue virus structural differences that correlate with pathogenesis. *J Virol* **73**, 4738-47.

Lewis, J. A., Chang, G. J., Lanciotti, R. S., Kinney, R. M., Mayer, L. W. & Trent, D. W. (1993). Phylogenetic relationships of dengue-2 viruses. *Virology* **197**, 216-24.

Li, K., Chen, Z. & Plagemann, P. (1999). High-frequency homologous genetic recombination of an arterivirus, lactate dehydrogenase-elevating virus, in mice and evolution of neuropathogenic variants. *Virology* **258**, 73-83.

Li, T. & Zhang, J. (2001). Retroviral recombination is temperature dependent. *J Gen Virol* **82**, 1359-64.

Littaua, R., Kurane, I. & Ennis, F. A. (1990). Human IgG Fc receptor II mediates antibody-dependent enhancement of dengue virus infection. *J Immunol* **144**, 3183-6.

Liu, H. M., Zheng, D. P., Zhang, L. B., Oberste, M. S., Pallansch, M. A. & Kew, O. M. (2000). Molecular evolution of a type 1 wild-vaccine poliovirus recombinant during widespread circulation in China. *J Virol* **74**, 11153-61.

Loke, H., Bethell, D., Phuong, C. X., Day, N., White, N., Farrar, J. & Hill, A. (2002). Susceptibility to dengue hemorrhagic fever in vietnam: evidence of an association with variation in the vitamin d receptor and Fc gamma receptor IIa genes. *Am J Trop Med Hyg* **67**, 102-6.

Loke, H., Bethell, D. B., Phuong, C. X., Dung, M., Schneider, J., White, N. J., Day, N. P., Farrar, J. & Hill, A. V. (2001). Strong HLA class I--restricted T cell

responses in dengue hemorrhagic fever: a double-edged sword? *J Infect Dis* **184**, 1369-73.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-60.

Lorono-Pino, M. A., Cropp, C. B., Farfan, J. A., Vorndam, A. V., Rodriguez-Angulo, E. M., Rosado-Paredes, E. P., Flores-Flores, L. F., Beaty, B. J. & Gubler, D. J. (1999). Common occurrence of concurrent infections by multiple dengue virus serotypes. *Am J Trop Med Hyg* **61**, 725-30.

Lorono-Pino, M. A., Farfan-Ale, J. A., Rosado-Paredes, E. P., Kuno, G. & Gubler, D. J. (1993). Epidemic dengue 4 in the Yucatan, Mexico, 1984. *Rev Inst Med Trop Sao Paulo* **35**, 449-55.

Marchette, N. J., Halstead, S. B., Falkler, W. A., Jr., Stenhouse, A. & Nash, D. (1973). Studies on the pathogenesis of dengue infection in monkeys. 3. Sequential distribution of virus in primary and heterologous infections. *J Infect Dis* **128**, 23-30.

Marianneau, P., Flamand, M., Deubel, V. & Despres, P. (1998). Induction of programmed cell death (apoptosis) by dengue virus in vitro and in vivo. *Acta Cient Venez* **49** Suppl 1, 13-7.

Marovich, M., Grouard-Vogel, G., Louder, M., Eller, M., Sun, W., Wu, S. J., Putvatana, R., Murphy, G., Tassaneetrithep, B., Burgess, T., Birx, D., Hayes, C., Schlesinger-Frankel, S. & Mascola, J. (2001). Human dendritic cells as targets of dengue virus infection. *J Investig Dermatol Symp Proc* **6**, 219-24.

Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562-3.

Martinez-Barragan, J. J. & del Angel, R. M. (2001). Identification of a putative coreceptor on Vero cells that participates in dengue 4 virus infection. *J Virol* **75**, 7818-27.

Marz L, A. F., Staudacher E, Kubelka V (1995). Protein glicosylation in insects. *in Montreuil J (Ed)* Glycoproteins, 543-553.

Maynard Smith, J. & Smith, N. H. (1998). Detecting recombination from gene trees. *Mol Biol Evol* **15**, 590-9.

Messer, W. B., Gubler, D. J., Harris, E., Sivananthan, K. & de Silva, A. M. (2003). Emergence and global spread of a dengue serotype 3, subtype III virus. *Emerg Infect Dis* **9**, 800-9.

Miagostovich, M. P., dos Santos, F. B., de Simone, T. S., Costa, E. V., Filippis, A. M., Schatzmayr, H. G. & Nogueira, R. M. (2002). Genetic characterization of dengue virus type 3 isolates in the State of Rio de Janeiro, 2001. *Braz J Med Biol Res* **35**, 869-72.

Miagostovich, M. P., dos Santos, F. B., Gutierrez, C. M., Riley, L. W. & Harris, E. (2000). Rapid subtyping of dengue virus serotypes 1 and 4 by restriction site-specific PCR. *J Clin Microbiol* **38**, 1286-9.

Mills, D. R., Dobkin, C. & Kramer, F. R. (1978). Template-determined, variable rate of RNA chain elongation. *Cell* **15**, 541-50.

Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc Natl Acad Sci U S A* **100**, 6986-91.

Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313-9.

Monath, T. P. (1994). Dengue: the risk to developed and developing countries. *Proc Natl Acad Sci U S A* **91**, 2395-400.

Monath, T. P., Wands, J. R., Hill, L. J., Brown, N. V., Marciniak, R. A., Wong, M. A., Gentry, M. K., Burke, D. S., Grant, J. A. & Trent, D. W. (1986). Geographic classification of dengue-2 virus strains by antigen signature analysis. *Virology* **154**, 313-24.

Mongkolsapaya, J., Dejnirattisai, W., Xu, X. N., Vasanawathana, S., Tangthawornchaikul, N., Chairunsri, A., Sawasdivorn, S., Duangchinda, T., Dong, T., Rowland-Jones, S., Yenchitsomanus, P. T., McMichael, A., Malasit, P. & Screaton, G. (2003). Original antigenic sin and apoptosis in the pathogenesis of dengue hemorrhagic fever. *Nat Med* **9**, 921-7.

Moreno-Altamirano, M. M., Sanchez-Garcia, F. J. & Munoz, M. L. (2002). Non Fc receptor-mediated infection of human macrophages by dengue virus serotype 2. *J Gen Virol* **83**, 1123-30.

Morens, D. M., Halstead, S. B. & Marchette, N. J. (1987a). Profiles of antibody-dependent enhancement of dengue virus type 2 infection. *Microb Pathog* **3**, 231-7.

Morens, D. M., Sather, G. E., Gubler, D. J., Rammohan, M. & Woodall, J. P. (1987b). Dengue shock syndrome in an American traveler with primary dengue 3 infection. *Am J Trop Med Hyg* **36**, 424-6.

Morl, M. & Schmelzer, C. (1990). Group II intron RNA-catalyzed recombination of RNA in vitro. *Nucleic Acids Res* **18**, 6545-51.

Moumen, A., Polomack, L., Roques, B., Buc, H. & Negroni, M. (2001). The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res* **29**, 3814-21.

Muller, H. J. (1964). The Relation of Recombination to Mutational Advance. *Mutat Res* **106**, 2-9.

Mune, M., Rodriguez, R., Ramirez, R., Soto, Y., Sierra, B., Rodriguez Roche, R., Marquez, G., Garcia, J., Guillen, G. & Guzman, M. G. (2003). Carboxy-terminally truncated Dengue 4 virus envelope glycoprotein expressed in Pichia pastoris induced neutralizing antibodies and resistance to Dengue 4 virus challenge in mice. *Arch Virol* **148**, 2267-73.

Nagy, P. D. & Bujarski, J. J. (1995). Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers. *J Virol* **69**, 131-40.

Nagy, P. D. & Bujarski, J. J. (1996). Homologous RNA recombination in brome mosaic virus: AU-rich sequences decrease the accuracy of crossovers. *J Virol* **70**, 415-26.

Nagy, P. D., Pogany, J. & Simon, A. E. (1999). RNA elements required for RNA recombination function as replication enhancers in vitro and in vivo in a plus-strand RNA virus. *Embo J* **18**, 5653-65.

Nagy, P. D. & Simon, A. E. (1997). New insights into the mechanisms of RNA recombination. *Virology* **235**, 1-9.

Nagy, P. D., Zhang, C. & Simon, A. E. (1998). Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase. *Embo J* **17**, 2392-403.

Navarrete, J., Vasquez, J.L., Vasquez, J.A., Gomez, H. (2002). Epidemiologia del dengue y dengue hemorragico en el Instituto Mexicano del Seguro Social (IMSS). *Revista Peruana de Epidemiologia* **10**, 1-7.

Navarro-Sanchez, E., Altmeyer, R., Amara, A., Schwartz, O., Fieschi, F., Virelizier, J. L., Arenzana-Seisdedos, F. & Despres, P. (2003). Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. *EMBO Rep* **4**, 723-8.

Negroni, M. & Buc, H. (2001). Mechanisms of retroviral recombination. *Annu Rev Genet* **35**, 275-302.

Negroni, M., Ricchetti, M., Nouvel, P. & Buc, H. (1995). Homologous recombination promoted by reverse transcriptase during copying of two distinct RNA templates. *Proc Natl Acad Sci U S A* **92**, 6971-5.

Nobuchi, H. (1979). The symptoms of a dengue-like illness recorded in a Chinese medical encyclopedia. *Kanpo Rinsho* **26**, 422-425.

Novella, I. S., Hershey, C. L., Escarmis, C., Domingo, E. & Holland, J. J. (1999). Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J Mol Biol* **287**, 459-65.

Odelberg, S. J., Weiss, R. B., Hata, A. & White, R. (1995). Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I. *Nucleic Acids Res* **23**, 2049-57.

Panaviene, Z. & Nagy, P. D. (2003). Mutations in the RNA-binding domains of tombusvirus replicase proteins affect RNA recombination in vivo. *Virology* **317**, 359-72.

Pandey, B. D. & Igarashi, A. (2000). Severity-related molecular differences among nineteen strains of dengue type 2 viruses. *Microbiol Immunol* **44**, 179-88.

Papaevangelou, G. & Halstead, S. B. (1977). Infections with two dengue viruses in Greece in the 20th century. Did dengue hemorrhagic fever occur in the 1928 epidemic? *J Trop Med Hyg* **80**, 46-51.

Peyrefitte, C. N., Couissinier-Paris, P., Mercier-Perennec, V., Bessaud, M., Martial, J., Kenane, N., Durand, J. P. & Tolou, H. J. (2003). Genetic characterization of newly reintroduced dengue virus type 3 in Martinique (French West Indies). *J Clin Microbiol* **41**, 5195-8.

Pinheiro, F. P. (1989). Dengue in the Americas. 1980-1987. *Epidemiol Bull* **10**, 1-8.

Posada, D. (2002). Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* **19**, 708-17.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **98**, 13757-62.

Posada, D. & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* **54**, 396-402.

Posada, D., Crandall, K. A. & Holmes, E. C. (2002). Recombination in evolutionary genomics. *Annu Rev Genet* **36**, 75-97.

Pringle, C. R. (1965). Evidence of genetic recombination in foot-and-mouth disease virus. *Virology* **25**, 48-54.

Pugachev, K. V., Guirakhoo, F., Trent, D. W. & Monath, T. P. (2003). Traditional and novel approaches to flavivirus vaccines. *Int J Parasitol* **33**, 567-82.

Putnak, R., Fuller, J., VanderZanden, L., Innis, B. L. & Vaughn, D. W. (2003a). Vaccination of rhesus macaques against dengue-2 virus with a plasmid DNA vaccine encoding the viral pre-membrane and envelope genes. *Am J Trop Med Hyg* **68**, 469-76.

Putnak, R., Porter, K. & Schmaljohn, C. (2003b). DNA vaccines for flaviviruses. *Adv Virus Res* **61**, 445-68.

Putnam, J. L. & Scott, T. W. (1995). The effect of multiple host contacts on the infectivity of dengue-2 virus-infected Aedes aegypti. *J Parasitol* **81**, 170-4.

Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M. & Holmes, E. C. (2001). Human immunodeficiency virus. Phylogeny and the origin of HIV-1. *Nature* **410**, 1047-8.

Rannala, B. & Yang, Z. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol* **43**, 304-11.

Repik, P. M., Dalrymple, J. M., Brandt, W. E., McCown, J. M. & Russell, P. K. (1983). RNA fingerprinting as a method for distinguishing dengue 1 virus strains. *Am J Trop Med Hyg* **32**, 577-89.

Rico-Hesse, R. (1990). Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology* **174**, 479-93.

Rico-Hesse, R. (2003). Microevolution and virulence of dengue viruses. *Adv Virus Res* **59**, 315-41.

Rico-Hesse, R., Harrison, L. M., Nisalak, A., Vaughn, D. W., Kalayanarooj, S., Green, S., Rothman, A. L. & Ennis, F. A. (1998). Molecular evolution of dengue type 2 virus in Thailand. *Am J Trop Med Hyg* **58**, 96-101.

Rico-Hesse, R., Harrison, L. M., Salas, R. A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M. T., Nogueira, R. M. & da Rosa, A. T. (1997). Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* **230**, 244-51.

Rico-Hesse, R., Pallansch, M. A., Nottay, B. K. & Kew, O. M. (1987). Geographic distribution of wild poliovirus type 1 genotypes. *Virology* **160**, 311-22.

Robertson, D. L., Hahn, B. H. & Sharp, P. M. (1995). Recombination in AIDS viruses. *J Mol Evol* **40**, 249-59.

Rodhain, F., Rosen, L. (1997). Mosquito vectors and dengue virus-vector relationships. *in Gubler DJ, Kuno G, (Eds.) Dengue and Dengue Hemorrhagic Fever.* Cab International, Wallingford UK.

Roehrig, J. T. (2003). Antigenic structure of flavivirus proteins. *Adv Virus Res* **59**, 141-75.

Roehrig, J. T., Bolin, R. A. & Kelly, R. G. (1998). Monoclonal antibody mapping of the envelope glycoprotein of the dengue 2 virus, Jamaica. *Virology* **246**, 317-28.

Romanova, L. I., Blinov, V. M., Tolskaya, E. A., Viktorova, E. G., Kolesnikova, M. S., Guseva, E. A. & Agol, V. I. (1986). The primary structure of crossover regions of intertypic poliovirus recombinants: a model of recombination between RNA genomes. *Virology* **155**, 202-13.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4.

Rosen, L. (1977). The Emperor's New Clothes revisited, or reflections on the pathogenesis of dengue hemorrhagic fever. *Am J Trop Med Hyg* **26**, 337-43.

Rosen, L., Roseboom, L. E., Gubler, D. J., Lien, J. C. & Chaniotis, B. N. (1985). Comparative susceptibility of mosquito species and strains to oral and parenteral infection with dengue and Japanese encephalitis viruses. *Am J Trop Med Hyg* **34**, 603-15.

Rothman, A. L. (1997). Viral pathogenesis of dengue infections. *in Gubler DJ, Kuno G, Dengue and Dengue Hemorrhagic Fever (Eds.)* Cab International, Wallingford UK.

Rothman, A. L. (2003). Immunology and immunopathogenesis of dengue disease. *Adv Virus Res* **60**, 397-419.

Rothman, A. L. & Ennis, F. A. (1999). Immunopathogenesis of Dengue hemorrhagic fever. *Virology* **257**, 1-6.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-7.

Ruiz, B. H., Sanchez, I., Ortega, G., Lopez, I., Rosales, L. & Medina, G. (2000). Phylogenetic comparison of the DEN-2 Mexican isolate with other flaviviruses. *Intervirology* **43**, 48-54.

Rush, A. (1789). An account of the bilious remitting fever, as it appeared in Philadelphia in the summer and autumn of the year 1780. *Medical Inquires and Observations* Prichard & Hall, Philadelphia PA, 104-117.

Russell, P. K. & McCown, J. M. (1972). Comparison of dengue-2 and dengue-3 virus strains by neutralization tests and identification of a subtype of dengue-3. *Am J Trop Med Hyg* **21**, 97-9.

Sabchareon, A., Lang, J., Chanthavanich, P., Yoksan, S., Forrat, R., Attanath, P., Sirivichayakul, C., Pengsaa, K., Pojjaroen-Anant, C., Chambonneau, L., Saluzzo, J. F. & Bhamarapravati, N. (2004). Safety and immunogenicity of a three dose regimen of two tetravalent live-attenuated dengue vaccines in five- to twelve-year-old Thai children. *Pediatr Infect Dis J* **23**, 99-109.

Sabchareon, A., Lang, J., Chanthavanich, P., Yoksan, S., Forrat, R., Attanath, P., Sirivichayakul, C., Pengsaa, K., Pojjaroen-Anant, C., Chokejindachai, W., Jagsudee, A., Saluzzo, J. F. & Bhamarapravati, N. (2002). Safety and immunogenicity of tetravalent live-attenuated dengue vaccines in Thai adult volunteers: role of serotype concentration, ratio, and multiple doses. *Am J Trop Med Hyg* **66**, 264-72.

Sabin, A. B. (1952). Research on dengue during World War II. *Am J Trop Med Hyg* **1**, 30-50.

Sala, M. & Wain-Hobson, S. (2000). Are RNA viruses adapting or merely changing? *J Mol Evol* **51**, 12-20.

Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* **11**, 1423-5.

Sambrook, J., Frisch, E.F., Maniatis, T. (1989). Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory Press* **2nd ED.**

Sanchez, I. J. & Ruiz, B. H. (1996). A single nucleotide change in the E protein gene of dengue virus 2 Mexican strain affects neurovirulence in mice. *J Gen Virol* **77 (Pt 10)**, 2541-5.

Sangkawibha, N., Rojanasuphot, S., Ahandrik, S., Viriyapongse, S., Jatanasen, S., Salitul, V., Phanthumachinda, B. & Halstead, S. B. (1984). Risk factors in

dengue shock syndrome: a prospective epidemiologic study in Rayong, Thailand. I. The 1980 outbreak. *Am J Epidemiol* **120**, 653-69.

