THE PHYLOGENETIC RELATIONSHIPS OF DENGUE VIRUS SEROTYPE 4 CIRCULATING IN AN ENDEMIC REGION OF COLOMBIA.

CINTHY LORENA JIMÉNEZ SILVA

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

THE PHYLOGENETIC RELATIONSHIPS OF DENGUE VIRUS SEROTYPE 4 CIRCULATING IN AN ENDEMIC REGION OF COLOMBIA.

CINTHY LORENA JIMÉNEZ SILVA

Trabajo de grado para optar al título de Biológo

Director
Daniel Rafael Miranda-Esquivel
Raquel Elvira Ocazionez Jimenez

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

To my family for their love.

CONTENT

FIGURES LIST

TABLES LIST

Pág.

ABSTRACT

The phylogenetic relationships of Dengue virus serotype 4 circulating in an endemic region of Colombia. *

Cinthy Lorena Jiménez Silva. †

Keywords: Dengue type 4 (DENV-4), Serotype, Genotype.

Dengue viral disease is an emergent global health problem considered a continuing threat for 2.5 billion people at risk of infection. The DENV viruses are grouped within four serotypes. Dengue virus type 4 (DENV-4) has been around for 30 years in Colombia. However, the phylogenetic relationships of DENV-4 in Colombia are unknown. In this study, we present a robust reconstruction of the phylogenetic relationships of DENV-4 using the complete E gene of 9 isolates sampled in a period covering from 2000 to 2005, and all available E gene sequences from the Genbank, which are circulated in Colombia. Through the parsimony criterion, the maximum-likelihood and the bayesian analysis, which reconstructed the phylogenetic trees. The analyses revealed that Colombian strains were grouped in three viral lineages showing different dispersal routes toward Colombia. However, the phylogenetic relationships of Colombian viruses isolated in the period 2000-2005 were represented in third lineages and seem to have come from Venezuela. This study also confirms previous reports showing the circulation of genotype II-subclade 2 of DENV-4 in Colombia. Additionally, We showed the phylogenetic reconstructions based on different short fragments of the E gene allow Dengue virus serotyping and genotyping, but they are inappropriate 363 to recover the phylogenetic relationships.

---

*Final degree project: The phylogenetic relationships of Dengue virus serotype 4 circulating in an endemic region of Colombia.

†Faculty: Ciencias Básicas. Escuela: Biología. Director: Daniel Rafael Miranda-Esquivel
Raquel Elvira Ocazionez Jimenez

RESUMEN

Relaciones Filogenéticas del virus del Dengue serotipo 4 que ha circulado en un area endémica en Colombia. *

Cinthy Lorena Jiménez Silva †

Palabras clave: Dengue tipo 4 (DENV-4), Serotipo, Genotipo.

El virus del Dengue es el agente causal de la Fiebre del Dengue. Se conocen cuatro serotipos del virus (DENV-1, 2, 3 y 4), y cada uno se agrupan en varios genotipos. En Colombia, el virus del dengue es altamente endémico presentando la circulación de los cuatros serotipos. No obstante, se desconocen aspectos relacionados con la filogenia del serotipo 4. El objetivo planteado en este estudio, fue determinar las relaciones filogenéticas del virus del Dengue tipo 4 que circuló en un área endémica en Colombia. Se secuenciarón 9 cepas de Santander y Norte de Santander del gen E, las cuales fuerón aisladas por el CINTROP-UIS. Y junto a las secuencias disponibles en el GenBank se analizaron bajo un algoritmo de agrupamiento y alineamiento multiple, después se reconstruyó las relaciones filogenéticas bajo Parsimonia, Máxima Verosimilitud e Inferencia Bayesiana, adicionalmente se contrastó la información filogenética entre el gen E ye sus particiones. Estos análisis permitierón establecer tres linajes virales del genotipos II han circulado en Colombia, en diferentes períodos de tiempo. Esto sugiere más de una ruta de introduccción del serotipo hacia el país. Por otro lado, Las cepas del Nororiente Colombiano aisladas del 2000 al 2005 han co-circulado con el Caribe y Occidente de Venezuela; por lo tanto, un límite político no es una barrera geográfica. Esta información de relevancia en la vigilancia conjunta entre estas regiones.

---

*Proyecto de Grado: Relaciones Filogenéticas del virus del Dengue serotipo 4 que ha circulado en un area endémica en Colombia.

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

INTRODUCTION

The Dengue virus (DENV) is an infectious agent belonging to the genus *Flavivirus* (family Flaviviridae). It is one of the most important arboviral pathogens in tropical and subtropical regions of the world [17]. The transmission of the dengue virus is given from vectors corresponding to *Aedes sp* mosquitoes [16]. Structurally, this virus is characterized by an icosahedral capsid which is surrounded by a lipoprotein membrane or envelope. Its genome is a single-stranded RNA in positive-sense that encodes a polyprotein organized in three structural proteins and seven nonstructural proteins [3]. The DENV viruses are grouped within four clusters that are antigenically and phylogenetically distinct [20] which are called serotypes (DENV1-4). Likewise, the DENV-4 has been grouped into four genetic types based on the divergence of E gene [13, 24].

