IDENTIFICATION OF THE INTRA-SEROTYPIC DIVERSITY IN DENV-1 BASED ON THE MONOPHYLY CRITERION

AYDA SUSANA ORTIZ BAEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS BÁSICAS
ESCUELA DE BIOLOGÍA
BUCARAMANGA
2012

IDENTIFICATION OF THE INTRA-SEROTYPIC DIVERSITY IN DENV-1 BASED ON THE MONOPHYLY CRITERION

AYDA SUSANA ORTIZ BAEZ

Trabajo de grado para optar al título de Biólogo

Director
Daniel Rafael Miranda-Esquivel
Raquel Elvira Ocazionez Jimenez

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS BÁSICAS
ESCUELA DE BIOLOGÍA
BUCARAMANGA
2012

ACKNOWLEDGMENTS

The authors wish to thank to Christian Julián Villabona-Arenas and Susy Echeverría Londoño for their valuable corrections and contributions to this project. This project corresponds to the first author's undergraduate thesis at Universidad Industrial de Santander.

To my family and especially my mother and sister for their love and support

CONTENT

FIGURES LIST

Pág.

TABLES LIST

Pág.

TÍTULO: Identificación de la Diversidad Intra-serotípica en DENV-1 con base en el Criterio de Monofilia *

AUTOR: Ayda Susana Ortiz Báez**

PALABRAS CLAVES: Dengue, clasificación, filogenia, genotipo, monofilia

CONTENIDO:

La clasificación de mayor prevalencia sobre la diversidad intra-serotípica en dengue tipo 1 reconoce cinco genotipos con base en un criterio de máxima divergencia nucleotídica del 6%. Aunque los genotipos difieren en su distribución geográfica, la variación intra-serotípica es frecuentemente asociada con dinámicas espaciales y temporales. En este estudio, evaluamos el criterio de monofilia para determinar la diversidad intra-serotípica en dengue tipo 1, además describimos la forma como patrones espacio-temporales estructuran la diversidad de secuencias y sus relaciones filogenéticas. Para direccionar estos componentes, comparamos 993 secuencias completas del gen E a partir de un muestreo extensivo de cepas aisladas en todo mundo e incluimos 51 nuevas secuencias para el nororiente de Colombia. El set de datos se seleccionó con base en un análisis de agrupamiento en el cual un umbral de identidad de 99.5% fue el valor de corte más informativo, representando los datos espaciales y temporales de los aislados virales muestreados. La reconstrucción de las topologías fue realizada a partir de acercamientos filogenéticos y basados en distancias. Así mismo las hipótesis construidas fueron comparadas a partir de un test de selección de árboles bajo el criterio de máxima verosimilitud. En este estudio, caracterizamos seis genotipos, los cuales fueron reconocidos bajo la mayoría de acercamientos como grupos monofiléticos. Estos genotipos compartieron patrones de distribución parciales aunque no se identificó un área común de circulación. El análisis de agrupamiento evidenció una asociación estructurada inicialmente por espacialidad y sub-estructurada por patrones temporales, los cuales correspondieron a la ocurrencia de períodos epidémicos generalmente evidentes a nivel local. Finalmente, identificamos las secuencias generadas en este estudio dentro del genotipo IV, agrupandose con secuencias correspondientes a América Latina.

* Trabajo de grado

** Facultad: Ciencias Básicas. Escuela: Biología.
Director: Daniel Rafael Miranda-Esquivel. Codirector: Raquel Elvira Ocazionez

TITLE: Identification of the Intra-serotypic Diversity in DENV-1 Based on the Monophyly Criterion*
AUTHOR: Ayda Susana Ortiz Báez**
KEYWORDS: Dengue, classification, phylogeny, genotype, monophyly

DESCRIPTION:

The most prevalent classification of the intra-serotypic diversity in dengue type 1 recognizes five genotypes based on a genetic distance criterion with maximum nucleotide divergence of 6%. Although the genotypes differ in their geographic distribution, the intra-genotypic variation is often associated with spatial and temporal dynamics. In this study, we tested the monophyly criterion to determine the intra-serotypic diversity in dengue type 1, also we described how spatio-temporal patterns structure the sequence diversity and phylogenetic relationships. To address these issues, we compared 993 complete E gene sequences from an extensive sampling of worldwide strains and including 51 new Colombian sequences from northeastern Colombia. We selected the dataset based on a clustering analysis in which a threshold of 99.5% was the most informative cut-off value, representing the spatial and temporal data of viral isolates sampled. Tree reconstruction was performed using phylogenetic and distance-based approaches whereas topological hypotheses were compared using a tree selection test under a maximum likelihood framework. Here we characterized six genotypes, which were recognized under most approaches as monophyletic groups. These genotypes shared partial distribution patterns but we did not identify a common area of circulation among all of them. Likewise, clustering analysis provided evidence that dengue sequences are associated first by spatial then temporal patterns. Temporal patterns corresponding to the occurrence of epidemic periods were usually evidenced at local level. We identified the Colombian sequences isolated in this study into the genotype IV which grouped together with other Latin American sequences.

* Research project
** Facultad: Ciencias Básicas. Escuela: Biología.
Director: Daniel Rafael Miranda-Esquivel. Codirector: Raquel Elvira Ocazionez

AUTHOR SUMMARY

Dengue viruses (DENV) are the infectious agents responsible of the most common mosquito-borne viral disease of humans. These viruses are classified in four distinct serotypes and each of them exhibits a genetic diversity that is further grouped into genotypes. Here, we focused on dengue type 1 (DENV-1) which traditionally has been classified based on a value of 6% of nucleotide divergence. This research assesses the monophyletic status of genotypes in DENV-1 and proposes monophyly as the criterion to delimit its intra-serotipic diversity, emphasizing on the phylogenetic relationships among strains and the recognition of natural groups. We explored different tree reconstruction approaches which consistently identified the same genotypes but differences amongst the topological arrangements. Further, we showed the interplay of spatial and temporal patterns and the level on which each one of them structure the intra-serotypic diversity and phylogenetic relationships in DENV-1. Our findings provide a stable classification of DENV-1, its possible evolutionary relationships and new insights of its spatio temporal dynamics, that could lead to advances in developing more effective surveillance programs and understanding of the evolution and the epidemiology of dengue.

INTRODUCTION

Dengue viruses (DENV) are infectious agents classified within the genus *Flavivirus*(family Flaviviridae), and are most widely spread throughout tropical and subtropical regions [1]. In last decades, changes in their geographic distribution indicate the movement of viruses to new localities and reemergence in previous epidemic areas [1, 2, 3, 4, 5]. These changes are influenced by the combination of various determinants related to ecological, human, vectorial and viral factors [6, 7]. Given that dengue is not contagious from person to person, the viral transmission in humans is mainly mediated by the bite of *Aedes* mosquitoes (family Culicidae) in urban and peridomestic cycles [8]. The distribution of these mosquitoes is critical to the distribution of arboviruses such dengue [9,10,11,12]. Structurally, these flavivirus are enveloped viruses, which are characterized by a positive-sense, single-stranded RNA genome, that encodes a polyprotein organized into three structural and seven nonstructural proteins [13]. DENV viruses are classified into four antigenically and phylogenetically distinct [14], but closely-related, serotypes (DENV1-4). The infection can lead to different manifestations of the disease ranging from non-severe dengue with/without warning signs to severe dengue [15]. As typically occurs with RNA viruses, dengue exhibits substantial intra-serotypic diversity that is arranged into groups, so-called genotypes [14, 16]. Conventionally dengue type 1 (DENV-1) has been classified into five genotypes characterized depending on the author [17, 18] and based on partial or complete nucleotide sequences of the envelope (E) gene. The traditional definition of a dengue genotype establishes groups of viruses having no more than 6% sequence divergence among them [17]. This definition, does not always identify historical groups in terms of monophyletic entities [19, 20]. The monophyly criterion establishes a monophyletic taxon as a group comprised of the most recent common ancestor and all its descendants [21, 22]. The concept is based on the recognition of systems of common ancestry [21], and is frequently used to delimit taxonomic units as natural groups that share a common history [23,24,25]. Although, the requirement that taxa be monophyletic can be considered merely semantic [26], the evidence of monophyly based on members of groups more closely related to each other than to anything

outside the group [26], can be a close approximation for grouping the intra-serotypic diversity within Dengue.