Sariol, C. A., Pelegrino, J. L., Martinez, A., Arteaga, E., Kouri, G. & Guzman, M. G. (1999). Detection and genetic relationship of dengue virus sequences in seventeen-year-old paraffin-embedded samples from Cuba. *Am J Trop Med Hyg* **61**, 994-1000.

Schierup, M. H. & Hein, J. (2000a). Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879-91.

Schierup, M. H. & Hein, J. (2000b). Recombination and the molecular clock. *Mol Biol Evol* **17**, 1578-9.

Schoepp, R. J. & Beaty, B. J. (1984). Titration of dengue viruses by immunofluorescence in microtiter plates. *J Clin Microbiol* **20**, 1017-9.

Scott, T. W., Clark, G. G., Lorenz, L. H., Amerasinghe, P. H., Reiter, P. & Edman, J. D. (1993). Detection of multiple blood feeding in Aedes aegypti (Diptera: Culicidae) during a single gonotrophic cycle using a histologic technique. *J Med Entomol* **30**, 94-9.

Seibert, S. A., Howell, C. Y., Hughes, M. K. & Hughes, A. L. (1995). Natural selection on the gag, pol, and env genes of human immunodeficiency virus 1 (HIV-1). *Mol Biol Evol* **12**, 803-13.

Shammas, F. V., Heikkila, R. & Osland, A. (2001). Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. *Clin Chim Acta* **304**, 19-28.

Sibold, C., Meisel, H., Kruger, D. H., Labuda, M., Lysy, J., Kozuch, O., Pejcoch, M., Vaheri, A. & Plyusnin, A. (1999). Recombination in Tula hantavirus evolution: analysis of genetic lineages from Slovakia. *J Virol* **73**, 667-75.

Stavrinides, J. & Guttman, D. S. (2004). Mosaic evolution of the severe acute respiratory syndrome coronavirus. *J Virol* **78**, 76-82.

Steinhauer, D. A. & Holland, J. J. (1987). Rapid evolution of RNA viruses. *Annu Rev Microbiol* **41**, 409-33.

Stephens, H. A., Klaythong, R., Sirikong, M., Vaughn, D. W., Green, S., Kalayanarooj, S., Endy, T. P., Libraty, D. H., Nisalak, A., Innis, B. L., Rothman, A. L., Ennis, F. A. & Chandanayingyong, D. (2002). HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. *Tissue Antigens* **60**, 309-18.

Sullenger, B. A. & Cech, T. R. (1994). Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* **371**, 619-22.

Sun, W., Edelman, R., Kanesa-Thasan, N., Eckels, K. H., Putnak, J. R., King, A. D., Houng, H. S., Tang, D., Scherer, J. M., Hoke, C. H., Jr. & Innis, B. L. (2003). Vaccination of human volunteers with monovalent and tetravalent live-attenuated dengue vaccine candidates. *Am J Trop Med Hyg* **69**, 24-31.

Suzuki, Y., Gojobori, T. & Nakagomi, O. (1998). Intragenic recombinations in rotaviruses. *FEBS Lett* **427**, 183-7.

Swofford, D. (1998). PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA: Sinauer Associates.

Tassaneetrithep, B., Burgess, T. H., Granelli-Piperno, A., Trumpfheller, C., Finke, J., Sun, W., Eller, M. A., Pattanapanyasat, K., Sarasombath, S., Birx, D. L., Steinman, R. M., Schlesinger, S. & Marovich, M. A. (2003). DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J Exp Med* **197**, 823-9.

Thein, S., Aung, M. M., Shwe, T. N., Aye, M., Zaw, A., Aye, K., Aye, K. M. & Aaskov, J. (1997). Risk factors in dengue shock syndrome. *Am J Trop Med Hyg* **56**, 566-72.

Tolou, H. J., Couissinier-Paris, P., Durand, J. P., Mercier, V., de Pina, J. J., de Micco, P., Billoir, F., Charrel, R. N. & de Lamballerie, X. (2001). Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *J Gen Virol* **82**, 1283-90.

Trent, D. W., Grant, J. A., Rosen, L. & Monath, T. P. (1983). Genetic variation among dengue 2 viruses of different geographic origin. *Virology* **128**, 271-84.

Troyer, J. M., Hanley, K. A., Whitehead, S. S., Strickman, D., Karron, R. A., Durbin, A. P. & Murphy, B. R. (2001). A live attenuated recombinant dengue-4 virus vaccine candidate with restricted capacity for dissemination in mosquitoes and lack of transmission from vaccinees to mosquitoes. *Am J Trop Med Hyg* **65**, 414-9.

Twiddy, S. S., Farrar, J. J., Vinh Chau, N., Wills, B., 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.

Twiddy, S. S. & Holmes, E. C. (2003). The extent of homologous recombination in members of the genus Flavivirus. *J Gen Virol* **84**, 429-40.

Twiddy, S. S., Holmes, E. C. & Rambaut, A. (2003). Inferring the rate and time-scale of dengue virus evolution. *Mol Biol Evol* **20**, 122-9.

Uchil, P. D. & Satchidanandam, V. (2003). Architecture of the flaviviral replication complex. Protease, nuclease, and detergents reveal encasement within double-layered membrane compartments. *J Biol Chem* **278**, 24388-98.

Uzcategui, N. Y., Camacho, D., Comach, G., Cuello de Uzcategui, R., Holmes, E. C. & Gould, E. A. (2001). Molecular epidemiology of dengue type 2 virus in Venezuela: evidence for in situ virus evolution and recombination. *J Gen Virol* **82**, 2945-53.

Uzcategui, N. Y., Comach, G., Camacho, D., Salcedo, M., Cabello de Quintana, M., Jimenez, M., Sierra, G., Cuello de Uzcategui, R., James, W. S., Turner, S., Holmes, E. C. & Gould, E. A. (2003). Molecular epidemiology of dengue virus type 3 in Venezuela. *J Gen Virol* **84**, 1569-75.

Van Dooren, S., Pybus, O. G., Salemi, M., Liu, H. F., Goubau, P., Remondegui, C., Talarmin, A., Gotuzzo, E., Alcantara, L. C., Galvao-Castro, B. & Vandamme, A. M. (2004). The Low Evolutionary Rate of Human T-Cell Lymphotropic Virus Type-1 Confirmed by Analysis of Vertical Transmission Chains. *Mol Biol Evol.*

van Vugt, J. J., Storgaard, T., Oleksiewicz, M. B. & Botner, A. (2001). High frequency RNA recombination in porcine reproductive and respiratory syndrome virus occurs preferentially between parental sequences with high similarity. *J Gen Virol* **82**, 2615-20.

Vazeille, M., Rosen, L., Mousson, L. & Failloux, A. B. (2003). Low oral receptivity for dengue type 2 viruses of Aedes albopictus from Southeast Asia compared with that of Aedes aegypti. *Am J Trop Med Hyg* **68**, 203-8.

Vorndam, V., Kuno, G. & Rosado, N. (1994). A PCR-restriction enzyme technique for determining dengue virus subgroups within serotypes. *J Virol Methods* **48**, 237-44.

Wain-Hobson, S. (1993). The fastest genome evolution ever described: HIV variation in situ. *Curr Opin Genet Dev* **3**, 878-83.

Wang, E., Ni, H., Xu, R., Barrett, A. D., Watowich, S. J., Gubler, D. J. & Weaver, S. C. (2000). Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J Virol* **74**, 3227-34.

Wang, W. K., Chao, D. Y., Lin, S. R., King, C. C. & Chang, S. C. (2003). Concurrent infections by two dengue virus serotypes among dengue patients in Taiwan. *J Microbiol Immunol Infect* **36**, 89-95.

Warke, R. V., Xhaja, K., Martin, K. J., Fournier, M. F., Shaw, S. K., Brizuela, N., De Bosch, N., Lapointe, D., Ennis, F. A., Rothman, A. L. & Bosch, I. (2003). Dengue virus induces novel changes in gene expression of human umbilical vein endothelial cells. *J Virol* **77**, 11822-32.

Watts, D. M., Porter, K. R., Putvatana, P., Vasquez, B., Calampa, C., Hayes, C. G. & Halstead, S. B. (1999). Failure of secondary infection with American genotype dengue 2 to cause dengue haemorrhagic fever. *Lancet* **354**, 1431-4.

Weaver, S. C. (1995). Evolution of alphaviruses. *In Gibbs A, Calisher CH, Garcia-Arenal F, Molecular Basis of Virus Evolution, Cambridge University Press, UK*, 501-530.

Weaver, S. C., Brault, A. C., Kang, W. & Holland, J. J. (1999). Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J Virol* **73**, 4316-26.

Weaver, S. C., Hagenbaugh, A., Bellew, L. A., Gousset, L., Mallampalli, V., Holland, J. J. & Scott, T. W. (1994). Evolution of alphaviruses in the eastern equine encephalomyelitis complex. *J Virol* **68**, 158-69.

Weaver, S. C., Kang, W., Shirako, Y., Rumenapf, T., Strauss, E. G. & Strauss, J. H. (1997). Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *J Virol* **71**, 613-23.

Wei, H. Y., Jiang, L. F., Fang, D. Y. & Guo, H. Y. (2003). Dengue virus type 2 infects human endothelial cells through binding of the viral envelope glycoprotein to cell surface polypeptides. *J Gen Virol* **84**, 3095-8.

Weiss, B. G. & Schlesinger, S. (1991). Recombination between Sindbis virus RNAs. *J Virol* **65**, 4017-25.

Whitehead, S. S., Hanley, K. A., Blaney, J. E., Jr., Gilmore, L. E., Elkins, W. R. & Murphy, B. R. (2003). Substitution of the structural genes of dengue virus type 4 with those of type 2 results in chimeric vaccine candidates which are attenuated for mosquitoes, mice, and rhesus monkeys. *Vaccine* **21**, 4307-16.

WHO (1999). Dengue haemorrhagic fever: diagnosis, treatment, prevention and control. *WHO, Geneve.*

Wierzchoslawski, R., Dzianott, A., Kunimalayan, S. & Bujarski, J. J. (2003). A transcriptionally active subgenomic promoter supports homologous crossovers in a plus-strand RNA virus. *J Virol* **77**, 6769-76.

Wittke, V., Robb, T. E., Thu, H. M., Nisalak, A., Nimmannitya, S., Kalayanrooj, S., Vaughn, D. W., Endy, T. P., Holmes, E. C. & Aaskov, J. G. (2002). Extinction

and rapid emergence of strains of dengue 3 virus during an interepidemic period. *Virology* **301**, 148-56.

Wiuf, C., Christensen, T. & Hein, J. (2001). A simulation study of the reliability of recombination detection methods. *Mol Biol Evol* **18**, 1929-39.

Worobey, M. & Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J Gen Virol* **80 (Pt 10)**, 2535-43.

Worobey, M., Rambaut, A. & Holmes, E. C. (1999). Widespread intra-serotype recombination in natural populations of dengue virus. *Proc Natl Acad Sci U S A* **96**, 7352-7.

Wu, S. F., Liao, C. L., Lin, Y. L., Yeh, C. T., Chen, L. K., Huang, Y. F., Chou, H. Y., Huang, J. L., Shaio, M. F. & Sytwu, H. K. (2003). Evaluation of protective efficacy and immune mechanisms of using a non-structural protein NS1 in DNA vaccine against dengue 2 virus in mice. *Vaccine* **21**, 3919-29.

Wu, S. J., Grouard-Vogel, G., Sun, W., Mascola, J. R., Brachtel, E., Putvatana, R., Louder, M. K., Filgueira, L., Marovich, M. A., Wong, H. K., Blauvelt, A., Murphy, G. S., Robb, M. L., Innes, B. L., Birx, D. L., Hayes, C. G. & Frankel, S. S. (2000). Human skin Langerhans cells are targets of dengue virus infection. *Nat Med* **6**, 816-20.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431-49.

Yazi Mendoza, M., Salas-Benito, J. S., Lanz-Mendoza, H., Hernandez-Martinez, S. & del Angel, R. M. (2002). A putative receptor for dengue virus in mosquito tissues: localization of a 45-kDa glycoprotein. *Am J Trop Med Hyg* **67**, 76-84.

Zanotto, P. M., Gould, E. A., Gao, G. F., Harvey, P. H. & Holmes, E. C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci U S A* **93**, 548-53.

Zhou, M. Y. & Gomez-Sanchez, C. E. (2000). Universal TA cloning. *Curr Issues Mol Biol* **2**, 1-7.

APPENDIX: ALIGNMENTS OF SEQUENCES.

In the following pages, the alignments of the sequences included in the phylogenetic analyses in chapter 3 of this dissertation are presented. They correspond to sequences of the E gene of DENV-1 to DENV-4 obtained in this study or in other works. The order of the sequences in each serotype is the same in which isolates appeared in the phylogenetic trees (Figures 3-3 to 3-6). This was done to facilitate visualizing the shared derived (synapomorphic) characters that define each clade in the tree. The complete name of each isolate is presented after an assigned number in the first page of each serotype. In the following pages only the assigned number is written preceding the sequence. The first three or four lines of numbers in each page correspond to the positions of each site in the sequence of E gene. They must be read vertically.

In order to keep this short, only the parsimony informative sites, which contain most of the phylogenetic information, are included. The complete sequences can be retrieved from GenBank where most sequences have already been submitted or will be submitted soon.

Alignment of DENV-1 parsimony informative sites

```
                                                  1111111111111111111222222
                                    1222233556677899900012334556666778890111111
                                   6961347364736387036589115670325681539210367 9

 1.MALAYSIA/72/P72 1244            ATGCACACGAGGTCGTCTCAAATCTCGTATCTATCTTACCTAAA
 2.THAILAND/54/TH SMAN             .CA....T..A......C......CT.C.ATCG....G......
 3.THAILAND/63/2543                .CA.......A......C......C..C.ATCG.T......G..
 4.THAILAND/64/16007               .CA.......A......C....C.CT.C.ATCG.T......G..
 5.AUSTRALIA/83/HCS1               GCA......T......T.T..GC.CT.C.C.CGC..C....G..
 6.INDONESIA/84/24 319             GCAA.....TA......CT..GC.CT.C.C.CGC....TT.G..
 7.INDONESIA/88/A88                GCAA.....T.......CT..GC..T.C.C.CGC..A.TT.G..
 8.PHILIPPINES/02/SA079            GCA......T.......CT..GC.C..C.A.C.C.......G..
 9.SRI LANKA95/SLMC28              GCA......T.......CT..GC.C..C.A.CGC..C...CG..
10.PHILIPPINES/74/228682           GCA......T.......CT..GC.CT.C.C.CGC..C....G..
11.NAURU/74/17646                  GCA......T.......CT..GC.CT.C.C.CGC.......G..
12.JAPAN/43/MOCHIZUKI              .CA..............C...G.TCT.C.C.CG...........
13.HAWAII/44                       .CA..............C...G.TCT.C.C.CG...........
14.CHINA99/GD05                    .CA.GTG...TA..A..C...GCTCTAC.C.CGC......C...
15.CAMBODIA/99                     ..A.G.G...TA..A..C...GCTC.AC.C.CGC......C...
16.LAOS/96/CH323                   ..A...G....A..A.TC...G.TCTAC...CGC......C...
17.DJIBOUTI98_606                  ..A...GT...A..ACTC...G.TCTAC.C.CGC..C.......
18.THAILAND/80/PUO359              ..A...G....A..A..C...G.TCTAC.C.CGC.C........
19.CHINA/80/GZ80                   ..A...G....A..A..C...G.TCTAC.C.CGC.C........
20.BURMA/76/PRS228686              .CA.....T.A....C.C.....TCTACGC..GC....T.....
21.IVORY COAST/98/1056             .CA.....T......C.C.GG.CT..A.GC..GC..A..T....
22.NIGERIA/68/IBH28328             .CA.....T......C.C.GG..T...CGC..GC..........
23.IVORY COAST/Dakar/A150          .CA.....T......C.C.GG..T.T.CGC..GC....T.....
24.JAMAICA/77/PRS288690            .CA.....T...CT.C.C.....TCT.CGC..GC...G......
25.COLOMBIA/96/371869              .CA..T..T...CT.C.C.....TCT.CGC..GC..........
26.COLOMBIA/85/347869              .C......T...CT.C.C.....TCT.CGC...C..........
27.BRAZIL/82/BE AR404147           .CA.....T...CT.C.C.....TCT.CGC..GC..........
28.ARGENTINA/00/293                ..A...G.T...CT.C.C...G.TCT.C.C..GC.........G
29.PERU/91/DEI0151                 ..A.....T...CT.C.C...G.TCT.CGC..GC.........G
30.MX/CAMPECHE/95/BC264            .CA.....T.A.CT.C.C.....TCT.CGC...C..........
31.MX/YUCATAN/94/BC126             .CA.....T.A.CT.C.C.....TCT..GC..GC..........
32.MX/COZUMEL/95/BC7               .CA.....T.A.CT.C.C.....TCT.CGC..GC..A.......
33.BRAZIL/97/BeH58452 6            .CA.....T..ACT.C.C.....TCT.CGC..GC..C.......
34.BRAZIL/01/BR01 MR               .CA.....C..ACT.C.C.....TCT.CGC..GC..C.......
35.VENEZUELA/94 125239             .CA.....T...CT.C.C.....TCT.CGC..GC..........
36.VENEZUELA/97 28164              .CA.....T...CT.C.......TCT.CGC..GC..........
37.MX/82/1379                      .CA....TT...CT.C.C.....TCT.CGC..GC........C.
38.MX/PUEBLA/84/1462               .CA....TT...CT.C.C.....TCT.CGC..GC........C.
39.MX/83/1378                      .CA....TT...CT.C.C.....TCT.CGC..GC........C.
40.MX/YUCATAN/80/1298              .CA....TT...CT.C.C.....TCT.CGC..GC..........
41.MX/SONORA/84/1463               .CA....TT...CT.C.C.....TCT.CGC..GC..........
42.MX/86/1756                      .CA....TT...CT.C.C.....TCT.CGC..GC..........
```