The DENV-4 was detected in the Americas for the first time in 1981 during an epidemic in the Puerto Rico. Starting this year, the virus spread rapidly invading neighboring countries and South America. After the event, the DENV-4 has been co-circulating with the other serotypes, with sporadically predominance but usually coinciding with the occurrence of an outbreak or epidemic of the other serotypes [23, 27]. In Colombia, the dengue virus is endemic in the majority of the departments with epidemics every 2-4 years (MPS â INS, 2011). Boschell and collaborators reported the presence of the serotype 4 in 1986 in Colombia [2]. Although the DENV-4 has been around for nearly 30 years, the information documented is limited [2, 15, 27]. In respect to Santander and Norte de Santander [6], has published information from 2006 on findings of DENV-4 in these departments.

In Colombia, the limited knowledge of the characteristics of the virus contributes to the focus on the control of the abundance of the mosquitoes as a prevention measure [27]. However, dengue outbreaks occur in localities with lower rates of mosquito infestation. This is due to hyperendemicity by the co-circulation of the 4 serotypes, the

population growth and the constant displacement of the people between regions [19,36]. This prevention strategy assumes that the lower the mosquito infestation the less the dengue fever; therefore, lower incidence of cases [36].

This study analysed the phylogenetic relationships and the genetic variability of the dengue virus type 4 in Colombia. We determined the complete sequence of the E gene of 9 Colombian strains isolated in 2000 and 2005, and we compared them with available envelope gene sequences retrieved from the GenBank. Subsequently, We evaluated whether the different fragments of the gene could recover the same relationships as the complete gene could. These results contribute to the knowledge of the molecular epidemiology of DENV-4 in Colombia.

MATERIALS AND METHODS

DATASET .

Sample collection.

Our study included 9 Colombian strains that were circulating in Santander/Norte de Santander between 2000 and 2005. The Centro de Investigaciones en Enfermedades Tropicales (CINTROP) and The Public Health Laboratory of the Health Department of Santander collected the samples. In both cases, we counted on the collection permits. We isolated these DENV-4 strains from patients' sera and identified the serotype through direct immunofluorescence with specific monoclonal antibody and RT PCR. [15, 27].

RNA extraction and RT-PCR.

The research team replicated the viral strains in *Aedes albopictus* cell culture (C6/36), in an L15 media supplemented with 10% FBS. Subsequently, the cell monolayers absorbed the virus in order to be infected. Through the column method, we extracted the total RNA using the QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's protocol. After the extraction, we used the SuperScript III reverse transcriptase (Invitrogen) to retrotranscribe the RNA.

PCR amplification and nucleotide sequencing.

We designed the oligonucleotide primers of the complete envelope gene; based on conservative regions from a representative collection of sequences from GenBank. With

these primers, we amplified the cDNA using the PCR method, employed the primers published by [13] and followed the protocol published by [35]. The Macrogen DNA Sequencing Service purified and sequenced the PCR-amplified DNA products. By using the CLC Genomics Workbench 4.5 (CLC Bio, Aarhus, Denmark), we assembled the local strain sequences.

Clustering analysis.

We retrieved the dataset, which contained the Colombian sequences with all the available E gene sequences, parcial E gene sequences and the total genome sequences of serotype 4 collected from the GenBank and published up to July 2011. We discarded mutant, chimeric, recombinant, and clone sequences based on information available in the database. We aligned these sequences preliminarily and selected the E gene region from them. Then, we detected and deleted the identical and chimera remaining sequences from the dataset using UCHIME in USEARCH package [12]. We wanted to build a free redundancy dataset, so that under the previous software and by implementing the clustering algorithm, we analyzed the total sequences with different identity thresholds. We evaluated from 95.0% to 99.9% of the identity thresholds. We selected the most appropriated identity threshold based on the trend of geographic and temporal structure of the clusters. Subsequently, the representatives of these clusters defined the final data base without redundancy.

Alignment and outgroups selection

We aligned the final dataset under the multiple sequence alignment algorithm implemented in MUSCLE v3.8 [11], following the default parameters. On the other hand, we selected the serotype 1,2 and 3 as outgroups. In this case, we excluded other viruses, different from Dengue, due to the generation of long branches in the analysis. Besides, these serotypes allow to root the topologies [26].

THE PHYLOGENETIC ANALYSIS.

The complete E gene:

We reconstructed the phylogenetic relationships with the complete E gene because it is considered one of the most conserved regions in the genome of DENV allowing the study of the spatiotemporal relationships between circulating strains [21]. Furthermore, it is recommended for phylogenetic studies due to the recognition of the viruses' cell host mediated by the envelope protein [25].