Depending of the general framework, the intra-serotipic variation in DENV-1 has been reconstructed from different phylogenetic or distance-based approaches [27,28,29]. Among the available methods, maximum likelihood (ML) is one of the most commonly accepted approaches given the implementation of evolutionary models to describe the patterns of DNA base substitution and the estimation of branch lengths [30, 31, 32]. In general, the likelihood function is well known by its powerful statistical basis for phylogenetic inference [32, 33]. Differences in score among the ML tree and an alternative hypothesis generated under any tree reconstruction method can be statistically compared, by performing congruence tests, such as the Shimodaira-Hasegawa (SH) test [34]. The test establishes whether differences between competing hypotheses are significant in likelihood or whether all tree hypotheses are equally good explanations of the data.

The intra-serotypic diversity in DENV-1 has been associated with spatial and temporal dynamics [35, 36, 28]. Genotypes have a varying geographical distribution, determined by the extensive movement of viruses among different localities, and their spatial range varies from widely distributed to geographically restricted genotypes [37]. Similar strains circulating during the same period or location tend to cluster together [38, 39]. However, these patterns can be influenced by different factors, such as time and geographic scales, viral dispersion, introduction of strains, local evolution, amongst others, which can form different grouping patterns in phylogenetic analyses [40, 39, 41]. Here we characterized the intra-serotypic diversity of DENV-1 based on the monophyly criterion. We build a dataset using worldwide sequences and including 51 new sequences from northeastern Colombia. Our results identified six genotypes as monophyletic groups which are stable units among methods although with differences in branching patterns. In addition, we showed the incidence of the spatio-temporal patterns in structuring the sequence diversity and phylogenetic relationships in DENV-1.

MATERIALS AND METHODS

COLOMBIAN ISOLATES

Fifty one dengue strains were obtained from serum samples of patients. The samples were collected by the Laboratorio de Arbovirus in the Centro de Investigaciones Enfermedades Tropicales (CINTROP) and taken from different localities of Santander and Norte de Santander in the Northeastern region of Colombia during the period 1998-2008 (Table 1). The viruses were isolated in *Aedes albopictus* cell monolayers (clone C6/36) maintained in L-15 Leibovitz Medium and supplemented with Fetal Bovine Serum (FBS), tryptose phosphate (10%) and Penicillin antibiotic. The procedure followed Ocazionez *et al* [42]. Infected cells were examined to identify the presence of cytopathic effect (CPE) and dengue viruses were typed by indirect immunofluorescence assay (IFA) using monoclonal antibodies (CDC, Puerto Rico).

RNA EXTRACTION AND RT-PCR

After the second passage to C6/36 cells, viral RNA was extracted from the medium supernatant of infected cultures using QIAamp Viral RNA Mini Kit (QIAGEN). The RNA reverse transcription to cDNA by RT-PCR was performed using SuperScript III Reverse Transcriptase (Invitrogen) and a mixture of random hexaprimers provided in the kit.

PCR AMPLIFICATION AND NUCLEOTIDE SEQUENCING

The E gene sequences were amplified by PCR reaction was the E gene (1485 bp). This gene was amplified in two overlapped fragments using GoTaq DNA polymerase (Promega, Madison, WI, USA), and oligonucleotide primers (Table S1) were designed by the authors based on conserved regions from a representative collection of sequences from GenBank. The size of PCR products were visualized in 2% agarose gel by electrophoresis and fragments were compared with Hyperladder II marker (Bioline). Amplified products were purified using the Wizard PCR Preps DNA Purification system (Promega), and sequencing reactions were conducted at Macrogen Corp. (USA), under BigDyeTM terminator cycling conditions and using 3730XL automatic sequencer. Nucleotide sequences were assembled with Lasergene software 9.0 (DNASTAR), and manually inspected. Finally, to confirm the taxonomic identity of all the sequences generated in this study, they were compared with available sequences using BLAST, and then submitted to the GenBank sequence database (Table S2).

CONSTRUCTION OF THE DATASET

The dataset was constructed from the 51 nucleotide dengue sequences generated in this study and combined together with all those DENV-1 sequences available in GenBank and published up to September 2011. The selection of reference sequences corresponded to strains isolated worldwide with the complete E gene and partial or full genome sequences containing this gene. Furthermore, the selection of sequences was based in the availability of spatial and/or temporal data. In vitro mutants, chimeric, recombinant, and clones sequences were discarded based on the information available in the database.

CLUSTERING ANALYSIS

To build a free redundancy dataset the analysis was conducted using USEARCH v5.0 software [43]. First, sequences were sorted by decreasing length and those identical and chimeric were deleted from the dataset. Based on UCLUST clustering algorithm, different identity threshold from 90 to 99.9% were evaluated. The selection of the appropriated identity threshold was addressed following the spatial and temporal trend of sequences to clustering (Figure 1). From these sequences the E gene region was selected for subsequent analyses.

ALIGNMENT AND MODEL SELECTION

From the representative dataset, sequences were aligned under the multiple sequence alignment algorithm implemented in MUSCLE v3.8 [44], under the default parameters. The statistical selection of best-fit model of nucleotide substitution was performed using Akaike and Bayesian information criteria implemented in jModeltest v0.1.1 [45]. The model was assessed including and excluding the outgroups in the dataset to determine if there was a bias by taking into account them in the selection of the model. HlRT was non considered due to some problems related with the model selection uncertainty as documented by Posada and Buckley, 2004 [46].

DISTANCE AND PHYLOGENETIC ANALYSIS

The phylogenetic relationships among viral sequences were reconstructed under the criteria of parsimony, maximum-likelihood and bayesian inference. The parsimony analysis was performed using a ratchet search and bisection-reconnection (TBR) branch swapping on the Wagner trees. To estimate the support values, bootstrapping was based on 1000 replications and the values were vizualized on the strict consensus. Trees were inferred using TNT v1.1 program [47]. The Maximum-likelihood phylogeny was esti-

mated under GTR+γ model and the starting tree was build using bioneighbor-joining method [48], implemented in PhyML v3.0 software package [49]. Standard and rapid non-parametric bootstrap with 1000 pseudo-replicates was used for estimating branch support on the phylogeny using RAxML v7.2.8 [50]. The Bayesian phylogenetic tree was reconstructed with MrBayes v3.1.2 program [51], and posterior probabilities were determined from two independent runs of four Markov chains each and $7x10^7$ generations. We used the default priors. The sampling frequency was every 7000 generations, the 25% of the generations were burn-in. Convergence was assessed using the average standard deviation of split frequencies across independent analyses and Bayesian MCMC runs were also analyzed using TRACER v1.5 [52]. The distance-tree was constructed from a genetic distance matrix based on the Neighbor-Joining method [53] with a bootstrap of 1000 replicates. The analysis was conducted in the R package Ape v3.0-3 [54].