Alignment of DENV-1 parsimony informative sites

```
  222222222222222222333333333333333333333334444444444444444444455555
  223344455666777889011112233334445577899900122223445556667888011244
  254703458247069587625891403670281735104935703698140392457239701803

 1 CAAACACAAGGCTCAGTACCTACACATCTAATTTCAATTACAGACCCTCAGATGCTACTCCGTCTC
 2 T.......G.......CG...G..T...C...C.......T.....TA.G.G............T
 3 T.........A.....C....G.....TC...C.......T......ATG.G............CT
 4 T..............C....G.....TC...C..............ATG.G........T...CT

 5 T....G...A...T..C..T.G.............G.C..T......A.GAT...........A.A
 6 TG...G.G.A...T..C..T..............TG.C..T..CT..A.GAT...C.T.....A.A
 7 TG...G...A...T..C..T...........C..TG.C.....CT..A.GAT...C.T...A.A.A
 8 T...TG.GGA.....AC..T.G.............G.C..T.C...AA..AT...C....T..A.A
 9 T....G.GGA......C..T.G.............G.C..T.C...AA..AT...C..C....A.A
10 T....G..GA...T..C..T.G......C......G.C..T.....AA.GAT...C.......A.A
11 T....G..GA...T..C..T.G.............G.C..T.....AA.GAT...C.......A.A
12 T...........CT.......G.T....C....C.GG..GTG.....A.G.G.....T..T....T
13 T.G.........CT.......G.T....C....C.G...GTG.......G.G...C.T..T....T
14 T.....T......T.......G.C..A.C....C.G....TGA....A.G.GCAGCG..TT.A..T
15 T...........T....T..G.C..A.C....C.G....TGA....A.G.G.AG.G..TT.A..T
16 T..........T.T.......G.T....C....C.G....TGA....A.G.G.AGC..AT.....T
17 T..........T.T.......G.T....C....CTG....TGA....A.G.G.AGCG.A......T
18 .......G...T.T.......G.C....C....C.G....TGA....A.G.G.AGCG.AT.....A
19 T..........T.T.......G.T....C....C.G....TGA....A.G.G.AGCG.AT.....T
20 T...T........T..C.T.......A.CCG..C.T....T....T.G.G.G.....T......C.
21 T.....T.....CT..C.T.......A.CTG..C..G..........G.G.G...C.T..TA..C.
22 T............T..C.T.......A.CCG..C............TG.G.G...C.T..T...C.
23 T.........T..T.AC........GA.CCG.CC............TTG.G.G...C.T..T...C.
24 T..GT........T...GT.C...T.A.CCG..C.G....T....T.G.G.GC....T..T...C.
25 T..GT........T....TTC...T.A.CCG..C......T....T.G.G.GC....T......C.
26 T..G........CT....T.C...T.A.CCG..C......T....T.G.G.GC....T..T...C.
27 T..GT.....A..T....T.C...T.A.CCG..C......T...TT.G.G.GC....T..T...C.
28 T..G.....T...T....T.C...T.A.CCG..C......T....T.G.G.GC....T.TT..TC.
29 T..GT........T....T.C...T.A.CCG..C......T....T.G.G.GC....T..T..TC.
30 T..GT........T....T.C...T.A..CG..C.G..C.T....T.G.GAGC....T..T...C.
31 T..GT........T....T.C...T.A..CG....G..C.T....T.G.G.GC....T..T...C.
32 T..GT........T....T.C...T.A..CG....G..C.T....T.G.G.GC....T..T..TC.
33 T..GT........T....T.C.T...A.CTG.CC......T......GTG.GC....T..T..TC.
34 T..GT........TC...T.C.T...A.CTG.CC......T......GTG.GC....T.....TC.
35 T..GT........TC...T.C....GA.CTG.CC......T....T.G.G.GC....T..T...C.
36 T..GT........TC...T.C....GATCTG.CC......T....T.G.G.GC..C.T..T...C.
37 T..GT........T..A.T.C...T.A.CCG..C......T....T.G.G.GC....T......C.
38 T..GT........T....T.C...T.A.CCGC.C......T....T.G.G.GC....T......C.
39 T..GT........T....T.C...T.A.CCG..C......T....T.G.G.GC....T......C.
40 T..G.........T....T.C...T.A.CCG..C......T....T.G.G.GC....T..T...C.
41 T.CG.........T....T.C...T.A.CCG..C......T....T.G.G.GC....T..T...C.
42 T.CG...G.....T....T.C..GT.A.CCG..C......T....T.G.G.GC....T..T...C.
```

Alignment of DENV-1 parsimony informative sites

```
   5555555555555556666666666666666667777777777777777788888888888888888
   4455667777889991224566677778889900122334555556899911133334444567778
   7925470369252455176236923891470658409281023698605803901473679540690

 1 TGTTAAGACTTGTGTATCCCCGGTTATCGGTCCGTTCTGGATCAAATTTACAACATTACCGCATGTT
 2 .A......T...C.....T..T......A...TAC.T....C.....C..T............CA.C
 3 .A..G...T.C...C...T..T......A...TAC.T.........A.........C...A..CA..
 4 .A..G...T.C...C...T..T......A...TAC.T..........C.........C...A..CA..
 5 C.C.............C.T.........A..T..C.T.....T.G.......T..............
 6 C...............C.T..A.C....A..T..C.T...............T....C........C
 7 C..C...G.C......C.T..A......A..T..C.T.....T.........T........T.....
 8 ......A.........C......C....A..T..C.T.A........C....T............C.
 9 ................C.T.........A..T..C.T.A........C....T............C.
10 ................C.T.........AA.TT.C.T.A............T.TA.............
11 ................C.T.........A..T..C.T.A.............T..............
12 ...C..............T...A...C.A...T...T....C.....CC...TA.....T.T.....
13 ...C..............T...A...C.A...T...T....C......C...TA.....T.T.C.C.
14 ......A....A....A.....A....AA...T.G.T...........C...T......T.TGC..C
15 ......A....A....A.....A....AAA..T.G.T......T.G..C...T......T.TGC..C
16 ...C..A..C.A....A.....A.AG.AA...T.GCT......G.G..C..CT......T.T.C..C
17 ..CC..A..C.A....A.....A.AG.AA...T.ACT...G.......C...T.G....T.T.C..C
18 ...C..A..C.A....A.....A.AG.AA...T.A.T...........C...T......T.T.C..C
19 ...C..A..C.A....A.T...A.AG.AA...T.A.T...........C...TA.....T.T.C..C
20 ..CC...G....C.CG..TT.......TA...T.C.TC..G..GG....G.G...C........C..
21 ..CC.G.G..C.C..G.TT........TA.C.T...........G....GTG...C..T.....A..
22 ..CC...G....C..G.TT........TA...T.C.T.......G....G.G...C..T........
23 ..CC...G.......G..TT.....G.TAA..T.C.T...G...G....G.G...C..T........
24 ..CC...G....CA....T.T.......A...T.CCT..A....GG...G.G...C...........
25 ..CC.G.G....CA....T.TA...T..A...T.CCTC.A....GG...G.G...C.......C...
26 ..CC...G....CA....T.T.......A.....CCT..A....GG...G.G...C...........
27 ...C...G.C..CA....T.T.......A...T.CCT..A....GG...G.G...C..T........
28 ..CC...G....CA....T.TAA....TA...T.CCT..A....GG...G.G..GC...........
29 ..CC...G....CA....T.T.A....TA...T.CCT..A....GG...G.G...C........A..
30 ..CC...G....CA....T.TA......A.....CCT..A....G....G.G...C.....T.....
31 ..CC...G....CA....T.T.............CCT..A....G....G.G...C...........
32 ..CC...G....CA....T.T.............CCT..A....G....G.G...C...........
33 ..CC...G....CA....T.T.A.....A...T.CCT..A....GGC..G.G...C.G.........
34 ..CC...G....CA....T.T.......A...T.CCT..A....G.C..G.G...C.G.........
35 ..CC...G....CA....T.T.......A.C.T.CCT..A....GG...G.G...C...........
36 ..CC...G....CA....T.T.......A.C.T.CCT..A....GG...G.G...CC..........
37 ..CC...G....CA....T.T.......A...T.CCT..A....GG...G.G...C...........
38 ..CC...G....CA....T.T.......A...T.CCT..A....GG...G.G...C...........
39 ..CC...G....CA....T.T.......A...T.CCT..A....GG...G.G...C...........
40 ..CC...G....CA....T.T.......A...T.CCT..A.C..GGC..G.G...CC..........
41 ..CC...G....CA....T.T.......A...T.CCT..A.C..GG...G.G...C...........
42 ..CC...G....CA....T.T.......A...T..CT..A.C..GG.C.G.G...C...........
```

Alignment of DENV-1 parsimony informative sites

```
                                  1111111111111111111111111111111111111111111
   8888899999999999999999900000000000000000000011111111111111111111
   8889901112233566667779900111122333556667788990001112333344556788 9
   5890102382703701360280358135669457132584503384570368124839267925 0

 1 AGATGGCTCTAAACTCAGGTTTACCTCACGATACTGCCCCACAGCGGAAATCCACGATAGAGGAG
 2 ..........G....TG....C.............A......G.T.....C...........A.T
 3 ..........G..T..G....C.T.......C...A........T.....C...........A..
 4 ..........G..T..G....C.T.......C...A........T.....C...........A..
 5 ..G.AA...C.G.T...A..C....C.T..G....A............GG...............
 6 .A..AA.C..GG.T...A..C....C.T..G................GG...............
 7 .A.CAA.C..GG.T...A.CC....C.T..GC...............GG...........G....
 8 ..G.AAG..CGG.T......C....C.T..G..............A.GG....CT..........
 9 ..G.AAG..CGG.T......C....C.T..G...C..........A.GG....CT..........
10 .A..AAG...GG.T......C....C.T..G..............A.GG.....T..........
11 .....AG...GG.T......C....C.T..G..............A.GG.....T..........
12 ..........G..T......C............T..T..............T..........A..
13 ..........G..T......C.................................T..........A..
14 .........CG.......A.C...T....A......T............G.T...A......A..
15 .........CG..T....A.C...T....A......T............G.T...A......A..
16 .........CG.G.......C...T....AG.TTG...T...G......C.T..........A.C
17 .........CG.G.C.....C...T....A....G...T...G........T..........A.C
18 .........CG.G.......C...T....A....G...T...G........T..........A.C
19 .........CG.G.......C...T....A...TG...T...G........T..........A.C
20 ........T.G...C.....C..A........G....T.T......A........AG.....A..
21 ........T.GG...T....C.......T...G....T.T.T....A.....T...G..A.AA..
22 ........T.GG...T....C.......T...G....T.T......A.....T...G..A.AA..
23 .........CGG...T..........T.T...G....T.T......A.....T..AG..A.AA.A
24 ........TCGG.T......C.G.........G....T.T......A.........G.....A..
25 ........TCGG.T......C.G.........G.....TT...A..A...C...........A..
26 ........TCGG.T......C.G.........G....T.T......A.........G.....AG.
27 ..G.....TCGG.T......C.G.........G....T.T......A.........G.....A..
28 .....A..TCGG.T......C.G.T..G....G..............A.........G...G.AG.
29 .....A..TCGG.T......C.G.T..G....G..............A.........G...G.AG.
30 ........TCGG.TC....CC.G...T.....G....T.T......A.........G.G...A..
31 ........TCGG.TC.....C.G.T.T.....G....T.T......A.........G.G...A..
32 ........TCGG.TC.....C.G...T.....G....T.T......A.........G.G...A..
33 G..C....TCGG.T..G..CC.G.........G....T.T......A.........G.....A..
34 G..C....TCGG.T.....CC.G...T.....G....T.T......A.........G.....A..
35 G..C....TCGG.T.....CCCG.........G....T.T......A.........GC....A..
36 G.......TCGG.T.....CCCG.........G....T.T......A.........GC....A..
37 ........TCGG.T......C.G.........G....T.TG..A..A.........G.....A..
38 ........TCGG.T......C.G.........G....T.TG.....A.........G.....A..
39 ....A...TCGG.T......C.G.........G....T.TG.....A.........G.....A..
40 ........TCGG.T......C.G.........G....T.TG.....A.........G.....A..
41 .A......TCGG.........C.G.....T...G....T.TG.....A.........G.....A..
42 .A......TCGG.........C.G.....T...G....T.TGT....A.........G.....A..
```

Alignment of DENV-1 parsimony informative sites

```
   1111111111111111111111111111111111111111111111111111111111111111111
   1222222222222222222222223333333333333333333333344444444444444444444
   9000112334445566778899990000122223334455668899011112233344555678 8
   1469591092352469281703463568534695894706281368756792834703025192 5

 1 CTCACCAACCTGACCTTAATGTGGCGGAAAATAGTTCCCCCAACCGTGCATCGCACCAAGCGTAG
 2 ..T.T.....C........C..A.....G....A..TTTT...T.A.A..G..TG........G.
 3 ..T.T....TC........C..A.....G....A..T.AT...T.A.A..G..TG........GA
 4 ..T.T....TC........C..A.....G....A..T.AT...T.A.A..G..TG........GA
 5 T.TGT.T...C..T.....C.......G....GA....TT...T.A....G.A.GT.......C.
 6 T.T.T.T...C....A...C............GA....TTA..T.AC...G.A.GT......CG.
 7 T.T.T.T..TC........C............GA....TTT..T.AC.T.GTA.GT......CG.
 8 T.T.T.T...C.......GC......A..G.CGAC...GA...T.A....G.A.GT.......G.
 9 T.T.T.T...C.......GC......A..G..GA....GT...T.A....G.A.GT..T....G.
10 T.T.T.T...C.......GC......A..G..GA....TT.G.T.AC...G.A.GT.......G.
11 T.T.T.T...C..T....GC......A..G..GA....TT.G.T.AC...G.A.GT.......G.
12 ..T.T.....C...TC...C.............A....T.T..T.A..T.G.....T...T..G.
13 .CT.T.....CA.......C.............A....T.T..T.A...GG...G.T...T..G.
14 ..TGTA...AC...T..G.C...TTA.......A....T...GT.AC...G...G.T.CAAA.G.
15 ..TGTA...AC...T....C...TTA.......A....T...GT.A....G...GTT.CAAA.G.
16 ..TG.....AC...T.C..CA..T.A.......A....T...GT.AC...GT..G.T.C.A..G.
17 ..TGT....AC...T.C..CA..T.A.......AC.......GT.AC...G...G.T.C.A..G.
18 ..TGT....AC...T....CA..T.A.......A....T...GT.AC...G...G.T.C.A..G.
19 ..TGT....AC...T....CA...TA.G.....A....T.T.GT.AC...G...G.T.C.A..G.
20 ...................CA...T...G.GC......TT...T.A....GA..G.T........
21 ...................C........G..C......TT.....A...GGA..G.T....A...
22 .C......TT.A.......C........G..C.....TTT.....A....GA..G.T........
23 ...G....TT.A.....G.C........G..C.....TTT.....A....GA..G.T........
24 ........T.C....C...C....T...G..C...C..TT...TT.....GA..G.T...T....
25 T.......T.C....CC..C....T...G.GC...C...T...TT.....GA..G.T........
26 ........T.C....C...CA.A.T...G..C...C..TT...TT.....GA..G.T........
27 T.......T.C....C...C....T...G..C...C..TT...TT.....GA..G.T........
28 ....T...T.C...TC...CAA.AT...G.GC...C..TT...TT.....GA..G.TG.......
29 ........T.C...TC...C.A..T...G..C...CT.TT...TT.....GA..G.TG.......
30 ....T..GT.C...TC...C....T...G..C...C..TT...TT.....GA..G.T......G.
31 .......GT.C...TC...C....T...G..C...C..TT...TT.....GA..G.T...T..G.
32 .......GT.C...TC...C....T...G..C...C..TT...TT.....GA..G.T......G.
33 ....T...T.C....C...CA...T...G..C.T.C..TT...TT.....GA..G.T...T....
34 ........T.C....C...CA...T...G..C.T.CT.TT...TT.....GA..G.T...T....
35 ........T.C....C...CA...T...G..C...C..TT...TT.....GA..G.T...T....
36 ........T.C.........A...T...G..C...C..TT...TT.....GA..G.T........
37 ........T.C.G..C...C....T...G..C...C..TT...TT.....GA..G.T........
38 ........T.C.G..C...C....T...G..C...C..TT...TT.....GA..G.T........
39 ........T.C.G..C...C....T...G..C...C..TT...TT.....GA..G.T........
40 ........T.C....C...C....T...G..C...C..TT...TT.....GA..G.T........
41 ........T.C....C...C....T...G..C...C..TT...TT.....AA..G.T........
42 ........T.C....C...C....T...G..C...C..TT...TT.....AA..G.T........
```