Parsimony:

We evaluated three phylogenetic methods (the criteria of Parsimony, Maximum-Likelihood and Bayesian Inference). First, we performed the parsimony analysis applying ratchet search and bisection-reconnection (TBR) by swapping branches of wagner trees. We estimated the bootstrap support based on 1000 replications, which provided the nodal support of the generated topology, and the values were projected over a strict consensus. We inferred this topology using the TNT v1.1 program [14].

Maximum-Likelihood and model selection:

We determined the best-fit nucleotide substitution model using the hierarchical Likelihood tests through the APE package [29] in the R sofware; we also employed Akaike and Bayesian information criteria implemented in Jmodeltest v0.1.1 [30]. In all cases, the general time-reversible of the data was the (GTR) + Γ + I model of base substitution. To determine any bias, we assessed the model by including and excluding the outgroups in the dataset. Based on the model of base substitution resultant, we estimated the maximum-likelihood tree and inference bayesian tree. In the Maximum-Likelihood analysis, we used bio-neighbor joining method to build the starting tree, and we estimated tree topologies with the BEST (NNI+SPR) approach. Standard and rapid non-parametric bootstrap with 1000 pseudoreplicates were used to construct reliable intervals over the phylogeny. We used the PhyML software [18] for it.

Bayesian Inference.

Finally, we implemented the bayesian analysis using the BEAST software [9]. We used Priors by default. We used 7000 generations as sampling frequency and the 25% of the generations were burn-in. We assessed the convergence through the TRACER v1.5 program [32].

Regions of the E gene:

We recovered the phylogenetic relationships using each region of E gene and the possible combinations between the regions; according to the regions of the envelope protein reported by [33]. We used the sequences encoding of the three domains (D1, D2, D3), a transmembrane region (TR) and the possible combinations (Table 2). Then, we analyzed these regions through the Parsimony criterion and the Maximum Likelihood analysis using the same procedure as the one used for the complete E gene. In all cases, the model of the nucleotide substitution was the same as the one we found in the complete E gene nucleotide model.

We estimated the topologies of the 4 regions and all the 10 possible combinations between the regions of E gene. The sectioning of regions took place based on the alignment of the dataset. As a consequence, we obtained 14 topologies and compared them with the topology of the complete E gene (Table 2). we consider that the assumption that the complete E gene provided the most reliable results, and that the regions of the E gene accounted for non-informative subsamples in contrast to the entire gene. We used three comparison measures. Firstly, we calculated the resolution of the topologies with the numbers of nodes using the APE package [29] in R software. Secondly, we compared the similarity between the topologies of the regions in E gene according to the complete E gene topology; this was accomplished through shared nodes using a function that we created in R language. Thirdly, the proportions of nodes missed were of two types: type 1 refered to the nodes that lacked recovering in the regions' topologies of E gene,

but that were present in the complete E gene topology; and type 2 refered to the nodes present in the topologies of the E gene regions that were not shared in the complete E gene topology.

RESULTS AND DISCUSSION

SET DATA

Our the dataset presented strains isolated from 1956 to 2011 in 39 countries around the world. The Colombian E gene sequences of fourteen isolates were 1485 base pairs (bp)(Table 1). We submitted the Santander/Norte de Santander sequences to the GenBank under accession numbers (KC009632-40). The total data were: 411 isolated composed by viruses of the complete E gene (1480-90 bp), partial E gene (218 bp) and the complete genome (10561 bp) for the phylogenetic analyses. We selected the 99.6% identity threshold, which generated unique and informative sample in a non-redundant database showing the structure of geographic and temporal clusters (Figure 1). Given this identity threshold, we obtained a set data with 209 representative sequences (Figure 2). We discarded thresholds lower than the 99.6% due to the loss of geographical localities and years of the isolations. Additionaly, with higher thresholds, the sample of a single locality and year of isolation increased showing the redundancy in the data set.

The clustering analysis based on identity thresholds helped us eliminate the redundant data, reduce the computational efforts in phylogenetic analysis tasks, aid the understanding of the data structure based on the distribution and temporality of the virus isolation, and correct the bias within the viruses dataset.

THE PHYLOGENETIC ANALYSIS.

The complete E gene:

We recovered four genotypes of DENV-4 showing monophyletic groups in agreement with previous researches [4, 13, 22, 24], regardless of the method. The high Bayesian posterior probabilities (0.99), the high non-parametric bootstrap values (70%) and the high boostrap (Parsimony criterion) supported these monophyletic groups 3. However, the general cladogram recovered the same inter-genotypic relationships but different intra-genotype resolutions.

Genotyping.