To test alternative phylogenetic hypotheses, we used the one-tailed Shimodaira-Hasegawa (SH) test [34]. One thousand replicates were performed by resampling the estimated log-likelihoods for each site (RELL model) as implemented in PAUP v4.0 [55]. Genotypes were defined by following the next steps: 1. Identification of monophyletic groups in tree reconstructions. 2. Stability of the groups identified among the different approaches assessed. The stability was evaluated in terms of topological congruence and recovery of groups. 3. Recognition of a characteristic temporal or spatial structure of each group. The spatial units identified in this study corresponded to geopolitical divisions.

RESULTS

SEQUENCES AND CLUSTERING ANALYSIS

The length of the amplified product corresponding to the E gene sequence from fifty-one Colombian isolates was 1485 base pairs (bp). These sequences were deposited in GenBank under accession numbers JQ581602 to JQ581652 (Table S2). The base composition expressed in terms of the overall GC content showed differences between Colombian sequences and the complete dataset, with values of 46.50% and 47.43% respectively. The comparison of sequences to establish unique and informative terminals yield an identity threshold of 99.5%. Given these results, 993 representative and unique sequences were selected (Table S3). At this threshold, it was possible to recognize geographic and temporal patterns among the different clusters. Lower thresholds were discarded due to the loss of fine-scale resolution within clusters, which were very inclusive and clustered strains from different geographical localities and sampled times (Figure 1A and B). Higher thresholds values identified most of the individual localities where dengue virus has been isolated, however, many clusters represented a single locality and time sampled, increasing the amount of redundant sequences into the data set (Figure 1C). Using the selected identity level (99.5%), we attempted to represent and maximize the geographic and temporal structure in the data set (Figure 1D).

MODEL SELECTION

The selected model of nucleotide substitution was GTR+γ under both Akaike (AIC) and Bayesian information criteria (BIC). The inclusion of the outgroups yield GTR+γ+I under AIC.

PHYLOGENETIC ANALYSES

Our findings, based on all DENV-1 worldwide strains showed six, but differentially-supported monophyletic groups that were, geographically defined under most the phylogenetic and distance-based methods evaluated. The six genotypes were defined as follows: (I) Asia and Djibouti, subdivided into: (IA) Southeast Asia and East Asia, (IB) Arabian Peninsula, East Asia, Indochina and Djibouti, (II) Japan and Hawaii, (III) Southeast Asia, East Asia and Oceania, (IV) Latin America including Caribbean islands, East and South East Asia, India (South Asia), Arabian Peninsula and Africa, (V) Malaysia of sylvatic origin (VI) Thailand in 60s (Figure 2).

According to the spatial distribution, genotypes IB and IV were widely distributed, but they did not circulate through continuous regions, and some localities with dengue transmission were distantly located from each other, for example, dengue transmission between the Arabian Peninsula and some localities of East Asia, where genotype IB circulated. However, temporal limits were noticeably well-defined among isolates from these regions. Groups of sequences isolated from the same or nearby localities and circulating in the same or similar periods, often showed be related by temporal patterns (Figure 2). Genotype IV was characterized as the cosmopolitan group, having a nearly worldwide distribution. Despite differences in the spatial distribution, genotypes had common geographic patterns amongst them (Figure 3). Even though, all genotypes shared part of its distribution range with at least another genotype, we did not identify a common isolation area where they were all represented.

Although there were topological incongruences, the most genotypes were recovered in all tree reconstructions (Figure 4), and most approaches showed the genotypes as monophyletic units. Often the placement of the genotypes varied through the topologies and only the genotypes I showed a stable position within all tree reconstructions. Genotype I was classified into subgroups IA and IB, which were always recovered as sister clades under all approaches evaluated and both together represented reciprocally monophyletic groups. This genotype was often related with genotype II except for the Parsimony analysis in which genotype II was related to the remaining intra-serotypic diversity in Dengue (Figure 4B). Genotype III was related in different clades with different genotypes depending of the method, whereas the genotypes IV and V were closely related between them as sister taxa or associated with Genotype VI. Bootstrap values for parsimony and ML usually were higher than 70%, and posterior probabilities

were in the range of 0.6 - 1.0. The phylogenetic relationships among DENV-1 sequences under maximum likelihood (ML) was the only approach in which sequences identified as the genotype VI, were not recovered as a monophyletic group (Figure 4A). Despite the Bayesian reconstruction was high congruent with the ML tree, there were differences in the monophyletic condition of genotype VI and the inclusion of one of the outgroups (DENV-4) into the genotype IV (Figure 4D).

Based on these differences in the branching order of genotypes among all the trees reconstructed, the results of the SH-test rejected most of the alternative tree hypotheses tested, given the significant differences in likelihood among them and the ML tree. However, the test supported the monophyletic composition of Thailand sequences of genotype VI as an equally good explanation of the data (Table 2).

Moreover, given the internal phylogenetic relationships in the genotype IV, we recognized four subgroups consisting of isolates from: (IV-A) Africa, (IV-B) Asia, (IV-C) Asia, Reunion island and a strains from Brazil, (IV-D) Latin America, Caribbean islands and Angola (Figure 5). Colombian sequences isolated in this study were grouped together with Latin American isolates into the subgroup IV-E and were closely related to Venezuelan sequences.

DISCUSSION

As shown in previous studies [39, 56], estimates of pairwise distances were relatively small among closely related strains at the spatial and/or temporal level. Nevertheless, some sequences isolated at different points in time or space were more similar than isolates from the same period or locality (Table S4). This suggests the movement of strains among areas and the circulation of multiples lineages within the same locality [57, 28, 58], or a much lower mutation rate for these dengue viruses [59].

Although, many of the sequences corresponded to serially isolated viral strains, the evaluation of identity thresholds between 95 - 98% in the clustering analysis, showed that the grouping of sequences was driven by their geographical distribution rather than their temporal proximity (Figure 1A). However, higher identity thresholds between 99 - 99.7% showed groups with temporal structure, following the trend of sequences sampled in similar times to cluster together (Figure 1B and C) [60]. It could be explained by an association among strains, where clusters of sequences represent a nucleotide consensus that is conserved through geography, while detailed variations among strains are reflected at the temporal level, representing the periods of circulation. In other words, sequence diversity is first structured by geography and then structured by temporality. These patterns of structuring can have implications in dengue vaccine development and the definition of the appropriate temporal and spatial scale for disease surveillance and control strategies [61, 4].

We characterized the intra-serotypic diversity in DENV-1 based on the monophyly criterion, in contrast to the maximum nucleotide divergence criterion proposed by Rico-Hesse [17] and Goncalvez *et al* [18] and the phylogenetic relationships were prioritized over how divergent and variable the sequences were. Although there were common genotypes recognized under both the divergence and monophyly criteria (Table S5), strains isolated from Japan and Hawaii which were previously identified as genotype I under the divergence criterion, were here identified as an independent evolutionary group and designated as a new genotype (Genotype II). Likewise, genotype II was the subtype with the oldest isolates that dated back to the nineteen forties and was re-

stricted to these Pacific islands [62]. The fact that all genotypes clearly shared part of their geographic distribution might be explained by the co-circulation of strains in particular areas [37]. However, differences in the isolation dates between genotypes showed that they could share the same dstribution but not always the same time (Figure 2). This temporal dynamic might be related with some sort of selective pressure [14,41]. In addition, the circulation of strains could be different at a micro-spatial scale in the same locality [63, 64, 65]. Although the evolutionary and population dynamics of inter/intro genotypes was not clear, it is influenced by the interplay of multiple factors related with the host, vector, and virus [66].