Alignment of DENV-2 parsimony informative sites

```
                                                    11111111111111111111
                                  11222333344556677778999900122233455556666777778
                                 69584570369254769235870369581136577367925681467O

 1 MALAYSIA/70/P8 1407           CCCACAACTGGATACCCCCAACTCAAAGCCACGTAAACAAACTGATCC
 2 GUINEA/81/PM33974             AT...C..CA.....TT.T...C...C..T..A.GG.....TGA....
 3 NIGERIA/66/IBH11208           AT...C.........T..T...C...C..TG.A.GG.....TGA....
 4 SENEGAL/90/DAKHD10674         AT.G.C......A..TT.T...C...C..TG.A.GG.....TGA...T
 5 THAILAND/64/16681             T.AGT.....AG....A.T....GGG.A.TGT....G.TCC.A.GA.T
 6 THAILAND/93/ThNH7             T.A.T....AA.....A.T....GGG.A.TGT....G.TTC.A.GA..
 7 MALAYSIA/87/M1                T.A.T...AAAG....A.T....GGG.A.TGT....G.TCC.A.GA.T
 8 TAIWAN/87                     T.A....T.AAG....A.T....GG.CA.TGT......TCT.A.GA.T
 9 PHILIPPINES/83                T.A....T.AAG....A.T....GG.CA.T.T......TCT.A.GA.T
10 CUBA/81/A15                   T.A.T....AAG....A.T....GGG.A.TGT......TCT.A.GA.T
11 VIETNAM/97/CTD113             T.A.T....AAG....A.T....GGG.A.TGT......TCT.A.GA.T
12 CHINA/87/43                   T.A.T....AAG....A.T....GGG.A.TGT......TCT.A.GA.T
13 NEW GUINEA/44/NGC             T.A.T....AAG....A.T....GGG.A.TGT......TCT.A.GA.T
14 MX/GUERRERO/97/C932           T.A.T....AAG....A.T....GGG.A.TGT.....TTCT.A.GA.T
15 MX/GUERRERO/97/C1077          T.A.T....AAG....A.T....GGG.A.TGT.....TTCT.A.GA.T
16 THAILAND/80/D80/141           T.A.T....AAG....A.T..T.GGG.A.TGT......TCT.A.GA.T
17 CHINA/85/04                   T.A.T....AAG....A.T..T.GGG.A.TGT......TCT.A.GA.T
18 VIETNAM/98/CT 29              T.A.T....AAG....A.T..T.GG..A.TGT......TCT.A.GA.T
19 JAMAICA/83/1409               T.A.T....AAG....A.T..T.GGG.ATTGT......TCT.A.GA.T
20 VENEZUELA/90/MARA4            T.A.T...CAAG....A.T..T.GGG.ATTGT......TCT.A.GA.T
21 BRAZIL/90/40247               T.A.T...CAAG....A.T..T.GGG.ATTGT......TCT.A.GA.T
22 COLOMBIA/96/PTCOL96           T.A.T...C.AG...TA.T..T.GGG.ATTGT......TCT.A.GA.T
23 VENEZUELA/97/1657             T.A.T...CAAG....A.T..T.GGG..TTGT......TCT.A.GA.T
24 MARTINIQUE/88/703             ..A.T...CAAG....A.T..T.GGG.A.TGT......TCT.A.GA.T
25 NICARAGUA/99/541              ..A.T...CAAG....A.T..T.GGG.A.TGT......TCT.A.GA.T
26 MX/YUCATAN/02/13381           ..A.T...CAAG....A.T..T.GGG.A.TGT......TCT.A.GA.T
27 MX/YUCATAN02/13404            ..A.T...CAAG....A.T..T.GGG.A.TGT......TCT.A.GA.T
28 MX/OAXACA/00/468              ..A.T....AAG....A.T..T.GGG.A..GT......TCT.A.GA.T
29 MX/YUCATAN/01/12914           ..A.T...CAAG....A.T..T.GGG.A..GT.C....TCT.A.GA.T
30 MX/YUCATAN/01/12021           ..A.T...CAAG....A.T..T.A.G.A..GT.C....TCT.A.GA.T
31 SOMALIA/94/S9                 TTA.T.....AG....A...G..GGG.AT.GT......CCT.A.GATT
31 INDONESIA/76                  TTA.T.....AG....A.T....GGG.A.TGT......TCT.A.GATT
33 MX/YUCATAN/96/BC17            TTA.T.....AG....A.T....GGG.ATTGT......CCT.A.GATT
34 AUSTRALIA/93/TSV01            TTA.T.G...AG....A.TG...GGGGATTGT....T.CCT.A.GATT
35 BORNEO/88/620                 TTA.T.....AG....A.T....GGGGA.TGT....T.CCT.A.GATT
36 SEYCHELLES/77/SEY42           TTA.T..TC.AG....A.T....G.G.A.TGT...GG.TCC.A.GA.T
37 SRI LANKA/85/1592             TTA.T..TC.AG....A.T....G.G.A.TGT...GG.TCC.A.GA.T
38 UGANDA/93/CAMR11              TTA.T.....AG....A.T....G.G.A.TG....G..TCT.A.GA.T
39 INDIA/94/CAMR10               TTA.T..T..AG....A.T....G.G.A.TG....G..TCT.A.GA.T
40 SAUDI ARABIA/92/CAMR16        TTA.T.....AG....A.T....G.G.A.TG....G..TCT.A.GA.T
41 TRINIDAD/53/TRI1751           .TA.T.G.C.AG....A.T....GGG.A..G.......CCT.A.GA..
42 INDIA/57                      ..A.T.G...A.GGT.A.T..T.GGG.AT.G.......CCT.A.GA.T
43 PUERTO RICO/69/159            ..A.T.G...A.GGT.ATT..T.GGG.AT.G.......CCCTA.GA.T
44 MX/83/200787                  ..A.T.G...A.GGT.ATT..T.GGC.AT.G.......CCCTA.GA.T
45 PUERTO RICO/77/1328           ..A.T.G...A.GGT.ATTG.T.GGG.AT.G.......CCCTA.GA.T
46 PERU/96/IQT2913               ..A.T.G...A.GGT.ATT..T.GGG.ATTG.......CCCTA.GA.T
47 MX/84/1482                    ..A.T.G...A.GGT.ATT..T.GGG.AT.G.......CCCTA.GA.T
48 VENEZUELA/87/VEN2             ..A.T.G...A.GGT.ATT.GT.GGG.AT.G.......CCCTA.GA.T
49 MX/83/1421                    ..A.T.G...A.GGT.ATT.GT.GGG.AT.G.......CCCTA.GA.T
50 MX/94/Q.ROO/BC139             ..A.T.G...A.GGT.ATT..T.GGG.AT.G.......CCCTA.GA.T
51 MX/92/SONORA/131              ..A.T.G...A.GGT.ATT..T.GGG.AT.G.......CCCTA.GA.T
52 MX/95/TAMAULIPAS/328298       ..A.T.G...A.GGT.ATT..T.GGG.AT.G.......CCCTA.GA.T
```

Alignment of DENV-2 parsimony informative sites (continued)

```
   1111112222222222222222222222222222223333333333333333333333333444444444444
   8889990011111234444444556667777788999001112233355666677788899000011112222
   3692351402369840236789281470138925147065891706917456926815723235814570l9

 1 TACATACGGAATTCTCGCAGTGGACAGCGCAACGGTATGTTCGGATTGCTTGGCGAGGTAATTAACCAACAT
 2 .....G.AA..CC....T...A.G..A..TG.T.....T......C.......A.GA...GCC.G......C
 3 A....G.A...CC.......CA.G..A..TGGT.....T..............A.GA...GCC.G......C
 4 A....G.A....C........A.G..A..TGGT.A...T....A.C.......A.GA...GCC.G......C
 5 AGAGC..AA...CAG....AATAG.........CAC.AA.A.AAGC.T.AAA....A....C.G....TG..
 6 A.AGC..AA....AG....AAAAGTG.......CAC.AA.ATAAGC.T.AAA.....AC..C.G....TGG.
 7 A.AGCG.AA....AG....AATAGTG.A.....CAC.AA.A.AAGC.T.AAA....AAC..C.G....TGG.
 8 AGAG...AA...CAG....AATAGT........CAC.AA.A.AA.C.TTAAA....AAC..CCG.T..CG..
 9 AGAG...AA...CAG....AATAG.........CAC.AA.A.AA.C.T.AAA.....AC..C.G.T..CG.C
10 AGAGCG.AA...CAA....AATAG.........CAC.AA.A.AA.C.T.AAA..A.A.CTGC.G....TG..
11 AGAGCG.AA...CAA....AATAG.........C.C.AA.ATAA.C.T.AAA..A.A.C..C.G....TG..
12 AGAGCG.AA.T.CAA....AATAG.........C.C.AA.ATAA.C.T.AAA..A.A.C..C.G....TG..
13 AGAGCG.AA.T.CAA....AATAG.........C.C.AA.ATAA.C.T.AAA..A.A.C..C.G....TG..
14 AGAGCG.AA...CAA....AATAG.........C.C.AA.ATAA.C.T.AAA..A.A.C..C.G....TG..
15 AGAGCG.AA...CAA....AATAG.........C.C.AA.ATAA.C.T.AAA..A.A.C..C.G....TG..
16 A.AGCGTAA....AG..T.AATAG.........CACGAA.ATAAGC.TTAAA....AAC....G....CG..
17 A.AGCGTAA....AG..T.AATAG.........CACGAA.ATAAGC.TTAAA....AAC....G....CG..
18 A.AGCGTAA....AG..T.AATAG.....T...CACGAA.ATAA.C.TTAAA....AAC....G....CG..
19 A.AGCGTAA....AG..T.AATAG....C....CAC.AA.ATAAGC.TTAAA....A.CTG..G....CG..
20 A.AGCGTAA....AG..T.AATAG....A....CAC.AA.ATAA.C.TTAA.....A.CTG..G....CG..
21 A.AGCGTAA....AG..T.AATAG....A....CAC.AA.ATAA.C.T.AAA....A.CTG.CG....CG..
22 A.AGCGTAA....AG..T.AATAG....A....CAC.AA.ATAA.C.TTAAA....A.CTG.CG....CG..
23 A.AGCG.AA....AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.CTG.CG....CG..
24 A.AGCG.AA..G.AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.C.G..G....CG..
25 A.AGCG.AA..G.AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.C.G..G....CG..
26 A.AGCG.AA..G.AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.C.G..G....CG..
27 A.AGCG.AA..G.AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.C.G..G....CG..
28 A.AGCG.AA..G.AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.C.G..G....CG..
29 A.AGCG.AA..G.AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.C.G..G....CG..
30 A.AGCG.AA..G.AG..T.AATAG....A....CAC.AA.ATAAGC.TTAAA....A.C.G..G....CG..
31 AGAGCG.AAC..CAG....AATAG.........TCAC.AACATAAGCCT.AAA....A.C..C.G....TG..
31 AGAGCGTAAC..CAG....AATAG.....T..TCAC.GACATAAGCCT.AAA....A.C....G....TG..
33 AGAGCGTAAC..CAG....AATAG.....T..TCAC.AACATAAGCCT.AAA....A.C..C.G....TG..
34 A.AGCG.AAC..CAAT...AATAG..AT.....CAC.AACATAAGCCA.AAA....A.C..C.GG...TG..
35 AGAGCG.AAC..CAAT...AATAG...T.....CAC.AACATAAGCCA.AAA....A.C..C.GG...TG..
36 AG.GCG.AAC...AG....AATAG.........TCAC.AACATAAGCCT.AAAA...A.C..C.G....TG..
37 AG.GCG.AAC...AG....AATAG.........TCAC.AACATAA.CCT.AAA....A.C..C.G....TG..
38 AGAGCG.AAC...AG....AATA..........TCAC.AA.ATAAGCCT.AAAA...AAC..CCG....TG..
39 AGAGCG.AAC..AAG....AATAG.........TCAC.AA.ATAAGCCT.AAAA...AAC..CCG....TGG.
40 AGAGCG.AAC..AAG....AATAG.........TCAC.AA.ATAAGCCT.AAAA...AAC..CCG....TG..
41 AGTGCG..A.CGCAG...GAATAG...T....TTAC.AACAT.A.CCT.AAA.TA.AA..GC.G....CG..
42 AGT.CG..A.CGCAG.C.GAACAG...T.....CAC.AA.A.AA.CCT.AAA....AA..GC.G..TGCG..
43 AGT.CG..A.CGCAG.C.GAATAG...T....TCAC.AA.ATAA.CCT.AAA....AA..GCCG.TTGCG..
44 AGT.CG..A.CGCAG.C.GAATAG...T....TCAC.AA.ATAA.CC..AAA...GA.C.G..G....CG..
45 AGT.CG..A.CGCAG.C.GAATAG...T....TCAC.A..ATAA.CCT.AAA....AA..GCCG..TGCG..
46 AG..CG..A.CG.AG.C.GAATAG...T....TCAC.AA.AT.A.CCT.AAA....AA..GCCG..TGCG..
47 AGT.CG..A.CGCAG.C.GAATAG...T....TCAC.AA.AT.A.CCT.AAA....AA..GCCG..TGCG..
48 AGT.CG..A.CGCAG.C.GAATAG...T....TCAC.AA.AT.A.CCT.AAA....AA..GCCG..TGCG..
49 AGT.CG..A.CGCAG.C.GAATAG...T....TCAC.AA.AT.A.CCT.AAA....AA..GCCG..TGCG..
50 AGT.CG..A.CGCAG.C.GAATAG...T....TCAC.AA.AT.A.CCT.AAA....AA..GCCG..TGCG..
51 AGT.CG..A.CGCAG.C.GAATAG...T....TTAC.AA.AT.A.CCT.AAA.....A..GCCG..TGCG..
52 AGT.CG..A.CGCAG.C.GAATAG...T....TTAC.AA.AT.A.CCT.AAA.....A..GCCG..TGCG..
```

Alignment of DENV-2 parsimony informative sites (continued)

```
   444444444444444444444455555555555555555555555555555566666666666666666666
   334444555667778888899900011111222344445556667777899900000112223334444555
   281457036281470346905814701369235101362581270369212504679251470390268124

 1 CGAGCTTAGTCGTTGGGCGAGCATCCATATACGCGATAGCTACACACTTCCCAGGAAAAAACAGCTGTATTA
 2 TAGA...G..G.CC...TA.A..CTTG......T..CGA.C...TCT.C...G...C.T.CT..TC....C.
 3 TAGA...G.CG.CC...T...T.CTTGC.....T..CGATC...TCTC........C...TT..TC....C.
 4 TAGA...G.CG..C...T..AT.CTTG......T...GA.C...TCTC........C...TT..TC......
 5 ......ACACAA.C.AA.A.AAG....A.A.T.TTGC.....AG.C.C.GT.G.A.T.TGG.GA.CACGG..
 6 ......GCACAA.C.AA.AG.AG....A.A.T.TCGCG...GA..C.C.GT.G.A.T.TGGTGATCACGA..
 7 ......GCACAA.C.AA.A.AAG....A.A.TATCGCG...GA..C.C.GT.G.A.T.T.GTGA.CA.GA..
 8 .....CATACGA.C.AA.A.AAG....A.AGT..AGCG.T.GAG.CTC.GTT..A.C.TGG.GA.CACGG.G
 9 .....CATACAA.C.AA.A.AAG....A.AGT..TGCG...GAG.C.CCGTT..A.T.TGG.GA.CACGG.G
10 ......ACA.AA.C.AA.A.AAG....A.AGT..TGCG...GAG.C.C.GT...A.T.TGGAGA.CACGG.G
11 ......ACACAA.C.AA.A.AAG....A.AGT..TGCG...GAG.C.C.GT...A.T.TGG.GA.CACGG.G
12 ......ACACAA.C.AA.A.AAG....A.AGT..TGCG...GAG.C.C.GT...A.T.TGG.GA.CACGG.G
13 ......ACACAA.C.AA.A.AAG....A.AGT..TGCG...GAG.C.C.GT...A.T.TGG.GA.CACGG.G
14 ......ACACAA.C.AA.A.AAG....A.AGT..TGCG...GAG.C.C.GT...A.T.TGG.GA.CACGG.G
15 ......ACACAA.C.AA.A.AAG....A.AGT..TGCG...GAG.C.C.GT..AA.T.TGG.GA.CACGG.G
16 .A...C..TCAA.C.AAAA.AAG....A.A....TGCG...GAG...C.G..G.AGC.TGG.GA.CACGG..
17 .A...C..TCAA...AA.A.AAG....A.A....CGCG...GAG.T.C.G..G.AGC.TGG.GA.CACGG..
18 .A...C..TCAA...AA.A.AAG....A.A....AGCG...GAG.T.C.G..G.AGC.TGG.GA.CACGA..
19 .A...C..TCAA.C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG.GA.CACGG..
20 .A...C..TCAA.C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGG.CGG.GA.CACGG..
21 .A...C..TCAA.C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGG.TGG.GA.CACGG..
22 .A...C..TCAA.C.A..A.AAG.T..A.A....TGCG...GAG.C.C.G..G.AGG.TGG.GA.CACGA..
23 .A...C..TCAA.C.AA.A.AAG....A.A....TGCT...GAG.C.C.G..G.AGC.TGG.GA.CACGG..
24 .A...C..CCA..C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG.GA.CACGG..
25 .A...C..TCA..C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG.GA.CACGG..
26 .A...C..TCA..C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG.GA.CACGG..
27 .A...C.GTCA..C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG.GA.CACGG..
28 .A...C..TCA..C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG.GA.CACGG..
29 .A...C..TCA..C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG..A.CACGG..
30 .A...C..TCA..C.AA.A.AAG....A.A....TGCG...GAG.C.C.G..G.AGC.TGG..A.CACGG..
31 ....A.ACACAACC.AA.A.AAG....A.A....TGCG...TAG.C.C.GT..AA.C.TGG..A..ACGG..
31 ....A.ATACGACC.AA.A.AAG....A.A....TGCG...GAG.C.C.GT...A.C.TGG..A.CACGA..
33 ....A.ATACGACC.AA.A.AAG....A.A....TGCG...GAG.C.C.GT...A.C.TGG..A.CACGG..
34 ....A.ACACAACC.AATAGAAG....A.A..A.TGCGA..GAG.C.C.GT...A.CGTGG.GA..ACGG..
35 ....A.ACACAA.C.AATAGAAG....A.A..A.TGCGA..GAG.C.C.GT...A.CGTGG.GA..ACGG..
36 ....A.ACACAA.CAAA.AGAAG....A.A....TGCG.T.GAG.C.C.GT...A.C.TGGTGA..ACGG..
37 ....A.ACACAA.CAAA.AGAAG....A.A....TGCG...GAG.C.C.GT...A.C.TGGTGA..ACGG..
38 ....A.ATACAA.C.AA.AGAAG..T.A.A..T.TGCG...GAG.T.C.GT...A.C.TGG.GA..ACGG..
39 ....A.ATACAA.C.AA.AGAAG..T.A.A....TGCG....AG.C.C.GT...A.C.TGG.GA..ACGG..
40 ....A.ATACAA.C.AA.AGAAG..T.A.A....TGCG...GAG.T.C.GT...A.C.TGG.GA..ACGG..
41 ...A..ACACAA.C.A..A.AAGC...A.G....TG.G...GAG.C.C.GT...AGT.TGG..A.CACGG..
42 ...A..ACACAA...A....AAGCT..AGG....TG.G...GAG.C.C.GT...AGC.TGG..A.CAC.GCG
43 T..A..ACACAA..AA....AAGC...AGG....TG.G....AG.C.C.GT..AAGC.TGG..A.CAC.G.G
44 ...A..ACACAA..AA....AAGC...ACG....TG.G....AG.C.C.GT..ATGC.TGG..A.CAC.G.G
45 T.GA..ACACAA..AA....AAGC...AGG....TG.G....AG.C.C.GT...AGC.TGG..A.CAC.G.G
46 ...A..ACACAA..AA....AAGC...AGG....TGAG....AG.C.C.GT...AGC.TGG..A.CAC.G.G
47 T..A..ACACAA..AA....AAGC...AGG....TG.G....AG.C.C.GT...AGC.TGG..A.CAC.G.G
48 T..A..ACACAA..AA....AAGC...AGG....TG.G....AG.C.C.GT...AGC.TGG..A.CAC.G.G
49 T..A..ACACAA..AA....AAGC...AGG....TG.G....AG.C.C.GT...AGC.TGG..A.CAC.G.G
50 T..A..ACACAA..AA....AAGC...AGG....TG.G....AG.C.C.GT...AGC.TGG..A.CAC.G.G
51 T..A..ACACAA..AA....AAGC...AGG....TG.G....AG.C.C.GT...AGC.TGG..A.CAC.G.G
52 T..A..ACACAA..AA....AAGC...AGG.T..TG.G....AG.C.C.GTT..AGC.TGG..A.CAC.G.G
```