The topology showed two restricted genotypes and two widely distributed genotypes [4]. The genotype sylvatic was only present in non-human primates corresponding to the only isolates from Malaysia and genotype III reported solely for Thail strains. These genotypes were geographically restricted as reviewed by [4]. In contrast, the genotypes I and II were widely distributed [4]. Isolated strains from Southeast Asia, India, Sri lanka and Japan predominated in genotype I. The results showed two subclades that divided a clade called genotype II. Villabona & Andrade, 2011 reported this subdivision in other studies. This two subclades were: subclade 1 which was represented by isolates from Southeast Asia, the French Polynesia and the Americas since 1977 and up to 2011; and Subclade 2 that corresponded to strains from Southeast Asia since the 2000s and two isolates from indonesia(1973) and China(1978). Therefore, out of the four genotypes, genotype II was the most diverse in country representatives (Figure 3).

DENV-4 circulating in America

Phylogenetic analysis revealed that the genotype II circulated in America from Southeast Asian countries as previously described by [13]. The American strains, isolated from 1981-2011, were segregated in eight different groups clustered according to the year of sampling. This temporal structure showed the oldest isolations were grouped at the basal position of the clades, which were continued by more recent isolates (Figure 3). [1, 13] also described this temporal pattern.

DENV-4 circulating in Colombia.

Our study reaffirmed the circulation of genotype II-subclade 2 in Colombia [7, 28, 34]. The colombian strains segregated into three sub-groupings or viral lineages (COI - III) that corresponded to three temporal clades. These lineages grouped at least three plausible circulation periods of DENV-4 in Colombia corresponding to isolates from 1982, 1996 - 1997, and 2000 - 2005.

We designated the first lineage (CO-I) as the Caribbean island/Caribbean South America from strains isolated since the 80's. This lineage appeared in 1981 isolated from Dominica [13] and we had the first report of DENV-4 in Colombia which was in 1982. Subsequently, circulated in DO/SR/CO/TT for this period of time, and then disappeared. The second lineage (CO-II), northern Central America/northern South America, appeared in 1984 from Mexico and circulated in MX/HN/EC/CO/PE for several years to then vanish in 2000. This lineage showed the introduction to northeastern Colombian in 1997.

The last lineage (CO-III) appeared in 1998 in solely northern South-America (CO/VE/EC/PE/BR and the last report of circulation was in 2011, which showed an introduction to northeastern Colombia in 2000 - 2005. The Santander/Norte de Santander isolates were very closely related to the Caribbean and the western Venezuela's isolates. This lineage showed 3 temporal groups among these regions that corresponded to (1997- 2001 CO/VE) group, (2000- 2005 CO/VE) group and (2000-2011 CO/VE/BR) group (Figure 4).

The data presented here provided strong evidence of the introduction of three different lineages of DENV-4 that have circulated in Colombia over the last 30 years; showing different histories by temporal and geographical clusters. These lineages appear and disappear over time showing lineage exchange by extinction. However, we require a larger sample size to validate these assertions [31]; and we discard that these genetic variations are still in circulation in Colombia. This introduction of different lineages to Colombia would be associated with dispersal routes of DENV-4 since the introduction

of the serotype into the Americas.

The first report of DENV-4 in northeast Colombia was in 1982; this strain is closely related to the Caribbean island and Caribbean South America strains isolated in the early 80's. The time period of this cluster coincided with the introductions of genotype II viruses in America that came from the Lesser Antilles to Caribbean South America. [13, 34]. This initial report represents a dynamic of virus movement based on the invasion phase that corresponded to a period when these viral lineages were absent or during their creation [5]. This reflects how the genotype II has been able to spread rapidly. Subsequently, the analysis showed an increase of new lineages of genotype II of DENV-4. [1, 5]. In contrast, the two new introductions of lineages into northeastern Colombia coincided with the emergence of lineages which were represented by clades that comprised strains from the Antilles, different regions of the American mainland countries or a mixture of both.

The second viral lineage of DENV-4 in northeastern Colombia refers to the appearance of two samples that were reported in 1996 and 1997. This viral lineage appeared with strains from Mexico/Honduras that may date back to 1984. Then it circulated in Central America towards South America and finally, it is possible that the viral lineage disappeared in 2000; according to the last report of isolates from Peru and Ecuador of that year.

The third viral lineage of DENV-4 in northeast Colombia involves viruses from northern South American regions. Our analysis suggested that this spreading took place from the western and the Caribbean of Venezuela to the northeast of Colombia, and a later traffic of strains toward the neighbor countries (Ecuador/Peru and north Brazil). The groups that were formed between Colombia and Venezuela were given in temporal clusters: (1997-2001 CO/VE), (2000-2005 CO/VE), (2000-2011 CO/VE/BR). According to the last report (BR/2011) , it is shown that the third viral lineage is spreading southwards.