In addition, there were differences in the spread range of the genotypes as ocurred with genotype IV, which was the most widely distributed, while others were more geographically restricted. Increases in the fitness or stochasticity could be plausible explanations [37,67]. Equally important was the continuity of the distribution pattern through space, for example, the genotype I subgroup IB with strains circulating mainly in East Asia and existing as an independent viral population in the Saudi Arabia. Previous studies have pointed to the importation of the virus into the region from different geographic sources and subsequent microevolution in the area [68, 69, 70].

Moreover, in the phylogenetic analyses for DENV-1 as well as in the clustering analysis, temporal patterns were evident (Figure 1 and 2). It is noteworthy, that strains circulating in different regions did not tend to be related just for being isolated in the same year. As the clustering analysis, there is an important geographical level of structuring in the relationships among sequences showing possible that related strains differ in space and time as a result of the continual traffic of viruses through vector and human movement [28, 71, 41].

Based on tree reconstructions, different hypotheses were formulated according to topological incongruences (Figure 4). Although it was no clear the recognition of an ancestral genotype as sister group of the remaining intra-serotypic diversity in DENV-1, we recovered different genotypes as potential candidates under the analyses, and based on their common geographic distribution the most recent common ancestor seemed to be in strains from Asia [64, 72]. In this study the oldest available isolate came from Hawaiii in the forties, which corresponds to the prototype strain, however is possible that strains belonging to another genotype, have circulated in previous periods without any record in public databases, and therefore this might explain the fact that genotype II was not the sister group of the remaining genotypes in DENV1.

The second hypotheses was associated with the variation between Thailand sequences

(genotype VI) as a monophyletic group or as independent lineages, and its relation with genotypes IV and V (Figure 4). This variation could be the result of the limited representation of this genotype or due to sequences with differences in the amount of accumulated change over time, relative to the isolation period and the rate of change [14]. Likewise, possible topological incongruences among phylogenetic reconstructions, regarding to the relative position of particular genotypes, could be influenced by the effect of very short branch lengths which represent few genetic differences among clades [72]. Then, any change could easily result in different groupings instead of reflecting a real genealogical discordance [73]. Although the analyses were based on the genealogical history of a gene, the incongruent phylogenetic signal could be associated with potential incomplete lineage sorting with the genealogical histories of other genes [74].
An interesting result, as has been shown previously [72] is the recovery of Malaysian sylvatic isolates from 33 years apart (1972, 2005) that form a monophyletic group (Figure 2B). Malaysian strains from 2005 may have been maintained in a sylvatic cycle through an enzootic forest cycle [72], but the sylvatic origin of Malaysian isolate from 1972 has been questioned given its proximity to endemic genotypes [75, 72].

Given the variable phylogenetic patterns reconstructed under the approaches evaluated, we initially hypothesized that the evolutionary relationships of DENV-1 could be better represented by the ML reconstruction than by the additional approaches evaluated. The implementation of SH-test showed that only the alternative topology that recovered the monophyly in the genotype VI was not significantly different from ML tree. Given that the only difference between both reconstructions was the monophyletic composition of the genotype VI (sequences from Thailand). The topological difference was not statistically well supported, and the relation between these sequences can change between a monophyletic and non-monophyletic status without a significant change in the likelihood score (Table 2) [34, 76]. Although, the SH-test rejected the bayesian reconstruction, this fact could reflect the bias in BI toward inaccurate branch-length estimates in response to multiple factors such as non-specific priors and the sensitivity of this method to place long-branches at every possible location with similar likelihood [77, 78, 79].
In summary, we suggest a criterion based on the identification of monophyletic groups to characterize the intra-serotipic diversity in DENV-1, in contrast with the maximum divergence criterion traditionally used. Although, we find common genotypes under both criteria, we tried to improve the previous approach based on the recognition of

monophyletic groups and the evolutionary content in the phylogenetic relationships. As noted, diversity can be structured by temporal and geographic components. This structure provides a preliminary way to identify possible relations among sequences. Spatial association holds a predominant place in the clustering and phylogenetic grouping of sequences, reflecting the continual virus dispersion among localities, while temporal clustering tends to happen locally over epidemic periods in affected areas. Our findings showed that the recognition of intra-serotypic diversity in DENV-1 was stable but the pattern of the phylogenetic relationships depended on the method. ML approach generated a close reconstruction of its phylogenetic relationships. Likewise, the evolutionary origin of DENV-1 is supported have occurred in Asia. In addition, although genotypes are mainly spatially defined, their distribution ranges are frequently overlapping showing the circulation of different genotypes in these regions. Future efforts should to establish whether there is a quantifiable association between spatio-temporal structure and phylogenetic patterns, and test the real effect of selective and stochastic process in the phylogenetic relationships of DENV-1.

REFERENCES

1. Gubler DJ (2002) Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends in Microbiology 10(2): 100-103.

2. Laferl H, Szell M, Bischof E, Wenisch C (2006) Imported dengue fever in Austria 1990-2005. Travel Medicine and Infectious Disease 4: 319-323.

3. Rico-Hesse R (2007) Dengue virus evolution and virulence models. Clinical Infectious Diseases 44: 1462-1466.

4. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, et al. (2010) Dengue: a continuing global threat. Nature Reviews Microbiology 8: S7-S16.

5. McElroy KL, Santiago GA, Lennon NJ, Birren BW, Henn MR, et al. (2011) Endurance, refuge, and reemergence of dengue virus type 2, puerto rico, 1986-2007. Emerging Infectious Diseases 17.

6. Holmes EC, Burch SS (2000) The causes and consequences of genetic variation in dengue virus. Trends in Microbiology 8(2): 74-77.

7. Ubol S (2007) Dengue: An emerging health challenge? Siriraj Med J 59: 197-200.

8. Gubler DJ (1998) Dengue and dengue hemorrhagic fever. Clinical Microbiology Reviews 11: 480-496.

9. Hoch AL, Gargan TP, Bailey CL (1985) Mechanical transmission of rift valley fever virus by hematophagous diptera. The American Journal of Tropical Medicine and Hygiene 34: 188-193.

10. Pagès F, Peyrefitte CN, Mve MT, Jarjaval F, Brisse S, et al. (2009) *Aedes*

albopictus mosquito: The main vector of the 2007 chikungunya outbreak in Gabon. PLoS ONE 4: e4691.

11. Jansen CC, Beebe NW (2010) The dengue vector *Aedes aegypti*: what comes next. Microbes and Infection 12: 272-279.

12. Puslednik L, Russell RC, Ballard JWO (2012) Phylogeography of the medically important mosquito *Aedes*(Ochlerotatus) *vigilax* (Diptera: Culicidae) in Australasia. Journal of Biogeography .

13. Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. Annual Review of Microbiology 44: 649-688.

14. Holmes EC, Twiddy SS (2003) The origin, emergence and evolutionary genetics of dengue virus. Infection, Genetics and Evolution 3: 19-28.

15. WHO (2009) Dengue guidelines for diagnosis, treatment, prevention and control. Geneva: World Health Organization, third edition edition.