Alignment of DENV-2 parsimony informative sites (continued)

```
   6666666667777777777777777777777777777778888888888888888888888888888889
   6667788990000011122233344455555667778899001122223333444445556777778889 90
   1362547692578914736925804703679251473628140325891479036792584034692584 70

 1 CTTTCGTCAAGTGCGAAATCTAGAGCTCCCGGCACCTATACAAAAATTAACCAATTGGCGACGCGACAAATT
 2 ...A..C.G.....A......GAGA....T.TGG..C.C.A.T...CA..TTGG....TAG.......GGC.
 3 ...AT...G.....A...C.CGAGA....T.AGG..C.C.A.T.G..A..TTGGC....AG.......G.CC
 4 ...A..C.G.A...A...C..GAGA....T.AGG..C.CGA.TGG.CA..TTGG....TAG.......GGCC
 5 TGCG..AAG..CAT.CT....G...T.TTTAA..G.C..G..C...C..C.....CC.......AG....C.
 6 .GCGT.AAG.ACAT.CT....G...T.TTTAA..G...C...C...C..C..T..CC.......AG....C.
 7 ..CG..AAG.ACAT.CT....G...TCT.TAA..G.C.C...C...C..C.....CC.......AG....C.
 8 .GCG.AAAG..CATACT.C..G..AT.TTTAAT.GTC..G..CG.GC..G.....CC...G.A.AG....CC
 9 TGCG.AAAGG.CAT.CT....G..AT.TTTAA..GTC.CG..CG.GC..G.....CC...G.A.AG....CC
10 .GCG.AAAG..CAT.CT...CG..AT.TTT.A..G.C.CG..CG..CG.G.....CC...G.A.AG....C.
11 .GCG.AAAG..CAT.CT....G...T.TTT.A..G.C.CG..CG..C..G.....CC...G.A.AG....C.
12 .GCG.AAAG..CAT.CT....G...T.TTT.A..G.C.CG..CG..C..G.....CC...G.A.AG....C.
13 .GCG.AAAG..CAT.CT....G...T.TTT.A..G.C.CG..CG..C..G.....CC...G.A.AG....C.
14 .GCG.AAAG..CAT.CT....G...T.TTT.A..G.C.CG..CG..C..G.....CC...G.A.AG....C.
15 .GCG.AAAG..CAT.CT....G...T.TTT.A..G.C.CG..CG..C..G.....CC...G.A.AG....C.
16 .ACG.AAAG..CAT.CT...C....TGTTTAA..G.CGCG..CG..C........CC.....A.AG....C.
17 .ACG.AAAG..CAT.CT...CG...TCTTTAAT.G.CGCG..CG..C........CC.....A.AG....C.
18 .ACG.AAAG..CAT.CT...CGA..TCT.TAAT.G.CGCG..CG..C........CC.....A.AG....C.
19 .ACG.AAAG..CAT.CT...CG..ATCTTTAAT.G.CGCG..CG..C..G.....CC.....A.AG....C.
20 .ACG.AAAG..CAT.CT...CG...TCTTTAAT.G.CGCG..CG..C..G.....CC.....A.AG....C.
21 .ACG.AAAG..CAT.CT...CG...TCTTTAAT.G.CGCG..CG..C..G.....CC.....A.AG....C.
22 .ACG.AAAG..CAT.CT...CG...TCTTTAAT.G.CGCG..CG.....G...G.CC.....A.AG....C.
23 .ACG.AAAG..CAT.CT...CG...TCTTTA.T.G.CGCG..CG..C..G.....CC.....A.AG....C.
24 .ACG.AAAG..CAT.CT...CC...T.TTTA.T.G.CGCG..CG..C..G.....CC.....A.AG....C.
25 .ACG.AAAG..CAT.CT...CG...TCTTTA.T.G.CGCG..CG..C..G.....CC.....A.AG....C.
26 .ACG.AAAG..CAT.CT...CG...TCTTTA.T.G.CGC...CG..C..G.....CC.....A.AG....C.
27 .ACG.AAAG..CAT.CT...CG...TCTTTA.T.G.CGCG..CG..C..G.....CC.....A.AG....C.
28 .ACG.AAAG..CAT.CT...CG...TC.TTA.T.G.CGCG..CG..C..G.....CC.....A.AG....C.
29 .ACG.AAAG..CAT.CT...CG...T..TTA.T.G.CGCG..CG..C..G.....CC.....A.AG....C.
30 .ACG.AAAG..CAT.CT...CG...T..TTA.T.G.CGCG..CG..C..G.....CC.....A.AG....C.
31 .GCG.AAA.G.CAT.CTG.T.G...T.T.TAA..GT...G.GCG..C..G....CCC.....A.AG.G....
31 .GCG.AAAGG.CAT.CT..T.G...T.TTTAA..GT...G.GCG..C..G.....CC.....A.AG......
33 .GCG.AAAGG.CAT.CT..T.G...T.TT.AA..GT...G.GCG..C..G.....CC.....A.AG....C.
34 .GCA.AAAGG.CAT.CT........T.TTTAA..GT..CG.GCG..C..G.....CT.....A.AG......
35 .GCA.AAAGG.CAT.CT....G...T.TTTAA..GT..CG.GCG..C..G.....CC.....A.AG......
36 TGCG.AAAGG.CAT.CT........T.TTTAA..TT...G.GCG..CC.G.....CC.....A.AG....C.
37 TGCG.AAA.G.CAT.CT........T.TTTAA..GT...G.GCG..CC.G....CCC.....A.AG..G.C.
38 TGCG.AAAGG.CAT.CT....G...T.TTTAA..GT...G.GCG..CC.G.....CC....TA.AG....C.
39 TGCG.AAAGG.CAT.CC....G...T.TTTAA..GT...G.GCG...C.G.....CC.....A.AG....C.
40 TGCG.AAAGG.CAT.CC....G...T.TTTAA..GT...G.GCG..CC.G.....CC.....A.AG....C.
41 .GCATAAAG..CAT.CT.....A..T.T.TAA.GG.C.CGTGCG..CC.G...T.CT.....A.AGT.G.C.
42 .GCA.AAAG..CAT.CC....GA..T.T.TAA.GG...CGTGCG..C.GG.....CCA....ATAGTGG.CC
43 .GCA.AAAG..CA..CC....GA..T.T.TAA.GG...CGTGCG..CCGG.....CC.....ATA.T.G.CC
44 .GCA.AAAG..CA..CC....GA..T.T.TAT.GG...CGTGCG..CCGG.....CT...G.ATA.T.G.CC
45 .GCA.AAAG..CA..CC....GA..T.T.TAA.GG...CGTGCG..C.GG.....CC....TATA.T.G.CC
46 .GCG.AAAG..CAT.CC....GA..T.T..AA.GG...CG.GC...CCGG.....CCA....ATA.T.G.CC
47 .GCA.AAAG..CAT.CCG...GA..T.T.TAA..G...CGTGCG..CCGG.....CTA....ATA.T.G.CC
48 .GCA.AAAG..CAT.CCG...GA..T.T.TAA..G...CGTGCG..CCGG.....CTA....ATA.T.G.CC
49 .GCA.AAAG..CAT.CCG...GA..T.T.TAA..G...CGTGCG..CCGG.....CTA....ATA.T.G.CC
50 .GCA.AAAG..CAT.CC....GA..T.T.TAA..GTC.CGT.CG..CCGG.....CCA...TATA.T.G.CC
51 .GCA.AAAG..CAT.CC....GA..T.T.TAA..G.C.CG.GC...CCGG.....CCA....ATA.T.G.CC
52 .GCA.AAAG..CAT.CC....GA..T.T.TAA..G.C.CG.GC...CCGG.....CCA....ATA.T.G.CC
```

Alignment of DENV-2 parsimony informative sites (continued)

```
                                  1111111111111111111111111111111111111111111111111111
   9999999999999999999999990000000000000000000000000000000001111111111111111
   0111222233356667788899990001112223333444555566677778888900122333334455555
   6258124703973462547903692584790342468147016925814690369347925124576925 89

 1 CGGCAGTAAATACATGGAAGTAACGTCGCTGCTGACCCAGACTGCTAAACAGACCAAATTCTACGTAAAAGT
 2 .A.....C.GG...C...TAC....CT.........TT...TC..C..G.....T.....TC.TACGG...C
 3 .A.....C.GG...C...TAC.....T.........TT.......C...T.......G..TC.TACGG...C
 4 .A.....C.GGG..C....AC.G..CT.........T.........C..GTG............T.CGG...C
 5 .A.T...GG.A.T.C.AGC.CT...CT.A..T.AG.T..TC..A.C.TG..A.T.G..CA.C..AC.GG.AC
 6 .A.T...GG.AGT.C.AGC.CT.TACT.A..T.AGTT..CC.CA.C.TG..A.T.G..CA.C..AC.GG.AC
 7 TA.T...GG.ACT.C.AGC.CT..ACT.A....AG.T..CC.CA.C.TG..AG..G..CA.C..AC.GG.AC
 8 .A.T..CGG.A.TGCAAGC..T.T....A..T.AG.T.GTC.CA.C.TG..AGT.G.CCA.CG.ACGGGGA.
 9 .A.T...GG.A.TGCAAGC..T.T....A..T.AG.T.GTC.CA.C.TG..AGT.G..CA.CG.AC.GGGA.
10 .A.T..CGG.A.T.CAAGC.CT.T.CT.A..T.AG.TT.TC..A.C.C..GA.T.G..CA.C..AC.GG.A.
11 .A.T...GG.A.T.CAAGC..T.T.CT.A..T.AG.TT.TC..A.C.C...A.T.G..CA.C..AC.GG.A.
12 .A.T...GG.A.T.CAAGC..T.T.CT.A..T.AG.TT.TC..A.C.C...A.T.G..CA.C..AC.GG.A.
13 .A.T...GG.A.T.CAAGC..T.T.CT.A..T.AG.TT.TC..A.C.C...A.T.G..CA.C..AC.GG.A.
14 .A.T...GG.A.T.CAAGC..T.T.CT.A..T.AG.TT.TC..A.C.C...A.T.G..CA.C..AC.GG.A.
15 .A.T...GG.A.T.CAAGC..T.T.CT.A..T.AG.TT.TC..A.C.C...A.T.G..CA.C..AC.GG.A.
16 TA.T.A.GG.A.T.CAAGC.CT.T.CT....T.AG.T..TC..ATCGC...A.T.G..CA.C..AC.GG.A.
17 TA.T.A.GG.A.T.CAA.C.CT.T.CTAA..T.AG.T..TC..ATCGC...A.T.G..CA.C..AC.GG.A.
18 TA.T.A.GG...T.CAA.C.CT.T.CTAA..T.AG.T..TC..ATCGC...A.T.G..CA.C..AC.GG.A.
19 TA.T.A.GG.A.T.CAAGC.CT.T.CT.A..T.AG....TC..ATCGC...A.T.G..CA.C..AC.GG.A.
20 TA.T.A.GG.A.T.CAAGC.CT.T.CT.A..T.AG....TC..ATC.C...A...G..CA.C..AC.GG.A.
21 TA.T.A.GG.A.T.CAAGC.CT.T.CT.A..T.AG....TC..ATC.C...A.T.G..CA.C..AC.GG.A.
22 .AAT.A.GG.A.T.CAAGC.CT.T.CT.A..T.AG....TC..ATC.C...A...G..CA.C..AC.G..A.
23 TA.T.A.GG.A.T.CAAGC..T.T.CT.ACAT.AG....TC..A.C.C...A.T.G..CA.C..AC.GG.A.
24 TA.T.ACGG.A.T.CAAGC..T.T.CT.ACAT.AG....AC..ATCGC...A.T.G..CA.C..AC.GG.AC
25 TA.T.A.GG.A.T.CAAGC..T.T.CT.ACAT.AG....TCT.ATC.C...A.T.G..CA.C..AC.GG.A.
26 TA.T.A.GG.A.T.CAAGC..T.T.CT.ACTT.AG....TCT.ATC.C...A.T.G..CA.C..AC.GG.A.
27 TA.T.A.GG.A.T.CAAGC..T.T.CT.ACAT.AG....TCT.ATC.C...A.T.G..CA.C..AC.GG.A.
28 TA.T.A.GG.A.T.CAAGC..T.T.CT.ACAT.AG....TCT.ATC.C...A.T.G..CA.C..AC.GG.A.
29 .A.T.A.GG.A.T.CAAGC..T.T.CT.ACAT.AG....TCT.ATC.C...A.T.G..CA.C..AC.GG.A.
30 TA.T.A.GG.A.T.CAAGC..T.T.CT.ACAT.AG....TCT.ATC.C...A.T.G..CA.C..AC.GG.A.
31 TA.T...GG.A.TGCAAGC..T..ACT.A..T.AG.T..TCT.A.C.TC..A...G..CA.C..A...G.AC
31 TA.T...GG.A.T.CAAGC..T..ACT.A..T.AG.T..TC..A.C.TT..A...G..CA.C..A...G.AC
33 TA.T...GG.A.T.CAAGC..T...CT.A..T.AG.T..TC..A.C..T..A...G..CA.C..A...G.AC
34 TA.T...GG.A.T.CAAGC..TG..CTAA..T.AG.T..TCT.A.C.TT..A...G..CA.C..A...G.AC
35 TA.T...GG.A.T.CAAGC..TG..CTAA..T.AG.T..TCT.A.C.TT..A...G..CA.C..A...G.AC
36 TA.TG..GG.A.T.C.AGT..T.TACTAA..T.AG.T..CC..A.C.TT..A...GG.CA.C..A...G.AC
37 TA.TG..GG.A.T.C.AGT..T.TACT.A..T.AG.T..CC..A.C.TT..A...GG.CA.C..A...G.AC
38 TA.T....G.A.T.CAAGT..T.TACT.A..T.AGTT..TC..A.C.TC..A...G.GCA.C..A...G..C
39 TA.T....G.A.T.CAAGT.CT.TACT.A..T.AG.T..TC..A.C.CC..A..TG..CA.C..A...G..C
40 TA.T....G.A.T.CAAGT..T.TACT.A..T.AG.T..TC..A.C.TT..A...G..CA.C..A...G.AC
41 T..T..CGG.A....AA.C.CT...CT.A..TCAG.T..TT..A.C.T...AG.GG..CA.C.TACGGG.AC
42 .A.T..CGG.A....AA.C.CT...C..A..TCAG.TTGCC.CA.C.T...AG.TG..CA.C..ACG...A.
43 .A.T...GG.A....AA.C.CT...C..A..TCAG.TTGCC.CA.C.T...AG.TG..CA.C..ACG...A.
44 .A.T...GG.A....AA.C.CT...CT.A..TCAG.TTGCC.CA.C.T...AG.TC..CA.C..ACG...A.
45 .A.T.A.GG.A.T..AA.C.CT...C..A..TCAG.TTGCC.CT.C.T...AG.TG..CA.C..ACG...A.
46 .AATGA.GG.A....AA.C.CT...C..A..TCAG.TTGCC.CA...T...AG.TG..CA.C..AC....A.
47 .A.T.A.GG.A....AA.C.CT...C..A..TCAG.TTGCC.CA.C.T...AG.TG..CA.C..ACG...A.
48 .A.T.A.GG.A....AA.C.CT...C..A..TCAG.TTGCC.CA.C.TG..AG.TG..CA.C..ACG...A.
49 .A.T.A.GG.A....AA.C.CT...C..A..TCAG.TTGCC.CA.C.TG..AG.TG..CA.C..ACG...A.
50 .A.T.A.GG.A....AA.C.CTT..C..A..TCAG.TTGCC.CA.C.T...AG.TG...A.C..ACG...A.
51 .A.T.A.GG.A....AA.C.CT...C..A..TCAG.TTGCC.CA.C.T...AG.TG..CA.C..ACG...A.
52 .A.T.A.GG.A....AA.C.CT...C..A..TCAG.TTGCC.CA.C.T...AG.TG..CA.C..ACG...A.
```