This circulation shows that the dispersion of the virus in northeastern Colombia that

circulated in 2000- 2005 is behaving as a region with Venezuela rather than as a locality. Villabona et al., 2009 showed this in the DENV-3. The shared cultural/historical and migrational factors comprised in these neighbouring countries are the most probable reason for constant circulation of the virus between these both countries. This eliminates the notion of a political division as a geographical barrier.

The regions of E gene:

Genotyping.

We obtained fourteen topologies that represented the E gene partitions. Ten of these regions recovered the four genotypes of DENV-4 (Figure 6). In all cases, the D2 and D3 were part of these regions (Figure 6). However, the phylogenetic relationships were different in respect to the complete E gene. Likewise, These ten topologies showed that the northeastern Colombian isolates were grouped with the countries of three previously described viral lineages but, we found unresolved relationships within these clusters.

The comparison measures.

The complete E gene topology presented 138 nodes of reference with Parsimony method and 158 nodes with Maximum Likelihood method, out of the 210 possible (total tips - 1). In the Parsimony criterion and Maximum Likelihood analysis, none of the regions E gene topologies recovered the same number of nodes than the ones in the complete E gene topology; in other words, the partial E gene topology showed a low-resolution compared to the complete E gene topology. The results showed that the resolution of topologies were correlated with the size of regions of the E gene; regardless of the method (Figure 6).

The highest similarity percentage of common nodes in respect to the complete gene topology was 66,2% with the D12TR topology in Parsimony and 70,5% with the D23TR. Besides, these regions were the lowest proportions of nodes missed. Therefore, we con-

sidered them the most informative region of the E gene and the TR topology presented 36,7% in Parsimony and 42,9% in Maximum-Likelihood showing the lowest percentage of nodes in common, in respect to the complete gene. This region was the highest proportion of nodes missed. The proportions of nodes missed ââwere proportional to the similarity value: the lower the similarity the more the proportions of nodes missed of type 1 and 2. All data demonstrated that the least informative region of E gene was the TR (Figure 6).

we recovered up to 70,5 % similarity of nodes of the regions of E gene topologies compared with the complete E gene topology. Therefore, the analysis showed the complete E gene results as the most reliable results compared to gene fragments. This assumption, based on low saturation exposed by the complete gene sequences, sugges better-resolved trees than the partitions of gene [10]. The partial gene topologies are not equal in number to the retrieved and common nodes, in respect to the complete E gene topology, leading to unreliable phylogenetic trees. However, we can recover the serotype in all cases and the 4 genotypes in ten topologies.

This study's results contradict the results reported by [8], as they implemented a methodology for phylogenetic reconstruction based on a short fragment in the carboxyl terminal of the E gene in order to establish phylogenetic relationships. In our analysis, the transmembrane region included the fragment used by [8], which was the least informative region. However , The D2 and D3 regions, or the possible combinations that include them, were informatives insomuch as we could retrieve the four genotypes. This could be due to the fact that the domain II and III presented most highly conserved residues in contrast with other domains [33]. Although along the protein it had a similar pattern of conservation, the sequences of these two regions are much larger than the TR region, therefore these two regions presented more conserved residues. Consequently, these more conserved regions may provide valuable phylogenetic information but still containing low resolution (polytomies). In contrast to the highly variable regions that can produce high resolution (fewer polytomies) for trees, but do not reflect the same topology of the complete E gene.

CONCLUSIONS

In summary, we suggest that the phylogenetic analysis provided evidence for the three viral lineages of Genotype II of DENV-4 within Colombia, in different periods of time; this suggests more than one route of movement through the continent. The Northeastern Colombian strains circulating between 2000 and 2005 are very closely related to the west and the Caribbean of Venezuela, which shows circulation among these regions; therefore, a political division is not a geographic barrier. This study revealed that the fragments of the E gene recover different phylogenetic relationships than the complete E gene. The findings suggest that the small regions of E gene grouped the genotypes, but without providing monophyletic relationships. Thus, the fragments allow Dengue virus serotyping and genotyping, but they are inappropriate to recover the phylogenetic relationships.

REFERENCES

1. Bennett, Shannon N. and Holmes, Edward C. and Chirivella, Maritza and Rodriguez, Dania M. and Beltran, Manuela and Vorndam, Vance and Gubler, Duane J. and McMillan, W. Owen Selection-driven evolution of emergent dengue virus. *Molecular Biology and Evolution*, 20(10):1650–1658, 2003.

2. Boshell J, Groot H, Gacharna M, Márquez G, González M, Gaitán M. Dengue en colombia. *Biomédica.*, 6:101-102., 1986.

3. Chambers, Thomas J. and Hahn, Chang S. and Galler, Ricardo and Rice, Charles M. Flavivirus genome organization, expression, and replication. *Annual Review of Microbiology*, 44(1):649–688, 1990. PMID: 2174669.