16. Kim J, Ahn Y, Lee K, Park S, Kim S (2010) A classification approach for genotyping viral sequences based on multidimensional scaling and linear discriminant analysis. BMC Bioinformatics 11: 434.

17. Rico-Hesse R (1990) Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. Virology 174: 479-493.

18. Goncalvez AP, Escalante AA, Pujol FH, Ludert JE, Tovar D, et al. (2002) Diversity and evolution of the envelope gene of dengue virus type 1. Virology 303: 110-119.

19. Wiley EO, Lieberman BS (2011) Phylogenetics: Theory and Practice of Phylogenetic Systematics. Wiley-Blackwell, 2nd edition edition, 432 pp.

20. Estes R, Pregill GK, Camp CL (1988) Phylogenetic relationships of the lizard families: essays commemorating Charles L. Camp. Stanford University Press.

21. Hennig W (1950) Grundzuge einer theorie der phylogenetischen. Systematik Deutscher Zentralverlag, Berlin .

22. Hennig W (1966) Phylogenetic Systematics. University of llinois Press, Urbana. Reissue ed.

23. Donoghue MJ (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. The Bryologist 88: 172-181.

24. Mishler BD (1985) The morphological, developmental and phylogenetic basis os species concepts in bryophytes. The Bryologist 88: 207-214.

25. de Queiroz K, Gauthier J (1990) Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. Systematic Biology 39: 307-322.

26. Velasco J (2009). When monophyly is not enough: Exclusivity as the key to defining a phylogenetic species concept. URL http://philsci-archive.pitt.edu/4397/. This paper is forthcoming in Biology and Philosophy.

27. Descloux E, Cao-Lormeau V, Roche C, Lamballerie XD (2009) Dengue 1 diversity and microevolution, french polynesia 2001-2006: connection with epidemiology and clinics. PLoS Neglected Tropical Diseases 3: e493.

28. Kukreti H, Chaudhary A, Rautela R, Anand R, Mittal V, et al. (2008) Emergence of an independent lineage of dengue virus type 1 (DENV-1) and its co-circulation with predominant DENV-3 during the 2006 dengue fever outbreak in Delhi. International Journal of Infectious Diseases 12: 542-549.

29. Schreiber MJ, Holmes EC, Ong SH, Soh HSH, Liu W, et al. (2009) Genomic epidemiology of a dengue virus epidemic in urban Singapore. Journal of virology 83: 4163-73.

30. Felsenstein J (1981) Evolutionary trees from dna sequences: a maximum likelihood approach. Journal of Molecular Evolution 17: 368–376.

31. Fisher RA (1922) On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character 222: 309–368.

32. Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J Mol Evol 54:

156-65.

33. Holder M, Lewis PO (2003) Phylogeny estimation: traditional and bayesian approaches. Nat Rev Genet 4: 275-284.

34. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16: 11-14.

35. Avilés G, Rowe J, Meissner J, Manzur Caffarena J, Enria D, et al. (2002) Phylogenetic relationships of dengue-1 viruses from Argentina and Paraguay. Arch Virol 147: 2075-2087.

36. Laille M, Roche C (2004) Comparison of dengue-1 virus envelope glycoprotein gene sequences from french polynesia. The American Journal of Tropical Medicine and Hygiene 71: 478-484.

37. Zhang C, Mammen MP Jr, Chinnawirotpisan P, Klungthong C, Rodpradit P, et al. (2005) Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. Journal of Virology 79: 15123-15130.

38. Domingo C, Palacios G, Jabado O, Reyes N, Niedrig M, et al. (April 2006) Use of a short fragment of the c-terminal e gene for detection and characterization of two new lineages of dengue virus 1 in India. Journal of Clinical Microbiology 44: 1519-1529.

39. Weaver SC, Vasilakis N (2009) Molecular evolution of dengue viruses: Contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infection, Genetics and Evolution 9: 523-540.

40. Wang E, Ni H, Xu R, Barrett ADT, Watowich SJ, et al. (2000) Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. Journal of Virology 74: 3227-3234.

41. Raghwani J, Rambaut A, Holmes EC, Hang VT, Hien TT, et al. (2011) Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission. PLoS Pathog 7: e1002064.

42. Ocazionez RE, Cortés F, Villar LA (2009) Vigilancia del dengue basada en el laboratorio: diferencias en el nÃºmero de casos y virus aislados según la recolección del suero y la prueba serológica. Colombia Médica 36: 65-72.

43. Edgar RC (2010) Search and clustering orders of magnitude faster than blast. Bioinformatics .

44. Edgar R (2004) Muscle: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

45. Posada D (2008) jmodeltest: phylogenetic model averaging. Molecular Biology and Evolution 25: 1253-6.

46. Posada D, Buckley TR (2004) Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches over Likelihood Ratio Tests. Systematic Biology 53: 793–808.

47. Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. Cladistics 24: 774-786.

48. Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol 14: 685-695.

49. Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696-704.

50. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-90.

51. Huelsenbeck JP, Ronquist F (2001) Mrbayes: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

52. Rambaut A DA (2007) Tracer v1.4. Available from http://beastbioedacuk/Tracer .

53. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

54. Paradis E, Claude J, Strimmer K (2004) Ape: analyses of phylogenetics and evolution in r language. Bioinformatics 20: 289-290.

55. Swofford DL (2003) PAUP: Phylogenetic Analysis Using Parsimony, version 4.0b10 .

56. Bennett CJ, Webb M, Willer DO, Evans DH (2003) Genetic and phylogenetic characterization of the type ii cyclobutane pyrimidine dimer photolyases encoded by leporipoxviruses. Virology 315: 10-19.

57. Foster JE, Bennett SN, Vaughan H, Vorndam V, McMillan WO, et al. (2003) Molecular evolution and phylogeny of dengue type 4 virus in the caribbean. Virology 306: 129-134.

58. Aquino VH, Amarilla AA, Alfonso HL, Batista WC, Figueiredo LTM (2009) New genotype of dengue type 3 virus circulating in brazil and colombia showed a close relationship to old asian viruses. PLoS ONE 4: e7299.

59. Zhang C, Mammen MP, Chinnawirotpisan P, Klungthong C, Rodpradit P, et al. (2006) Structure and age of genetic diversity of dengue virus type 2 in thailand. Journal of General Virology 87: 873-883.

60. Gray RR, Pybus OG, Salemi M (2011) Measuring the temporal structure in serially sampled phylogenies. Methods in Ecology and Evolution 2: 437-445.

61. Mammen MP Jr, Pimgate C, Koenraadt CJM, Rothman AL, Aldstadt J, et al. (2008) Spatial and temporal clustering of dengue virus transmission in thai villages. PLoS Med 5: e205.

62. Gilbertson WE (1945) Sanitary aspects of the control of the 1943-1944 epidemic of dengue fever in Honolulu. Am J Public Health Nations Health 35: 261â270.

63. Mendez J, Usme-Ciro J, Domingo C, Rey G, Sanchez J, et al. (2010) Phylogenetic history demonstrates two different lineages of dengue type 1 virus in colombia. Virology Journal 7: 226.

64. Patil J, Cherian S, Walimbe A, Patil B, Sathe P, et al. (2011) Evolutionary dynamics of the american african genotype of dengue type 1 virus in india

(1962-2005). Infect Genet Evol 11: 1443-8.

65. Lee K, Lo S, Tan S, Chua R, Tan L, et al. (2011) Dengue virus surveillance in singapore reveals high viral diversity through multiple introductions and in situ evolution. Infect Genet Evol .