Alignment of DENV-2 parsimony informative sites (continued)

```
   1111111111111111111111111111111111111111111111111111111111111111111111111111
   1111111111111112222222222222222222222222333333333333333333333333333334444444
   6666667777889990022233444456677778899990001112222333444556678888801112222
   1457890689281476914709235873623584703492581470369258147362840346970392 38

 1 GACCAATCAAACCAATAAACACCTGATTTTCAGGCGTGGGGGTGATTACTTTTTCCGCGCGGATATTCCCCA
 2 ..TA..CT....T......T.......C..A....AA..AAA.AT.C...CCC..TAT....G...C.T.T.
 3 ..TA..C.....T......T.......C.CA....AA..AAA.ATCCG..CCCC.TAT....G..CC...T.
 4 ..TA..C.....TT.....T.T.....C.CA.....A..AAA.AT.CG..CCC..TAT....G..CC.T.T.
 5 .G....CT.G.TTCC.GGGGG.T.ATC..CTGA.TA.AA.TCCAC.A...A.CCTTATA.AC..C.A..A.T
 6 .G....C..G.TTCC.GGGGG.T..TC.CC.GA.TA.AAATCCAC.A...A.CCTTATA.AC..CCA..A.T
 7 .G....CT.G.TTCC.G.GGG.T..TC..C.GA.TA.AAATCCACCA...A.CCTTATA.AC..C.A..A.T
 8 .G.G...T.G.TTCC.GG.GGTTCAT...C.GAA.A.AA.TCC.T.A..CG.CC...TA.AC..C.AT.A.T
 9 .G.T...T.G.TTCC.GG.GGTTCAT...C.GA..A.AA.TCC.T.A..CG..CT..TA.AC..T.AT.A.T
10 .G....CT.G.TTCC.GG.GG.T.AT...C.GA.TA.AA.TCCATCA...G.CCT.ATA.AC..C.A..A.T
11 .G....CT.G.TTCC.GG.GG.T.AT...C.GA.TA.AA.TCCATCA...G.CCT.ATA.AC..C.A..A.T
12 .G....CT.G.TTCC.GG.GG.T.AT...C.GA.TA.AA.TCCATCA...G.CCT.ATA.AC..C.A..A.T
13 .G....CT.G.TTCC.GG.GG.T.AT...C.GA.TA.AA.TCCATCA...G.CCT.ATA.AC..C.A..A.T
14 .G....CT.G.TTCCAGG.GG.T.AT...C.GA.TA.AA.TCCATCA...G.CCT.ATA.AC..C.A..A.T
15 .G....CT.G.TTCCAGG.GG.T.AT...C.GA.TA.AA.TCCATCA...G.CCT.ATA.AC..C.A..A.T
16 A.....CT.G.T.CC.G..GG.T.ATC..C.GA.TA.AA.TCCATCA...G.C.T.ATAAAC.CC.AT.A.T
17 ......CT.G.T.CC.G..GG.T.ATC..C.GA.TA.AA.TCCATCA...G.C.T.ATA.AC.CC.AT.A.T
18 ......CT.G.T.CC.G..GG.T.ATC..C.GA.TA.AA.TC.ATCA...G.C.T.ATA.AC.CC.A..A.T
19 ......CT.G.T.CC.G..GG.T.ATC..C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
20 .......T.GGT.CC.G..GG.T.ATC..C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
21 .......T.GGT.CC.G..GG.A.ATC..C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
22 ......CT.GGT.CC.G..A..T.ATC..C.GA.TA.AA.TCCATCA.......T.ATA.AC.CC.ATTA.T
23 .......T.GGT.CCCG..GG.T.ATC..C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
24 .......T.GGT.CC.G..AG.T.ATC..C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
25 .......T.GGT.CC.G..AG.T.ATCC.C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
26 .......T.GGT.CC.G..AG.T.ATCC.C.GA.TACAA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
27 .......T.GGT.CC.G..AG.T.ATCC.C.GA.TACAA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
28 .......T.GGT.CC.G..AG.T.ATCC.C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
29 .......T.GGT.CC.G..AG.T.ATCC.C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
30 .......T.GGT.CC.G..AG.T.ATCC.C.GA.TA.AA.TCCATCA...G...T.ATA.AC.CC.AT.A.T
31 AG....C..G.TTCC.G..GG.T.ATC.CC.GA.TA.AA.TCCAT.A.T.G.CCT.ATA.ATGCT.A..ATT
31 .G....CT.G.TTCC.G..GG.T.ATC..C.GAA.A.AA.TCCAT.A.T.G.CCT.ATA.ATGCT.A..A.T
33 .G....CT.G.TTCC.G..GG.T.ATC..C.GA..A.AA.TCCAT.A.T.G.CCT.ATA.ATGCT.A..A.T
34 .G...GCT.G.TTTC.G..GG.T.AT...C.GA..A.AA.CCCAC.A...G.CCTTATA.ACGCC.A..A.T
35 .G...GCT.G.TTTC.G..GG.T.AT...C.GA..A.AA.CCCAC.A...G.CCTTATA.ACGCC.A..A.T
36 .G...GCT.G..ACC.G..G..T..TC..C.GA..A.AA.TCCAT.A...G.CCT.A.A.ATGCC.A..A.T
37 .G...GCT.G..ACC.G..G..T..TC..C.GA..A.AA.TCCAT.AT..G.CCT.ATA.ATGCC.A..A.T
38 .G...GCT.G..ACCCG..GG.T.ATC..C.GA..ACAA.CCCAT.A...G.CCT.ATA.ATGCT.ATTA.T
39 .G...GCT.G..A.CCG..GG.T.ATC..C.GA..A.AA.CCCAT.A...G.CCT.ATA.ATGCC.AT.A.T
40 .G...GCT.G..ACCCG..G..T.ATC..C.GA..A.AA.CCCAT.A...G.CCT.ATA.ACGCC.AT.A.T
41 .G.TG.C..G.TTCT.G..G..T.ATC....GA.TAAAA.TCC.C.AG..G.CCT.AT..AT.CCCATTA.C
42 .G.TG.C..G.T.CC.GG.G..T..TC....GA..AAAA.TCC.T.A...G..CT.AT..AT.CCCATAA.T
43 .G.GG.C..G.T.CC.GG.G..T..CC....GA..AAAA.TCC.T.A..CG..CT.AT..AT.CCCATAA.T
44 .G.GG.C..G.T.CC.G.GTG.T..CC....GA..A.AA.TCC.T.A...GGGCT.AT..AC.C..AT.A.T
45 .G.GG.C..G.T.CC.GG.G..T..CC....GA..AAAA.TCC.T.A..CG..CT.AT..AT.CCCATAA.T
46 AG.GG.C..G.T.CT.GG.G.....CC....GA.TAAAAATCC.TCA..CG.ACTTAT..AT.CCCA.AA.T
47 .G.GG.C.GG.T.CC.GG.G.....CC....GA..AAAAATCC.T.A..CG..CT.AT.TAT.CCCA.AA.T
48 .G.GG.C.GG.T.CC.GG.G.....CC....GA..AAAAATCC.T.A..CG..CT.AT.TAT.CCCA.AA.T
49 .G.GG.C.GG.T.CC.GG.G.....CC....GA..AAAAATCC.T.A..CG..CT.AT.TAT.CCCA.AA.T
50 .G.GG.C..G.T.TC.GG.G.....CCC...GA..AAAAATCC.T.A..CG..CT.A..TAT.CTCATAA.T
51 .G.AG.C..G.T.CC.GG.G.....CC....GA..AAAAATCC.T.A..CG..CT.AA.TAA..CCATAA.T
52 .G.GG.C..G.T.CC.GG.G.....TC....GA..AAAAATCC.T.A..CG..CT.AA.TAA..TCATAATT
```

Alignment of DENV-2 parsimony informative sites (continued)

```
   1111111111111111111111111
   4444444444444444444444444
   3333344445555556667777888
   1245701369023589145234925

 1 GAACTGATAAATGAACCCTTGAGGG
 2 ..C.G.TA...C.G..A.C....A.
 3 ..C.G.CA...C.G..A.C....A.
 4 ..C.G.TA...C.G..A.C....A.
 5 ....AATGG....G..GT..C...C
 6 .T..AATGG....G..GT..C.A.C
 7 .T..AATGG....G..GT..C.T.C
 8 .T..AATGG.GC.G..G...CG..C
 9 AT..AATGG.GC.G..G.C.CG..C
10 .T..AATGG.GC..G.GT..T...C
11 .T..AATGG.GC.GG.GT..T...C
12 .T..AATGG.GC.GG.GT..T...C
13 .T..AATGG.GC.GG.GT..T...C
14 .T..AATGG.GC.GG.GT..T...C
15 .T..AATGG.GC.GG.GT..T...C
16 .T..AATGG.GC.G..G..CT...A
17 .T..AATGG.GCA...G..CC...T
18 .T..AATGG.GCA...G..CC...T
19 .T..AATGG.GC.G..G.CCT...T
20 .T..AATGG.GC.G..G.CCT...T
21 .T..AATGG.GC.G..G.CCT...T
22 .T..AATGG.GC.G..G.C.T...A
23 .T..AATGG.GC.G..G.CCT...T
24 .TG.AATGG.GC.G..G.CCT...T
25 .TG.AATGGCGC.G..G.CCT..AT
26 .TG.AATGGCGC.G..G.CCT..AT
27 .TG.AATGGCGC.G..G.CCT..AT
28 .TG.AATGGCGC.G..G.CCT..AT
29 .TG.AATGGCGC.G..G.CCT..AT
30 .TG.AATGGCGC.G..G.CCT..AT
31 .T..AATAG.GC.G.TGT..C...C
31 .T..AATAG.GC.G.TGTC.T...C
33 .T..AATAG.GC.G.TGTC.T...C
34 .T..AATAGGGC.G.TAT..T...C
35 .T..AATAGGGC.G.TGT..C...C
36 .T.TAATAG.GC.G.TGT..T...C
37 .T.TAATAG.GC.G.TGT..T...C
38 .T.TAATAG.GC.G.TGT..C...T
39 .T.TAATAG.GC.G.TGT..C...T
40 .T.TAATAG.GC.G.TGT..C...T
41 .T..AATAG..C.G..A...C...C
42 .T..AATAG..C.G..A...T...C
43 .T..GATAG..C.G..G...T...C
44 .T..AATGG.GC.G..G.CCT...T
45 AT..GACAG..C.G..G...T...C
46 .T..GATAG..C.G..G...T..AC
47 .T..GATAG..C.G..G...T..AC
48 .T..GATAG..C.G..G...T.TAC
49 .T..GATAG..C.G..G...T.TAC
50 .T..GATAG..C.G..G...T..AT
51 .T..GATAG..C.G..G...T..AC
52 .T..GATAG..C....G...T..AC
```

Alignment DENV-3 parsimony informative sites

```
                                    1111111111111111112222222222222222222
                                  1444677889991122344557778888990011222334444456789
                                 9825865814069471654937145058925230925817123627321

 1 PUERTO RICO/63/BS             CGCAGACGTTCTCACTCTTTCTGGCCGCAGTACCGTCTAGACTAATGCG
 2 PUERTO RICO/77/1339           ....................C.........................
 3 PHILIPPINES/56/H87            TA...C..C.T..G.C...CTC....AG.A..T.A..C.A.T..GC.T.
 4 CHINA/80/2                    TA...C..C.T..G.C..CCTC....AG.A..T.A..C.A.T..GC.T.
 5 PHILIPPINES/83/168            TA..AC..C.T..G.CTCCC.C.A..AG.A.GT....C.A.TC.GC..A
 6 MALAYSIA/81/29586             TA..AC..C.T..G.C.CCC.C....AG.A.GT..C.C.A.T..GC...
 7 INDONESIA/73/22876            TA..AC..C.T..G.CTCCC.C.A.TAG.A.GT..C.C.A.T..GC...
 8 FIJI/92/29472                 TA..AC..C.T.TG..TCCC.C.A.TAG.A.GT..C.C.A.T..GC...
 9 TAHITI/89/2167                TA..AC..C.T.TG..TCCC.C.A.TAG.A.GT..C.C.A.T..GC...
10 INDONESIA/78/1280             TA..ACT.C.T..G.CTCCC.CAA..AG.A.GT..C.C.A.T..GC...
11 INDONESIA/85/159              TA..AC.AC.T..G.CTCCC.C.A..AG.A.GT..C.C.A.T..GC...
12 MALAYSIA/74/1300              TA..AC....T..G.CTCCC.C.A..AG.A.GT..C.C.A.T..GC...
13 THAILAND/73/CH3489D73         .AT..C..C.TC.G.C..CC.C..T.AG.A..TT..TC.A.T..GCAT.
14 MALAYSIA/93/LN1746            TATG.T..C.TC.GTC..CC.C..T.AG.A..TT..TCGA.T..GCA..
15 THAILAND/87/MK315             .ATG.TT.C.TC.GTC..CC.C..T.AG.A..TT..TCGA.T..GCA..
16 MYANMAR/98/31985KLA           .ATG.T.AC.TC.G.C..CC.C..T.AG.AC.TT..TCGA.T..GCA..
17 THAILAND/93/044               .ATGAT..C.TC.GTC..CC.C..T.AG.A..TT..TCGA.T..GCA..
18 THAILAND/97/0106              ..TG.C..C....GTC..CC.C..T.AG.A..TT..TCGA.T..GCA..
19 SAMOA/86/1696                 TAT.AC..C.T..G.CT.CC.C.A..AGGA..TTA.TC.AGT.GGC..A
20 SRI LANKA/81/1326             .AT.AC..C.T..G.CT.CC.C....AGGA..TTA.TC.AGT.GGC.TA
21 SRI LANKA/85/1594             TAT.ACT.C.T..G.CT.CC.C....AGGA..TTACTC.AGTCGGC.TA
22 INDIA/84/1416                 TAT.AC..CGT..G.CT.CC.C....AGGA..TTA.TC.AGT.GGC.TA
23 SRI LANKA/00/1266             TAG.AC..CGT..G.CT.CC.C....AGGA..TTA.TC.AGT.GGC.TA
24 SRI LANKA/91/2783             TAT.AC..CGT..G.CT.CC.C....AGGA..TTA.TC.AGT.GGC.TA
25 SOMALIA/93/S142               TAT.AC..CGT..G.CT.CC.C....AGGA..TTA..C.AGT.GGC.TA
26 MARTINIQUE/99/1243            TAT.AC..CGT..GA.T.CC......AGGA..TTA.TC.AGT.GGC.TA
27 BRAZIL/00/68784               TAT.AC..CGT..GA.T.CC......AGGA..TTA.TC.AGT.GGC.TA
28 MARTINIQUE/01/2336            TAT.AC..CGT..GA.T.CC......AGGA..TTA.TC.AGT.GGC.TA
29 MARTINIQUE/00/1706            TAT.AC..CGT..GA.T.CC..A...AGGA..TTA.TC.AGT.GGC.TA
30 PANAMA/94                     TAT.AC..CGT..G..T.CC.C....AGGA..TTA.TC.AGT.GGC.TA
31 MX/YUCATAN/95/BC4             TAT.AC..CGT..G..T.CC.C....AGGA..TTA.TC.AGT.GGC.TA
32 VENEZUELA/00/5990             TAT.AC..CGTC.G..T.CC.C....AGGAC.TTA.TC.AGT.GGC.TA
33 VENEZUELA/01/7984             TAT.AC..CGTC.G..T.CC.C....AGGAC.TTA.TC.AGT.GGC.TA
34 VENEZUELA/01/6668             TAT.AC..CGTC.G..T.CC.C....AGGAC.TTA.TC.AGT.GGC.TA
35 MX/95/6097                    TAT.TC..CGTC.G..T.CC.C....AGGA..TTA.TC.AGT.GGC..A
36 MX/OAXACA/00                  TAT.TC..CGTC.G..T.CC.C....AGGA..TTA.TC.AGT.GGC..A
37 MX/QUINTANA ROO/97/BC177      TAT.TC..CGTC.G..T.CC.C....AGGA..TTA.TC.AGT.GGC..A
38 MX/YUCATAN/96/BC285           TAT.TC..CGTC.G..T.CC.C....AGGA..T.A.TC.AGT.GGC...
39 MX/YUCATAN/97/BC192           TAT.TC..CGTC.G..T.CC.C....AGGA..TTA.TC.AGT.GGC..A
40 MX/QUINTANA ROO/97/BC197      TAT.TC..CGTC.G..T.CC.C....AGGA..TTA.TC.AGT.GGC..A
```