4. Chen, Rubing and Vasilakis, Nikos Dengue â quo tu et quo vadis? *Viruses*, 3(9):1562–1608, 2011.

5. Carrington, Christine V. F. Foster, Jerome E. Pybus, Oliver G. Bennett, Shannon N. and Holmes, Edward C. Invasion and maintenance of dengue virus type 2 and type 4 in the americas. *JOURNAL OF VIROLOGY*, 79(23):14680–14687, 2005.

6. Cortés, F. M., Gómez, S. Y., Ocazionez, R.E. Subtipos de virus dengue serotipos 2, 3 y 4 aislados en el departamento de santander colombia. *Revista Cubana de Medicina Tropical.*, 57(3):5, 2007.

7. Díaz, Francisco J. and Black, William C. and Farfán, José A. and Loroño-Pino, María A. and Olson, Kenneth E. and Beaty, Barry J. Dengue virus circulation and evolution in mexico: A phylogenetic perspective. *Archives of Medical Research*, 37(6):760 – 773, 2006.

8. Domingo C, Palacios P, Jabado P, Reyes N, Niedrig M, Gasco J, Cabrerizo M,

Lipkin W, Tenorio A, 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(4):1519–1529, 2006.

9. Drummond A J & Rambaut A. Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology.*, 7:214, 2007.

10. DuchÃªne, Sebastián AND Archer, Frederick I. AND Vilstrup, Julia AND Caballero, Susana AND Morin, Phillip A. Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS ONE*, 6(11):e27138, 11 2011.

11. Edgar, Robert Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5(1):113, 2004.

12. Edgar, Robert C. Search and clustering orders of magnitude faster than blast. *Bioinformatics.*, 2010.

13. Foster, Jerome E. Bennett, Shannon N. Vaughan, Helen. Vorndam, Vance. McMillan, W.Owen and Carrington, Christine V.F Molecular evolution and phylogeny of dengue type 4 virus in the caribbean. *Virology*, 306(1):126 – 134, 2003.

14. Goloboff, Pablo A. and Farris, James S. and Nixon, Kevin C. Tnt, a free program for phylogenetic analysis. *Cladistics*, 24(5):774–786, 2008.

15. Gómez, Sergio Yebrail Rangel. Arenas, Christian Julián Villabona. Pimiento, Flor Angela Torres. Esquivel, Daniel Rafael Miranda. Jimenez, Raquel Elvira Ocazionez Dengue virus serotype 3 (genotype iii) from colombia: A perspective of its pathogenic potential. *Dengue Bulletin*, 32:126–37, 2008.

16. Gubler, Duane J. Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews*, 11(3):480–496, 1998.

17. Gubler, Duane J. Reiter, Paul. Reiter, Eduard. Sanders, J. and Vorndam, AVance . Dengue and dengue haemorrhagic fever. *The Lancet*, 352(9132):971 – 977, 1998.

18. Guindon, Stéphane and Gascuel, Olivier A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5):696–704, 2003.

19. Holmes EC. The evolutionary biology of dengue virus. *Novartis foundation Symposium*, 277:177–187, 2006.

20. Holmes Edward C. and Twiddy S.Susanna The origin, emergence and evolutionary genetics of dengue virus. *Infection, Genetics and Evolution*, 3(1):19 – 28, 2003.

21. Klungthong, C and Putnak, R and Mammen, M P and Li, T and Zhang, C Molecular genotyping of dengue viruses by phylogenetic analysis of the sequences of individual genes. *Journal of virological methods*, 154(1-2):175–81, December 2008.

22. Klungthong, Chonticha Zhang, Chunlin Mammen P. Mammen Jr. Ubol, Sukathida and Holmes, Edward C. The molecular epidemiology of dengue virus serotype 4 in bangkok, thailand. *Virology*, 329(1):168 – 179, 2004.

23. Kouri GP, Guzman MG, Bravo JR, Triana C. Dengue haemorrhagic fever/dengue shock syndrome: lessons learned from the cuban epidemic, 1981. *Bulletin of the World Health Organization.*, 67:375â8, 1989.

24. Lanciotti, R S and Gubler, D J and Trent, D W Molecular evolution and phylogeny of dengue-4 viruses. *Journal of General Virology*, 78(9):2279–84, 1997.

25. Modis, Yorgo Ogata, Steven Clements David & Harrison Stephen C. Structure of the dengue virus envelope protein after membrane fusion. *nature*, 427:313–319, 2004.

26. Nixon, Kevin C. and Carpenter, James M. On outgroups. *Cladistics*, 9(4):413–426, 1993.

27. Ocazionez, Raquel Elvira AND CortÃ, FabiÃ¡n Mauricio AND Villar, Luis Angel AND GÃ, Sergio Yebrail Temporal distribution of dengue virus serotypes in Colombian endemic area and dengue incidence: re-introduction of dengue-

3 associated to mild febrile illness and primary infection. *MemÃdo Instituto Oswaldo Cruz*, 101:725 – 731, 11 2006.