66. Kuno G (1997) Factors influencing the transmission of dengue viruses. IN: Gubler, DJ and Kuno, G (eds) Dengue and Dengue Hemorrhagic Fever : 61-88.

67. Holmes E (2009) The evolution and emergence of RNA viruses. Oxford series in ecology and evolution. Oxford University Press, 288 pp.

68. Fakeeh M, Zaki A (2003) Dengue in jeddah, Saudi Arabia, 1994-2002. Dengue Bulletin 27: 13-18.

69. Ayyub M, Khazindar A, Lubbad E, Barlas S, Alfi A, et al. (2006) Characteristics of dengue fever in a large public hospital, Jeddah, Saudi Arabia. Journal of Ayub Medical College Abbottabad 18: 9-13.

70. Zaki A, Perera D, Jahan SS, Cardosa MJ (2008) Phylogeny of dengue viruses circulating in Jeddah, Saudi Arabia: 1994 to 2006. Tropical Medicine and International Health 13(4): 584-592.

71. Vazquez-Prokopec GM, Kitron U, Montgomery B, Horne P, Ritchie SA (2010) Quantifying the spatial dimension of dengue virus epidemic spread within a tropical urban environment. PLoS Negl Trop Dis 4(12): e920.

72. Teoh B, Sam S, Abd-Jamil J, AbuBakar S (2010) Isolation of ancestral sylvatic dengue virus type 1, malaysia. Emerg Infect Dis 16: 1783-5.

73. Hipp AL, Hall JC, Sytsma KJ (2004) Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. Systematic Biology 53: 81-89.

74. Rosenfeld JA, Payne A, DeSalle R (2012) Random roots and lineage sorting. Molecular Phylogenetics and Evolution 64: 12–20.

75. Vasilakis N, Fokam EB, Hanson CT, Weinberg E, Sall AA, et al. (2008) Genetic and phenotypic characterization of sylvatic dengue virus type 2 strains. Virology 377: 296-307.

76. Buckley TR, Simon C, Shimodaira H, Chambers GK (2001) Evaluating hypotheses on the origin and evolution of the New Zealand Alpine Cicadas (Maoricicada) using multiple-comparison tests of tree topology. Molecular Biology and Evolution 18: 223-234.

77. Brown JM, Hedtke SM, Lemmon AR, Lemmon EM (2010) When trees grow too long: Investigating the causes of highly inaccurate bayesian branch-length estimates. Systematic Biology 59: 145-161.

78. P G, POL D (2005) Parsimony and bayesian phylogenetics. In: Parsimony, Pylogeny, and Genomics : 148-159.

79. Kolaczkowski B, Thornton JW (2009) Long-branch attraction bias and inconsistency in bayesian phylogenetics. PLoS ONE 4(12): e7891.

TABLES

Table 1. Identity thresholds assessed in clustering analyses

Id Thresholds	seeds
90	5
95	13
98	101
98.5	247
99	446
99.5	996
99.7	1368
99.9	1764

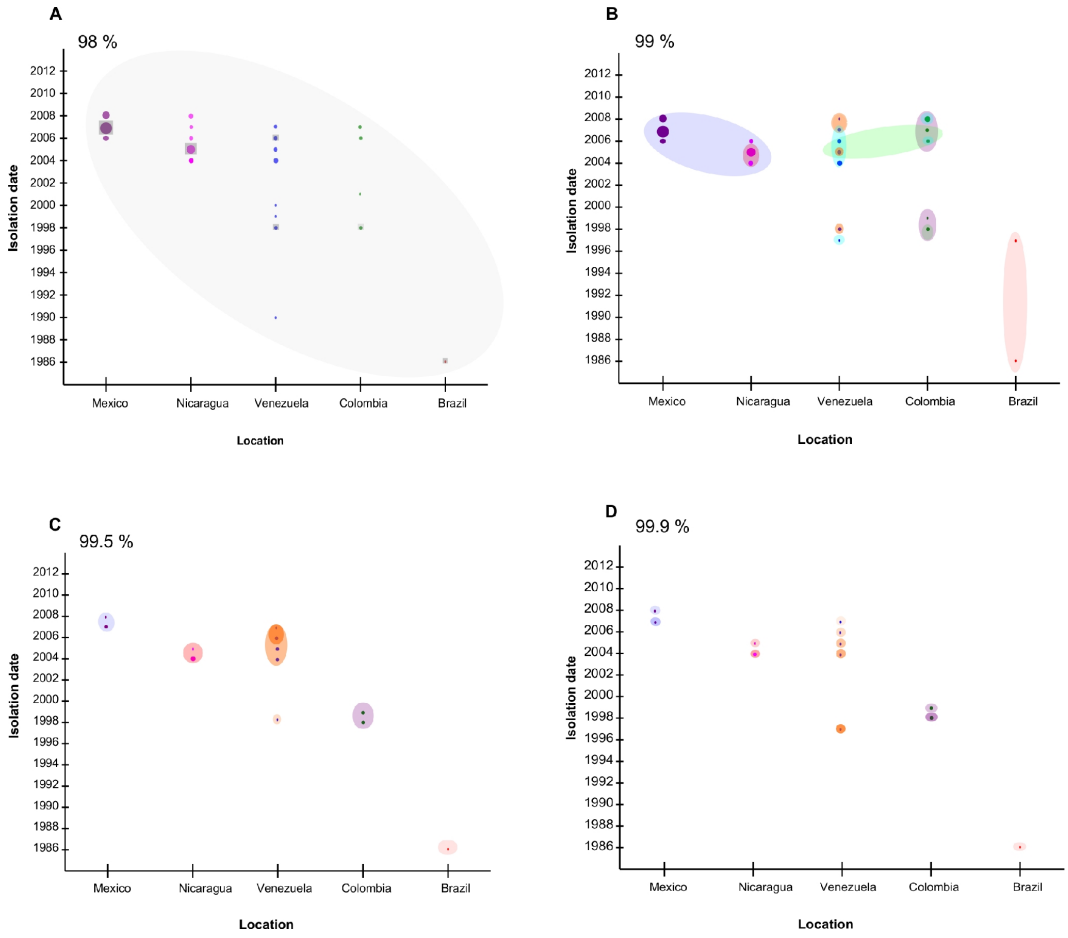
Identity thresholds tested in the clustering analysis. The seeds are the representative sequences recognized under each identity threshold.