Alignment of DENV-3 parsimony informative sites (continued)

```
   2233333333333333333333344444444444444555555555555555555555556666666666666
   9900011233444566667789990111224677889900111112234446667888890122344446777
   4706928707258703460163498457391249030256235695870194570568919814302563023

 1 CGTATTCCGTGATCCTTACTCACAACATCTCACCTACCGCTTCCAACTATCTCGCCCACAAGCACCATCAGAG
 2 ..................T.........T......A.................................
 3 .AC.C.T.....CTA...TCT..G..GC.G..T..GT...CCTT..TCGC..T.TTT.TG.AT..T.C..A.A
 4 .AC.C.T.....CTA...TCT..G..GC.G..T..GT...CCTT..TCGC..T.TTT.TG.AT..T.C..A.A
 5 .AC.C.T.....CTA...TCTG.G..GC.G.....G...TCCT...T.GC..TAT--GTG..T.......A..
 6 .AC.C.T...A.CTA..GTCTG.G..G..G.........TCCT...T.GC.CTATTTGTG.A..T.....A..
 7 .AC.C.T...A.CTA..GTCTG.G..GC.GT....G...TCCT...T.GC..TATTTGTG.AT.T.....A..
 8 .AC.C.T..CAGCTA..GTC.G.G..GC.GT....G...TCCT.G.T.GCT..ATTTGTG.AT.T.G.T.A..
 9 .AC.C.T..CA.CTA..GTC.G.G..GC.GT....G...TCCT.G.T.GCT..ATTTGTG.AT...G.T.A..
10 .AC.C.T...A.CTG..GTCTG.G..GC.G.....G...TCCT...T.GC..TAT.TGTG.AT.T....GA..
11 .AC.C.T...A..TA..GTCTG.G..GC.G.....G...TCCT...T.GC..T.TTTGTG.AT.T....GA..
12 .AC.C.T...A.CTA..GTCTG.G..GC.G.....G...TCCT...T.GC..TATTTGTG.AT.T....GA..
13 TAC.C.TA....CTAC..TCT..G.TGCTG.....G...TC.T.G.T.GC..T.T.T.TG.AT....C..A..
14 TAC.C.TA....CTA.C..CT..G.TGC.G.C.T.G...TCCT.G.T.GC..T.TTT.TG.ATG...C..A..
15 TA..C.TAA...CTA.C..CT..G.TGCTG.C.T.G...TC.T.G.T.GC..T.TTT.TG.ATG.T.C..A..
16 TA..C.TAA...CTA.C..CT..G.TGCTG.C.T.G...TC.T.G.T.GC..T.TTT.TG.ATG...C..A..
17 TA..C.TAA...CTA.C..CT.TG.TGC.G.C.T.G...TC.T.GGT.GC.CT.TTT.TGCATG...C..A..
18 TA..C.TAAC..CTA.C..CT..G.TGC.G.C.T.G...TCCT.GGT.G...T.TTT.TGCATG...C..A..
19 .ACGC.T..CA..TACCGTCT..G..GCTG.....G.TA.CCT.G.T.GC..T.T.T....AT.......A..
20 .ACGC.T...A..TACCGTCT.TG..GCTG.....G.TA.CCT.G.T.GC..T.T.T....AT.......A..
21 .ACGC.T...A..TACCGTCT.TG..GCTG.....G.TA.CCT.G.T.GC..T.T.T....AT.......A..
22 .ACGC.T...A..TACCG.CT.TG..GCTG.....G.TA.CCT.G.T.GC..T.T.T....AT.......A..
23 .A.GC.T.A.A..TACCG.CT.TG..GCTG.....G.TA.CCT.G.T.GC..T.T.T....AT.......A..
24 .ACGC.....AG.TACCG.CT.TG..GCTG.....G.TA.CCT.G.T.GC..T.T.T....AT.......A..
25 .ACGC.T...A..TACCG.CT.TG..GCTG.....G.TA.CCT.G.T.GC..T.T.TG...AT.T.....A..
26 .ACGCAT...A..TACCG.CT.TG..GC.GT....G.TA.CCT.G.T.GC..T.T.TG...AT.T.....AG.
27 .ACGCAT...A..TACCG.CTGTG..GC.GT....G.TA.CCT.G.T.GC..T.T.TG...AT.T...T.AC.
28 .ACGCAT...A..TACCG.CT.TG..GC.GT....G.TA.CCT.G.T.GC..T.T.TG...AT.T.....AG.
29 .ACGCAT...A..TACCG.CT.TG..GC.GT....G.TA.CCT.G.T.GC..T.T.TG...AT.T.....AG.
30 .ACGCAT...A..TACCG.CT.TG..GC.G.....G.TA.CCT.G.T.GC.CT.T.TG...AT.T.....AG.
31 .ACGCAT...A..TACCG.CT.TG..GC.G....CG.TA.CCT.G.T.GC.CT.T.TG...AT.T.....AG.
32 .ACGCAT...A..TACCG.CT.TG..GC.G.....G.TA.CCT.G.T.GC..T...TG...AT.T.....AG.
33 .ACGCAT...A..TACCG.CT.TGG.GC.GT....G.TA.CCT.G.T.GC..T...TG...AT.T.....AG.
34 .ACGCAT...A..TACCG.CT.TGG.GC.G.....G.TA.CCT.G.T.GC..T...TG...AT.T.....AGA
35 .ACGCAT...A..TACCG.CT.TG..GC.G.....G.TA.CCT.G.T.GC..T.T.TG...AT.T.....AG.
36 .ACGCAT...A..TACCG.CT.TG..GC.G....CG.TA.CCT.G.T.GC.CT.T.TG...AT.T.....AG.
37 .ACGCAT...A..TACCG.CT.TG..GC.G.....G.TA.CCT.G.T.GC..T.T.TG...AT.T.....AG.
38 .ACGCAT...A..TACCG.CT.TG..GC.G.....G.TA.CCT.G.T.GC..T.T.TG...AT.T.....AG.
39 .ACGCAT...A..TACCG.CT.TG..GC.G.....G..A.CCT.G.T.GC..T.T.TG...AT.T.....AG.
40 .ACGCAT...A..TACCG.CT.TG..GC.G.....G..A.CCT.G.T.GC..T.T.TG...AT.T.....AG.
```

Alignment of DENV-3 parsimony informative sites (continued)

```
                                                                111111111111
   666666667777777777777778888888888888889999999999999999999999000000000000
   777899990122334455788890001233444455670000000112333455566689013334555667
   567402368409254709703451499517013958801234567261039217803660812581069351

 1 AGTTCGAAGCTAGAGTTGCAACCGTTCTCATCATTACGTCAGGCACCATTGGGTATATCTAGTCCTTTCGAC
 2 ............................T....C...............................C....G.
 3 .AC...G..T..AGA.CAT...T.CC..TGCT..CGT..TGAAT.TTGCC..AC..G..CC...TCCCT.G.
 4 .AC...G..T..AGA.CAT...T.CC..TGCT..CGT..TGAAT.TTGCC..AC..G..CC...TCCCT.G.
 5 .ACCTAG...C.AGAA........CC..TGCT..CGT..TGAATGTTGCC..AC..G...C.C.TCC...G.
 6 .ACCTAG...C.AGAAC.....T.CC..TGCTG.CGT..TGAATG.TGCC.AAC..G...C.C.TCC...GT
 7 .ACCTAG...C.AGAAC.....T.CC..TGCT..CGT..TGAATG.TGCC.AAC..G...C.C.TCC...G.
 8 .ACCTAG...C.A.AAC.....T.CC..TGCT..CGT..TGAATG.TGC..A.C..G...C.C.TCC..A..
 9 .ACCTAG...C.AGAAC.....T.CC..TGCT..CGT..TGAATG.TGC..A.C..G...C.C.TCC..A..
10 .ACCTAG...C.A.AAC.....T.CC.CTGCT.CCGT..TGAATG.TGCC.AAC..G...C...TCC...G.
11 .ACCTAG...C.A.AAC....TT.CC..TGCT.CCGT..T.AATGTTGCCAAA...G...C.C.TCC...G.
12 .ACCTAG...CGA.AAC.....T.CC..TGCT.CCG...TGAATGTTGCC.AAC..G...C.C.TCC...G.
13 GAC...G.....AGA.C...G.T.CA.CTGCTG..GT..TGAA..TTGCC..AC..G...C...T.CC..G.
14 GAC...G.....AGA.C...G.T..ATCTGCTG..GT..TGAAT.TTGCCA.AC..G...CA..T.CC..G.
15 GAC...G.....AGA.C...G.T..ATCTGCTG..GT..TGAAT.TTGCC..AC..G...CA..T.CC..G.
16 GAC...G.....AGA.C...G.T..ATCTGCTG...T..TGAAT.TTGCC..AC..G...CA..T.CC..G.
17 GAC...G.A...AGA.C...G.T..ATCTGCTG..GT..TGAAT.TTGCC..AC..G...CAC.T.CC..G.
18 GAC...G.A...AGA.C...G.T..ATCTGCTG..GT..TGAAT.TTGCC..AC.CG...CAC.T.CC..G.
19 .ACC..GG..CGA.A...T....ACA.CTGCT...GT.A.GAAT.TTGCA..ACG.GC.CCA..TCCC..GT
20 .ACC..GG..C.A.A.C.TT...ACC.CTGCT...GT.A.GAAT.TTGCA..ACG.GC..CA..TCCC..GT
21 .ACC..GG..C.A.A.C.T....ACC.CTGCT...GT.A.GAAT.TTGCA..AC..GC..CA..TCCC..GT
22 .ACC..GG..CGA.A.C.TC...ACA.CTGCT...GT.A.GAAT.TTGCA...C..GC..CA..TCCC..G.
23 .ACC..GG..CGA.A.C.TC..TACA.C.GCT...GT.A.GAAT.TTGCA..AC..G...CA.TTCCC..GT
24 .ACC..GG..C.A.A.C.T....ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GC..CA..TCCC..GT
25 .ACC..GG..CGA.A.C.TC...ACA.CTGCT..CGT.A.GAAT.TTGCA..AC..GC..CA.TTCCC..GT
26 .ACC..GG.TCGA.A.C.TC.T.ACA..TGCT...GT.A.GAAT.TTGCA..AC..GC..CA..TC.C..GT
27 .ACC..GG.TCG..A.CTTC.T.ACA..TGCT...GT.A.GAAT.TTGCA..ACG.GC..CA..TC.C..GT
28 .ACC..GG.TCG..A.C.TC.T.ACA..TGCT...GT.A.GAAT.TTGCA..AC..GC..TA..TC.C..GT
29 .ACC..GG.TCG....C.TC.T.ACA..TGCT...GT.A.GAAT.TTGCA..AC..GC..CA..TC.C..GT
30 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
31 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
32 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GTAA.GAAT.TTGCA..AC.CGCT.CA..TC.C..GT
33 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GTAA.GAAT.TTGCA..AC.CG.T.CA..TC.C..GT
34 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GTAA.GAAT.TTGCA..AC.CG.T.CA..TC.C..GT
35 .ACC..GG.TCGA.A.C.TC.T.ACA..TGC....GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
36 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
37 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
38 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
39 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
40 .ACC..GG.TCGA.A.C.TC.T.ACA.CTGCT...GT.A.GAAT.TTGCA..AC..GCT.CA..TC.C..GT
```

Alignment of DENV-3 parsimony informative sites (continued)

```
   11111111111111111111111111111111111111111111111111111111111111111
   00011111111111111122222222222233333333333333333333344444444444444
   89902223345567778135666778899001123334455567777888011334466667 7
   02512581993710265231039891403284792581704654570361361669146736

 1 ACTTTCTGCATGTCAGGTTCTGGGTGCTAGACGATTTCTTAATTATCATCTCCTCTAACTGG
 2 ..............A.............G.....................C......
 3 G....TAAT.CAC.G..CC..A.T..TAGC.T..CA.TCC....GC..CT.T.C..GGTC.A
 4 G....TAAT.CAC.G..CC..A.T..TAGC.T..CA.TCC....GC..CT.T.C..G.TC.A
 5 GTCA.TA.T..A.T...C...AAT.ATAG..TA.CA.TCC....GCT.CT.T....G..C.A
 6 GT.A.TAAT..AC....CC.CAAT.ATA...TA.CA.TCC...AGCT.CT.T....G..C.A
 7 GT.A.TAAT..AC....CC..AAT.ATA...TA.CA.TCC....GCT.CT.T....G..C.A
 8 G..A.TAAT..AC..A.CC..AAT.ATA...TA.CA.TCC....GCT.CT.T....G..C.A
 9 G.CA.TAAT..AC....CC..AAT.ATA...TA.CA.TCC....GCT.CT.T....G..C.A
10 GT.A.TA.T..AC....CC..AAT.ATA...TA.CA.TCC....GCT.CT.T....G..C.A
11 GT.A.TAA...AC....CC..AAT.ATAG..TAGCA.TCC....GC..CT.T...CG..C.A
12 GT.A.TAAT..AC....CC..AAT.ATA...TAGCA.TCC....GCT.CT.T....G..C.A
13 G.....A.T..A.....CC..A.TC.TAG..TA.CG.TCA....GC..CT..TCT.G..CAA
14 G.....ATT..A.....CC..A.TCATAG...A.CG.TCA....GC..CTCT.CT.GGTC.A
15 G.....A.T..A.....CC..A.TCATAG...A.CG.TCA....GC..CT...CT.G..CAA
16 G.....A.T..A.T...CC..T.TCATAG..TA.CG.TCA....GC..CT...CT.G..CAA
17 G...C.A.T..A.....CC..T.TCATAG.G.A.CG..CA....GC..CT....T.G..CAA
18 G...C.A.T..A......C.CT.TCATAG...A.CG..CA....GC..CT....T.G..CAA
19 GT...TAA.C.AC....CC..A.TC..AG..TAGCA.T..GG..GC.G.TA.TC..G..CAA
20 GT...TAATC.AC....CC..A.TC..AC..TA.CACTC.GG..GC.G.TA..C..G..CAA
21 GT...TAA.C.AC....CC..A.TC..GC..TA.CACTC.GG..GC.G.TA..C..G..CAA
22 GT...TAATC.ACT...CC..A.TC..AC..TA.CACTC.GG..GC.G.TA..C..G...AA
23 GT..CTAATC.AC....ACT.A.TC..AC..TA.CGCTC.GG..GC.G.TA..C..G...AA
24 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.GG..G..G.TA..C..G...AA
25 GT...TAATC.AC...TCCT.A.TC..AC..TA.CACTC.GG..G..G.TA..C..G...AA
26 GT...TA.TC.ACT.ATCCT.A.TC..AC...A.CACTC.G.CAG..G.TA..C.CG....A
27 GT...TA.TC.ACT.ATCCT.A.TC..AC...A.CACTC.G.CAG..G.TA..C..G....A
28 GT...TA.TC.ACT.ATCCT.A.TC..AC...A.CACTC.G.CAG..G.TA..C..G....A
29 GT...TA.TC.ACT.ATCCT.A.TC..AC...A.CACTC.G.CAG..G.TA..C..G....A
30 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
31 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
32 GT...TAATC.AC...T.CT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
33 GT...TAATC.AC...T.CT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
34 GT...TAATC.AC...T.CT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
35 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
36 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
37 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
38 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
39 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
40 GT...TAATC.AC...TCCT.A.TC..AC...A.CACTC.G..AG..G.TA..C..G....A
```

Alignment of DENV-4 parsimony informative sites

```
                                                  1111111111111111112222222222222
                                  1122335566789900012333555666788901223344456 67
                                  2814061567510325841378069256436270284703681 43

 1 MALAYSIA/75/P75 215            GGGTTGAACCGCTTGAGGTTTCAGATATTAGACGTAAATTTACGT
 2 MALAYSIA/75/P75 514            .............................................
 3 MALAYSIA/73/P73 1120           .............................................
 4 PHILIPPINES/56/H241            ...CC.TGTTAT.C.C.A.C..CAG.G.CTAG...G....C.TAC
 5 PHILIPPINES/84/12123           A..CC.TG.TAT.C.CAA.C..CAG.G.CTAG..CG....C.TA.
 6 SRI LANKA/78/S44750            ...CC.TGT.AT.C.CAA.C..CAG.G.CTAC...G.GC.CGTAC
 7 MALAYSIA/69/1006               A..CC.TGTTAT.C.C.A.C..CAG.G.CTAT...G.G.CC.TAC
 8 THAILAND/78/D78 017            A..CCATGT.AT.C.C.A.C..CAG.G.CTAC...G.G.CC.TAC
 9 THAILAND/84/D84 024            AA.CCATGT.AT.C.C.ACC..CAG.G.CTAC...G.G.CC.TAC
10 INDONESIA/77/1132              AAACC.TGT.ATCC.C.A.CCTCAG.G.CTA....G.G..A..A.
11 TAHITI/79/S44754               AAACC.TGT.ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
12 PUERTO RICO/82/M03             AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
13 MX/YUCATAN/84/1503             AAACC.TG..ATCC.C.A.CCTCAG.GCCTA.T..G.G..GG.A.
14 MX/83/1420                     AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
15 MX/83/1414                     AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
16 BRAZIL/82/1385                 AAACC.TG..ATCC.C.A.CC.CAG.G.CTA.T..G.G..GG.A.
17 TRINIDAD/82/82211757           AAACC.TG.TATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
18 JAMAICA/83/830886              AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
19 SURINAM/82/824182              AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
20 DOMINICA/81/M44                AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
21 NEW CALEDONIA/81/5489          AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
22 HONDURAS/91                    AAACC.TGT.ATCC.C.A.CCTCAGCGCCTA.T..G.G.CGG.A.
23 COLOMBIA/96/371813             AAACC.TGT.ATCCAC.A.CCTCAGCG.CTA.T..G.G.CGG.A.
24 ECUADOR/94/109                 AAACC.TGT.ATCC.C.A.CCTCAGCGCCTA.T..G.G.CGG.A.
25 EL SALVADOR/93/110             AAACC.TG..ATCCAC.A..CTCAG.G.CTA.T..G.G..GG.A.
26 EL SALVADOR/94/BC6494          AAACC.TG..ATCCAC.A..CTCAG.G.CTA.T..G.G..GG.A.
27 MX/85/1554                     AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.TA.G.G..GG.A.
28 MX/85/1551                     AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.TA.G.G..GG.A.
29 MX/YUCATAN/95/BC5              AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.TA.G.G..GG.A.
30 MX/91                          AAACC.TG..ATCC.C.A.CCTCAG.G.CTA..A.G.G..GG.A.
31 MX/95/111                      AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.TA.G.G..GG.A.
31 MX/Q. ROO/95/BC8               AAACC.TG..ATCA.C.A.CCTCAG.G.CTA.TACG.G..GG.A.
33 MX/YUCATAN/97/BC287            AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.TACG.G..GG.A.
34 MX/YUCATAN/96/BC26             AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.TACG.G..GG.A.
35 PUERTO RICO/90/96              A.AC..CG..ATCC.C.A.CCTCAG.G.CTA.T..G.GC.GG.A.
36 PUERTO RICO/85/M32             AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
37 PUERTO RICO/87/73              AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.A.
38 PUERTO RICO/94/483             AAACC.TG..ATC..C.A.CCTCAG.G.CTA.T..GGGC.GG.A.
39 TRINIDAD/94/9400264            AAA.C.TG..ATCC.CAA.CCTCAG...CTA.T..GGGC.GG.A.
40 BARBADOS/93/9312102            A.ACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.GC.GG.A.
41 COSTA RICA/96/108              A.ACCATG..ATCC.C.A.CCTCA..G.CTA....G.G..GG.A.
42 PUERTO RICO/98/13              AAACC.TG..ATCC.C.A.CCTCAG.GCCTA.T..G.G..GG.AC
43 BAHAMAS/98/08654               .AACC.TG..ATCC.CAACCCTCAG.G.CTA.T..G.G..GG.AC
44 VENEZUELA/95/113               AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.AC
45 MONSERRAT/94/9411972           AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.AC
46 TRINIDAD/99/9908820            AAACC.TG..ATCC.C.A.CCTCAG.G.CTA.T..G.G..GG.AC
```