28. Ocazionez, Raquel E. Gómez , Sergio. Cortés, Fabián M. Dengue hemorrhagic fever serotype and infection pattern in acolombian endemic area. *Revista de Salud Publica*, 9:262 – 274, 06 2007.

29. Paradis E. and Claude J. and Strimmer K. Ape: analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20:289–290, 2004.

30. Posada, David jmodeltest: Phylogenetic model averaging. *Molecular Biology and Evolution*, 25(7):1253–1256, 2008.

31. Pybus, O. G. and Rambaut, A. and Holmes, E. C. and Harvey, P. H. New inferences from tree shape: Numbers of missing taxa and population growth rates. *Systematic Biology*, 51(6):881–888, 2002.

32. Rambaut A. Tracer v1.4., 2007.

33. Rey, Felix A. Heinz, Franz Mandl, Christian Kunz, Christian Harrison, Stephen C. The envelope glycoprotein from tick-borne encephalitis virus at 2. *Nature*, 375:291 – 298, 1995.

34. Villabona-Arenas, C. J. de Andrade Zanotto, P. M. Evolutionary history of dengue virus type 4: Insights into genotype phylodynamics. *Infection, Genetics and Evolution*, 11(5):878 – 885, 2011.

35. Villabona-Arenas, C. J. de Andrade Zanotto, P. M. Phylogeny of dengue virus type 3 circulating in colombia between 2001 and 2007. *Tropical Medicine & International Health*, 14(10):1241–1250, 2009.

36. Weaver, S. C. Vasilakis, N. Molecular evolution of dengue viruses: Contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. *Infection, Genetics and Evolution*, 9(4):523 – 540, 2009.

TABLES

Table 1. The Colombian E gene sequences of the gene analysis. The Colombian E gene sequences of fourteen isolates were 1485 base pairs. CO: Colombia

Table 2. The statistical table of the complete E gene and the regions of the gene analysis. The sectioning of regions of E gene represented by the D1, D2, D3, a transmembrane region (TR) and the possible combinations obtaining fourteen topologies. M: Four monophyletic groups. I: Genotype I, II: Genotype II, III: Genotype III and S: Genotype Selvatic. *: Genotype present. NODES: Number of nodes. To: the similarity between the topologies of the regions in E gene according to the complete E gene topology. ERR1: the proportions of nodes missed type 1. ERR2: the proportions of nodes missed type 2. In this section the percentages were according with the 210 possible nodes (total tips - 1).

Country	Locality	Year
Co	--	1982
Co	Santander	1996
Co	--	1997
Co	Santander	2000
Co	Santander	2001
Co	Santander	2001
Co	Santander	2001
Co	Santander	2004
Co	Santander	2004
Co	Norte de Santander.	2005
Co	Norte de Santander.	2005

Table 1. The Colombian E gene sequences of the gene analysis.

THE COMPLETE E GENE	SIZE OF SEQ		M	I	II	III	S	PARSIMONY								MAXIMUM LIKELIHOOD							
								#NODES		T0		ERR1		ERR2		#NODES		T0		ERR1		ERR2	
	1485	100,0	*	*	*	*	*	138	65,7	138	65,7	0	0,0	0	0,0	158	75,2	158	75,2	0	0,0	0	0,0
D1	375	25,3					*	65	31,0	20	9,5	118	56,2	45	21,4	97	46,2	54	25,7	61	29,0	43	20,5
D2	525	35,4	*	*	*	*	*	**95**	45,2	**55**	26,2	***83***	39,5	***40***	19,0	109	51,9	**83**	39,5	***49***	23,3	***26***	12,4
D3	300	20,2	*	*	*	*	*	***63***	30,0	24	11,4	114	54,3	39	18,6	90	42,9	58	27,6	68	32,4	32	15,2
TR	288	19,4						77	36,7	***5***	2,4	**133**	63,3	**72**	34,3	**90**	42,9	***51***	24,3	**68**	32,4	**39**	18,6
D12	900	60,6	*	*	*	*	*	112	53,3	24	11,4	114	54,3	88	41,9	134	63,8	92	43,8	24	11,4	42	20,0
D13	675	45,5	*	*	*	*	*	94	44,8	32	15,2	106	50,5	62	29,5	125	59,5	87	41,4	33	15,7	38	18,1
D1TR	663	44,6						117	55,7	30	14,3	108	51,4	87	41,4	126	60,0	89	42,4	32	15,2	37	17,6
D23	825	55,6	*	*	*	*	*	104	49,5	60	28,6	78	37,1	44	21,0	133	63,3	105	50,0	25	11,9	28	13,3
D2TR	813	54,7	*	*	*	*	*	109	51,9	***19***	9,0	***119***	56,7	90	42,9	132	62,9	113	53,8	26	12,4	19	9,0
D3TR	588	39,6					*	***87***	41,4	20	9,5	118	56,2	67	31,9	123	58,6	**77**	36,7	**35**	16,7	**46**	21,9
D123	1200	80,8	*	*	*	*	*	125	59,5	73	34,8	65	31,0	52	24,8	**147**	70,0	**130**	61,9	11	5,2	***17***	8,1
D12TR	1188	80,0	*	*	*	*	*	**139**	66,2	30	14,3	108	51,4	**109**	51,9	147	70,0	112	53,3	11	5,2	35	16,7
D23TR	1113	74,9	*	*	*	*	*	131	62,4	**99**	47,1	**39**	18,6	***32***	15,2	148	70,5	118	56,2	***10***	4,8	30	14,3
D13TR	963	64,8	*	*	*	*	*	118	56,2	37	17,6	101	48,1	81	38,6	139	66,2	100	47,6	19	9,0	39	18,6
		%							%		%		%		%		%		%		%		%