Table 2. Alternative topological hypotheses tested using the SH-test

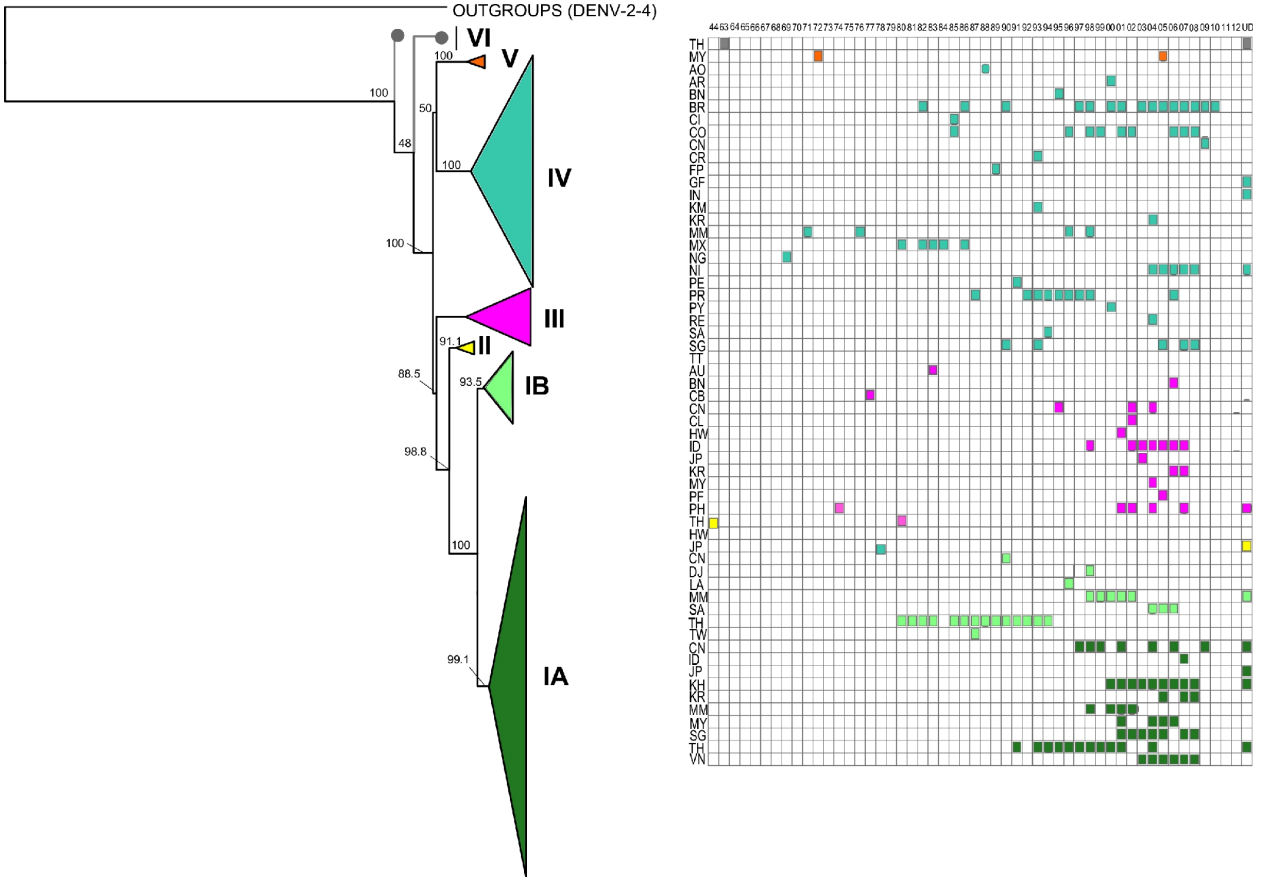
Hypothesis	-ln L	Diff -ln L	P
ML	36289.20955	(best)	—
VI (MONO)	36293.03662	3.82707	0.167
(V,VI (MONO)	54576.07288	18286.86332	0.000*
(IV,(V,VI (MONO)	54415.64761	18126.43806	0.000*
PAR	52338.03171	16048.82216	0.000*
NJ	37026.12215	736.91260	0.000*
IB	36379.24089	90.03134	0.012*

Log likelihoods and difference of log lokelihoods from the Maximum Likelihood reference topology (best) and alternative tree topologies. The monophyletic status of the genotype VI and variations in the placement among genotypes IV, V and VI were tested (remaining topology kept equal). Tree reconstructions recovered under Bayesian inference, Neighbor joining and Parsimony were also contrasted.