Alignment of DENV-4 parsimony informative sites (continued)

```
   222333333333333333334444444444444444444444445555555555555555555566
   778000122244567888990122334445556666778888991112223344456667788990 0
   795069912728132457498739281270390125140489580393581713951473625145 7

 1 AACATTCATAACACACCGGAAGGTTGTGCAGCAGCGTTAACACCGATCAATGCGCCCGCCCCACGAG
 2 ....................................................................
 3 ....................................................................
 4 CGTGC.TGCGGTG.GTT.AG.AACCACAT.ATGA.AC.GG.G.....T.G.ATA.T...TTTGT..A
 5 CGTGC.TG.GGTG.GTT.AG..ACCACAT.ATGA.AC.GG.G.T...T.G.ATATT...T.TGT..A
 6 CG.GC..GCGGTG.G.TAAG..ACCACAT.ATGA.ACCGG.G.....T.GCAT..T...TTTGT.GA
 7 CG.GC..G.GGTG.G.TAAG..A.CACAT.ATGATACCGG.G.....T.GCATA.TT..TTTGT..A
 8 CG.GCCTGCGGTG.G.TAAG..ACCACATTATGATACCGG.G.....T.GCATATTT..T.TGT..A
 9 CG.GCCTG.GGTG.G.TAAG..ACCACATTATGATACCGG.G.....T.GCATATTT..TTTGT..A
10 CGTGC.TG..GTGTGTT.AG.AACCACAT.ATGA.AC.TG.GT..GCTGGCA.A.T..TT.TGT..A
11 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TG.GT..GCTGGCA.A.T..TT.TG..GA
12 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
13 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
14 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..CGTGT..GCTGGCA.A.T..TT.TGT..A
15 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
16 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
17 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
18 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
19 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
20 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
21 CGTGC.TG..GTGTGTT.AG..ACCACAT.ATGA.A..TGTGT..GCTGGCA.A.T..TT.TGT..A
22 CGTGC.T...GTGTGTT.AG..ACCACAT.ATGA.A..CGTGTTAGCTGGCA.A.T..TT.TGT..A
23 CGTGC.T...GTGTGTT.AG..ACCACAT.ATGA.A..CGTGTTAGCTGGCA.A.T..TT.TGT..A
24 CGTGC.T...GTGTGTT.AG..ACCACAT.ATGA.A..CGTGTTAGCTGGCA.A.T..TT.TGT..A
25 CGTGG.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..TG.GT..GCTGGCA.A.T.ATTTTGT..A
26 CGTGG.TGC.GTGTGTT.AG..ACCA.AT.ATGA.A..TG.GT..GCTGGCA.A.T.ATTTTGT..A
27 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..TGTGT..GCTGGCA.A.T..TTTTGT..A
28 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..TGTGT..GCTGGCA.A.T..TTTTGT..A
29 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..CG.GT..GCTGGCA.A.T..TTTTGT..A
30 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..CG.GT..GCTGGCA.A.T.ATTTTGT..A
31 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..CG.GT..GCTGGCA.A.T.ATTTTGT..A
31 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..CG.GT..GCTGGCA.A.T..TTTTGT..A
33 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..CG.GT..GCTGGCA.ATT..TTTTGT..A
34 CGTGC.TG..GTGTGTT.AG..ACCA.AT.ATGA.A..CG.GT..GCTGGCA.A.T..TTTTGT..A
35 CGTGC.TG..GTGTGT..AGG.ACCACAT.ATGA.A..TG.GT..GCTGGCA.A.T..TT.TGTA.A
36 CGTGC.TG..GTGTGT..AGG.ACCACAT.ATGA.AC.TG.GT..GCTGGCA.A.T..TT.TGTA.A
37 CGTGC.TG..GTGTGT..AGG.ACCACAT.ATGA.AC.TG.GT..GCTGGCA.A.T..TT.TGTA.A
38 CGTGC.TG..GTG.GT..AGG.ACCACAT.ATGA.A..TG.GT.AGCTGGCA.A.T..TT.TGTA.A
39 CGTGC.TG..GTGTGT..AGG.ACCACAT.ATGA.A..TG.GT..GCTGGCA.A.T..TT.TGTA.A
40 CGTGC.TG..GTGTGT..AGG.ACCACAT.ATGA.A..TG.GT..GCTGGCA.A.T..TT.TG.A.A
41 CGTGC.TG..GTGTGT..AGG.ACCACAT.ATGA.A..TG.GT..GCTGGCA.A.T..TT.TG.A.A
42 CGT.C.TG..GTGTGT..AG..ACCACAT.ATGA.A..TG.GT...CTGGCA.A.T..TT.TG.A.A
43 CGT.C.TG..GTGTGT..AG..ACCACAT.ATGA.A..TG.GT...CTGGCA.A.T..TT.TG.A.A
44 CGT.C.TG..GTGTGT..AG..ACCACAT.ATGA.A..TG.GT...CTGGCA.A.T..TT.TG.A.A
45 CGT.C.TG..GTGTGT..AG..ACCACAT.ATGA.A..TG.GT...CTGGCA.A.T...T.TG.A.A
46 CGT.C.TG..GTGTGT..AG..ACCACAT.ATGA.A..TG.GT...CTGGCA.A.T...T.TG.A.A
```

Alignment of DENV-4 parsimony informative sites (continued)

```
   6666666666666666666666777777777777888888888888888888999999999999999
   0011223344455677888999122235568899000112234444556677012223345668889
   8925170925814128047079703620953635147362813679251409681470351691450

 1 GAAATCACATACGATACACTTTGCAATCACCTATTGGTTCCCGCCAGCGCAGCTCCCGAGCGGACAG
 2 .......................................................................
 3 .......................................................................
 4 AG.G..GTG..TAGA...T..CA.GT.G.TT..C.AA.CT..AT.GA.T.GA.CAT.AGATAAGTGA
 5 AG.G..GTG.GTAGA...T..CA.GT.G.TT..C.AA.CT..AT.GA.T.GA.CAT.AGATTA.TGA
 6 AG.G..GTG..TA.AG..T.CC..GC.G.TTA.CCAA.CT..ATTG..T.GA.CATTAGATAAGTGA
 7 AG.G..GTG..TA.AG..T.CCA.GC.G.TTA.CCAA.CT..ATT...TTG..CATTAGATAAGTGA
 8 AG.GC.GTG..TA.AG..TCCCA.GT.G.TTA.CCAA.CT.TAT.T.TTTG..CATTAGATAA.TGA
 9 AG.GC.GTG..TA.AG..TCCCA.GT.G.TTAGCCAA.CT.TA.T..TTTGA.CATTAGATA....A
10 AG...TGTG..TA.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGATCAT.AGAT.A.TGA
11 AGG.CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
12 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
13 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
14 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
15 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
16 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
17 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
18 AG..CTGTG.GTT.A..GTCCCATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
19 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
20 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.T..CTGAT.AT.AGAT.A.TGA
21 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
22 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
23 AG..C.GTG.GTT.A..GTC.CATGT.GGTT.GC.AACCTT.AT.C..C.GAT.AT.AGAT.A.AGA
24 AG..CTGTG.GTT.A..GTC.CATGT.GGTT...CAACCTT.AT.C..C.GAT.AT.AGAT.A.TGA
25 AG..CTGTG.GTT.A..GTC.CA.GT.GGTT.G.CAACCTT.AT.C..C.GAT.AT.AGAT.A.TGA
26 AG..CTGTG.GTT.A..GTC.CA.GT.GGTT.G.CAACCTT.AT.C..C.GAT.ATTAGAT.A.TGA
27 AG..CTGTG.GTT.A..GTC.CA.GT.G.TT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
28 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
29 AG..CTGTG.GTT.A..GTC.CATGT.GGT..G.CAA.CTT.AT.C..CTGATCAT.AGAT.A.TGA
30 AG..CTGTG.GTT.A.TGTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGATCAT.AGAT.A.TGA
31 AGG.CTGTG.GTT.A.TGTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGATCAT.AGAT.A.TGA
31 AG..CTGTGCGTT.A..GTC.CATGTCGGTT.G.CAACCTT.AT.C..CTGATCAT.AGAT.A.TGA
33 AG..CTGTGCGTT.A..GTC.CATGTCGGTT.G.CAACCTT.AT.C..CTGATCAT.AGAT.A.AGA
34 AG..CTGTGCGTT.A..GTC.CATGTCGGTT.G.CAACCTT.AT.C..CTGATCAT.AGAT.A.TGA
35 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CA.CCTT.AT.C..CTGAT.AT.AGAT.A.TGA
36 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..C.GAT.ATTAGAT.A.TGA
37 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..C.GAT.ATTAGAT.A.TGA
38 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CT.ATCA..AGAT.A.TGA
39 AG..CTG.G.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CT.A.CA..AGAT.A.TGA
40 AG..C.GTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..C.GAT.AT.AGAT.A.TGA
41 AG..C.GTG.GTT.A..GTCCCATGC.GG.T.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
42 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A..GA
43 AG..CTGTG.GTT.G..GTC.CATGT.GGTT.G.CAACCTT.AT.C..C.GAT.AT.AGAT.A.TGA
44 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
45 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
46 AG..CTGTG.GTT.A..GTC.CATGT.GGTT.G.CAACCTT.AT.C..CTGAT.AT.AGAT.A.TGA
```

Alignment of DENV-4 parsimony informative sites (continued)

```
    1111111111111111111111111111111111111111111111111111111111111111111
    9900000000000000000000000000011111111111111111111222222222222222233
    9900011222244555566667788899900011234445567888999112333445678889900
    6935849346947013624581703902815736573560516258147584013357922576928

 1  ACATAAACAGTTACATTTTTCTACCCAATTCCCCCGCCTGTGTGATTCGTTTAAATGACCTTTAGAG
 2  ..............................T...............................
 3  ..............................T...............................
 4  .TG.CGGTG.CAG..CC.CCTGGT.TG.CATA...ATT...ACAG.CTCACCGC.CACTTCCCGAGA
 5  G.G.CGGTG.CAGTGCC.CCTGGT.TG.CATA.T.ATT...ACAG.CTCAC.GC.CACTTCCCGAGA
 6  .TG.CGGTG.C.G..CC.CCT.GT.TGGCATG..TATTAA.A.AG.CTCACCGCCCATTTCCCGAGA
 7  .TG.CGGTG.C.GT.CC.CCT.GT.TGG.ATG..TATTA..ACAG.CTCACCGCCCATTTCCCGAGA
 8  .TGCCGGTG.C.GT.CC.CCT.GT.TGG.ATG.TTATTA..ACAGCCTCACCGCCCATTTCCCGAGA
 9  .TGCCGGTG.C.GT.CC.CCTGGT.TGG.ATGTTT.TTG..ACAGCCTCAC.GCCCATTTCCCGAGA
10  G.GCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.A.ACAG.CTCAC..C.CATT.C.CGAG.
11  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
12  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
13  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATAT...TT.ACACAG.CTCAC..CGCATT.C.CGAG.
14  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
15  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
16  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT...CGAG.
17  GTGCCGGTGAC.GT.CCCCCTGG..TG.CATA....TT.ACA.AG.CTCAC..C.CATT.C.CGAG.
18  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
19  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
20  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
21  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
22  GTGCCGGTGAC.G..CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
23  GTGCCGGTGAC.GT.CCCC.TGGT.TG.CA.A....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
24  GTGCCGGTGAC.GT.CCCC.TGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
25  GTGCCGGTGAC.GT.CC.CCTGGT.TG.CATA...ATT.ACACAG.CTCAC..C.CATT.C.CGAG.
26  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TTCACACAG.CTCAC..C.CATT.C.CGAG.
27  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
28  GTGCCGGTGAC.GT.CCCCCTGGT.TG.CATA....TT.ACACA..CTCAC..C.CATT.C.CGAG.
29  GTGCCGGTGA..GTGCCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CACT.C.CGAG.
30  GTGCCGGTGAC.GTGC.CCCTGGTTTG.CATA....TT.ACACAG.CTCAC..C.CACT.C.CGAG.
31  GTGCCGGTGAC.GTGCCCCCTGGTTTG.CATA....TT.ACACAG.CTCAC..C.CACT.C.CGAG.
31  GTGCCGGTGA..GTGCCCCCTGGTTTG.CATA....TT.ACACAG.CTCAC..C.CACT.C.CGAG.
33  GTGCCGGTGA..GTGCCCCCTGGTTTG.CATA....TT.ACACAG.CTCAC..C.CACT.C.CGAG.
34  GTGCCGGTGA..GTGCCCCCTGGTTTG.CATA....TT.ACACAG.CTCAC..C.CACT.C.CGAG.
35  GTGCC.GTGAC.G.GCCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.C.TT.C.CGAG.
36  GTGCCGGTGAC.G.GCCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
37  GTGCCGGTGAC.G.GCCCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
38  GTGCCGGTGAC.G.GCCCCCTGGT.TG.CATA....TTCA.ACAG.CTCAC..C.CATT.C.CGAG.
39  GTGCCGGTGAC.G.GCCCCCTGGT.TG.CATA..T.TT.ACACAG.CTCAC..C.CATT.C.CGAG.
40  GTGCCGGTGAC.G.G.CCCCTGGT.TG.CATA....TT.ACACAGCCTCAC..C.C.TT.C.CGAG.
41  GTGCCGGTG.C.G.G.CCCCTGGTTTG.CATA....TT.ACACAGCCTCAC..C.C.TT.C.CGAG.
42  GTGCCGGTGAC.G.G.CCCCTGGT.TG.CATA....TT.ACACAG.CTTAC..C.CATT.C.CGAG.
43  GTGCCGGTGAC.G.G.CCCCTGGT.TGGCATA.T..TT.ACACAG.CTCAC..C.CATT.C.CGAG.
44  GTGCCGGTGAC.G.G.CCCCTGGT.TG.CATA....TT.ACACAG.CTCAC..C.CATT.C.CGAG.
45  GTGCCGGTGACCG.G.CCCCTGGT.TG.CATA....TT.AC.CAG.CTCAC..C.CATT.C.CGAG.
46  GTGCCGGTGACCG.G.CCCCTGGT.TG.CATA....TT.AC.CAG.CTCAC..C.CATT.C.CGAG.
```

Alignment of DENV-4 parsimony informative sites (continued)

```
   1111111111111111111111111111111111111111
   3333333333333333334444444444444444444444
   1222336677778888990011123355556667777 88
   1369283512470139284703982425891570369 25

 1 TCTCCAACTCCTGATAGCGCAATATAACCTACTTTGCAA
 2 ....................................
 3 ....................................
 4 CT.GTTGTC.A..TCG.T.TG.C.AG..TCG.G.CATC.
 5 CT.GTTGTC.A..TCG.T.TG.C.AG..TCGAG.CATC.
 6 CT.GTTGTC.AC.TC..T.T.....G.TT.G.G.CATT.
 7 CT.GTTGTC.AC.TCG.T.T.....G.TT.G.G.CATT.
 8 C..GTTGTC.AC.TCG.T.T.....G..TCG.GCCATT.
 9 C..GTTGTC.AC.TCG.T.T...T.G.TTCG.GCCATT.
10 CA.GTC.TC.A.CTCG.T...G.TAG..TCG.GCCAT..
11 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
12 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
13 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
14 CA.GTC.TC.A..TCG.T...GCTAG..TCG.GCCAT..
15 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
16 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
17 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
18 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
19 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
20 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
21 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
22 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCA...
23 CA.ATC.T..A..CCG.T...G.TAG..TCG.GCCAT..
24 CA.GTC.TC.A..CCG.T...G.TAG..TCG.GCCAT..
25 CA.GTC.TC.A..TCGAT...G.TAG..TCG.GCCA...
26 CA.G.C.TC.A..TCGAT...G.TAG..TCG.GCCAT..
27 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
28 CA.G.C.TC.A..TCG.T...G.TAG..TCG.GCCAT..
29 CA.G.C.TC.A.TTCG.T...G.TAG..TCG.GCCAT..
30 CA.G.C.TC.A.TTCG.T...G.TAG..TCG.GCCAT..
31 CA.G.C.TC.A.TTCG.T...G.TAG..TCG.GCCAT..
31 CA.G.C.TC.A.TTCG.T...G.CAG..TCG.GCCAT..
33 CA.G.C.TC.A.TTCG.T...G.CAG..TCG.GCCAT..
34 CA.G.C.TC.A.TTCG.T...G.CAG..TCG.GCCAT..
35 CACGTC.TCTA..TCG.T...G.TAG..TCG.GCCAT.G
36 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT.G
37 CA.GTC.TC.A..TCG.T...G.TAG..TCG.GCCAT..
38 CACGTC.TC.A..TCG.TA..G.TAGT.TCG.GCCAT.G
39 CACGTC.TCTA..TCG.TA..G.TAGT.TCG.GCCAT.G
40 CA.GTC.TC.A..TCG.T...G.CAG..TCG.GCCAT.G
41 CA.GTC.TC.A..TCG.TA..G.CAG..TCG.GCCAT.G
42 CACGTC.TC.A..TCG.T...G.CAG..TCGTGCC.T.G
43 CA.GTC.TC.A..TCG.T...G.CAG..TCGTGCCAT.G
44 CA.GTC.TC.A..TCG.T...G.CAG..TCGTGCCAT.G
45 CA.GTC..C.A..TCG.T...G.CAG..TCGTGCCAT.G
46 CA.GTC..C.A..TCG.T...G.CAG..TCGTGCCAT.G
```