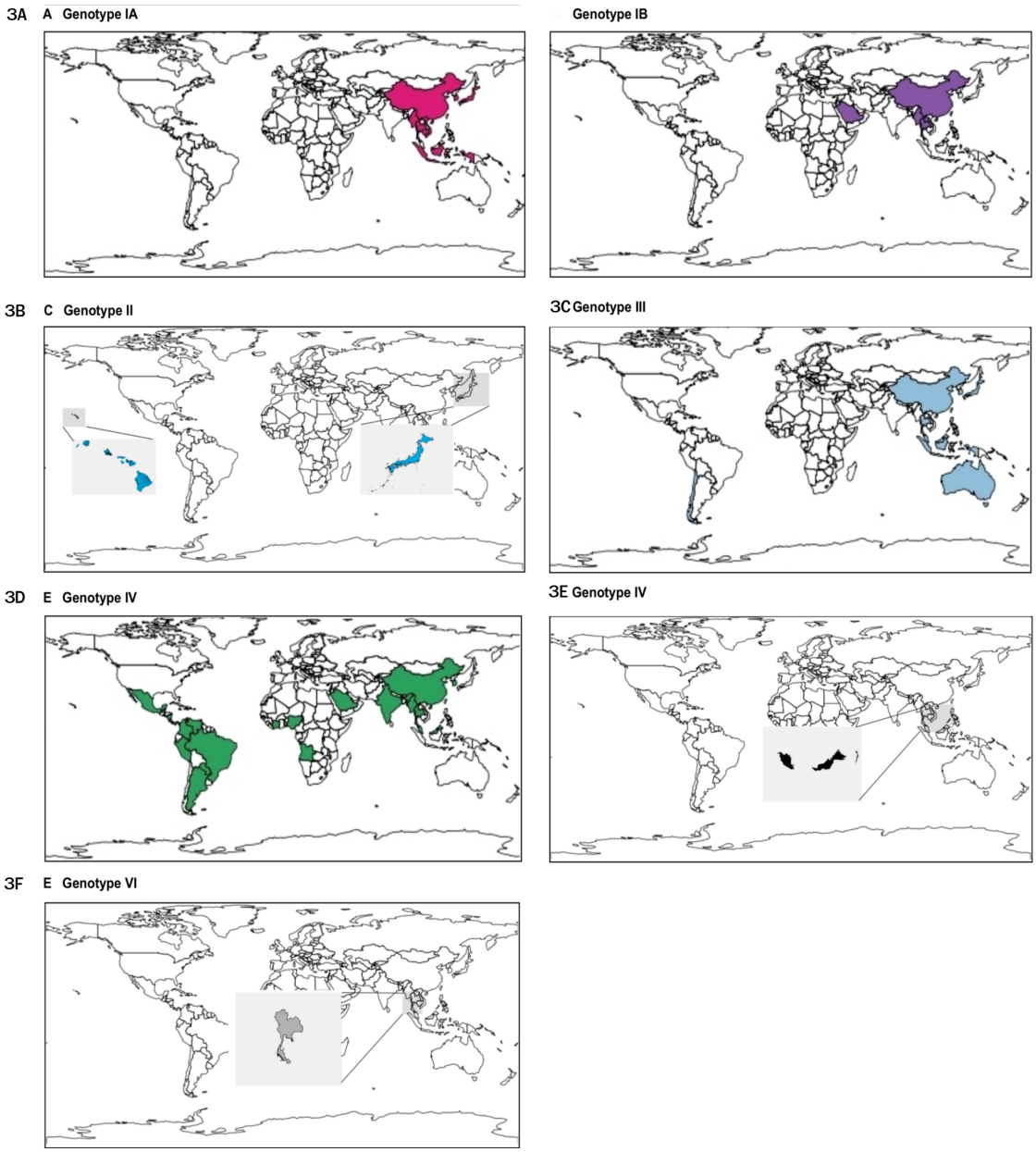
FIGURES



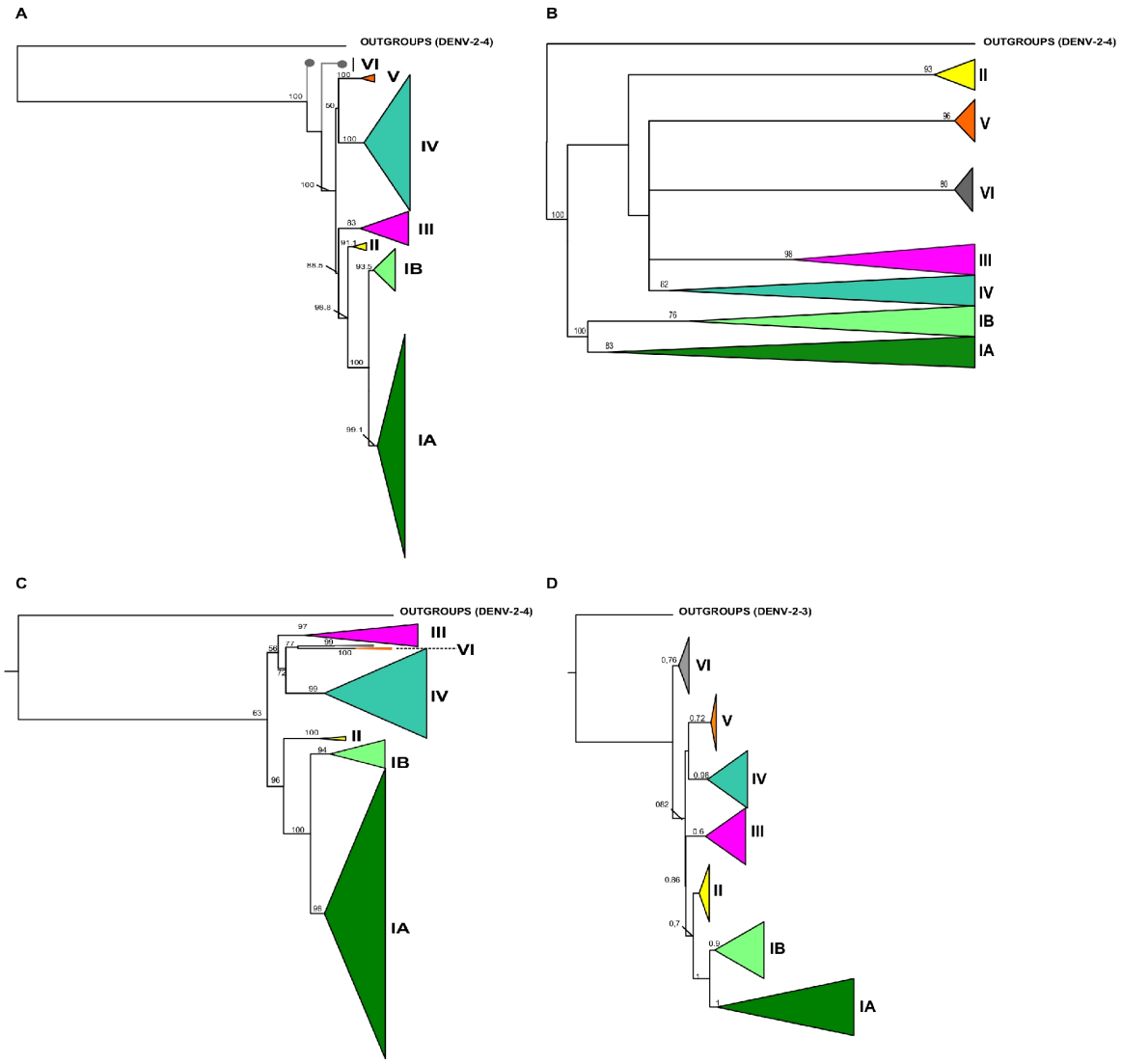
Figures 1. Association among identity thresholds and the spatio-temporal structure in sequences of DENV-1. A. 98% B. 99%, C. 99.5% and D. 99.9%. The groups show patterns of temporal and spatial clustering. Ovals with the same color represent the same cluster. Gray boxes represent sequences evaluated in this example through the different identity thresholds. The size of the circles is proportional to the number of sequences that circulated in the same period or location. Variations in the color shades in C and D represent new clusters.



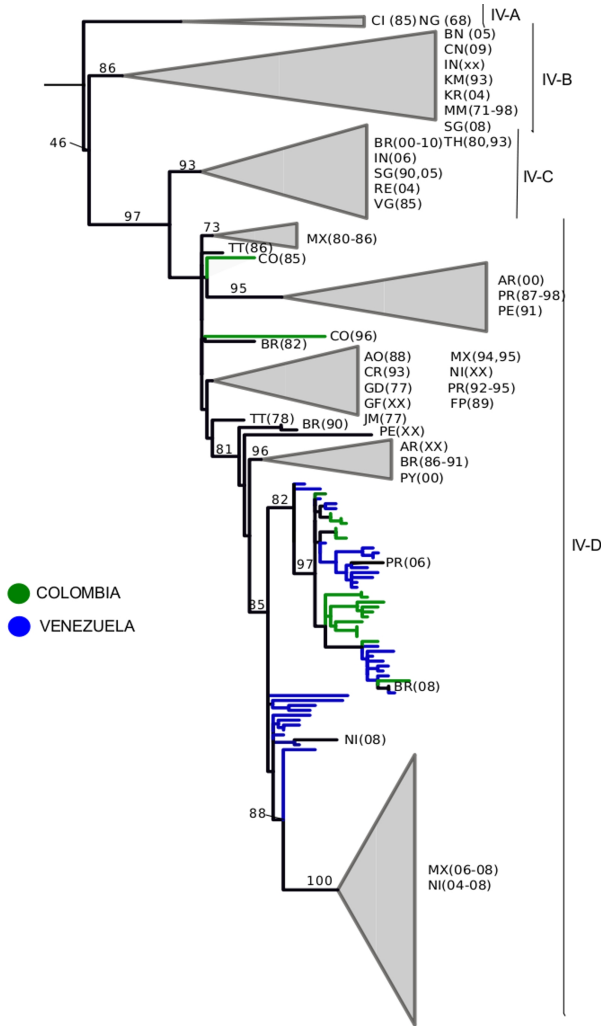
Figures 2. Phylogenetic tree reconstruction under the Maximum Likelihood approach. Genotypes are represented as cartoons in different colors. The sister groups within genotype I are represented in different shades. Isolation dates are shown in colors on the sidebar. Bootstrap confidence values are indicated as percentages above nodes. The tree was rooted using representative strains of DENV2-4.



Figures 3. Geographic distribution of DENV-1 genotypes based on sampling data. 3A. Genotype I, subgroups IA and IB, 3B. Genotype II, 3C. Genotype III, 3D. Genotype IV, 3E. Genotype V, and 3F. Genotype VI.



Figures 4. Phylogenetic relationships patterns among genotypes of DENV-1. A. Maximum Likelihood, B. Parsimony, C. Neighbor joining, D. Bayesian Inference.



Figures 5. Phylogenetic relationships of DENV-1 sequences in genotype IV under Maximum LIkelihood approach. Subgroups are indicated as clades IV-A to IV-D and included strains isolated from Africa, Asian and America. Colombian sequences are shown in green whereas venezuelan strains are shown in blue. Sequences with label XX, represent inderteminated isolation date. Boostrap values are represented above nodes.