Table 2. Complete E gene and the regions of the gene.

FIGURES

Figure 1. Association among identity thresholds and the spatio-temporal structure in sequences of DENV-4. A. 99.2% B. 99.6%, C. 99.7% and D. 99.9%. The cones represent sequences in the date and location of isolation: STRAINS. The size of the cones is proportional to the number of sequences that circulated in the same period or location. the shaded regions represent the same cluster. These groups show patterns of temporal and spatial clustering. Areas with more cones represent samples more informative.

Figure 2. Identity thresholds VS numbers of seeds. 95.0% to 99.9% represent the identity thresholds that we evaluated with the number of representative sequences (seeds) . We selected the 99.6% identity threshold obtaining a set data with 209 representative sequences.

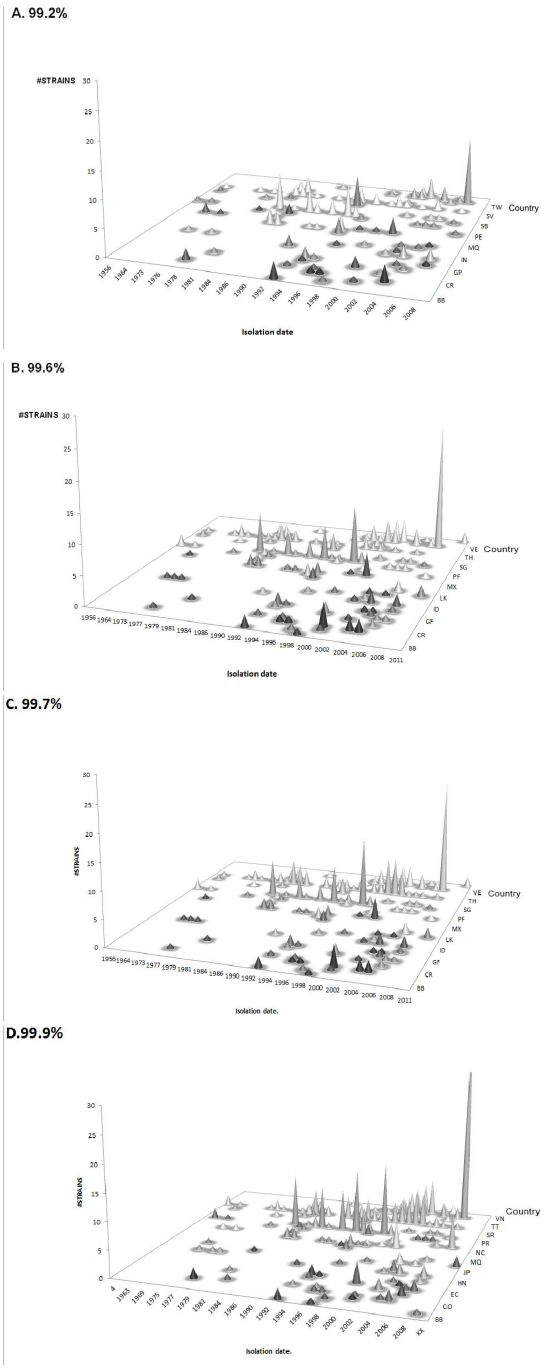
Figure 3.Phylogenetic tree reconstruction of four genotypes recuperated with the Parsimony (PP), Maximum Likelihood (ML) and Bayesian Inference (BB) analyses. Genotypes are represented as cartoons in different colors. The tree was rooted using representative strains of DENV1-3

Figure 4. Phylogenetic relationships of Dengue virus type 4 under Maximum Likelihood approach. Boostrap confidence values are indicated as percentages above nodes. The tree was rooted using representative strains of DENV1-3. The Colombian lineages represented as CO I-III and the American Lineages by I-VIII. each clade represented as cartoons a showing the date and country of isolation.

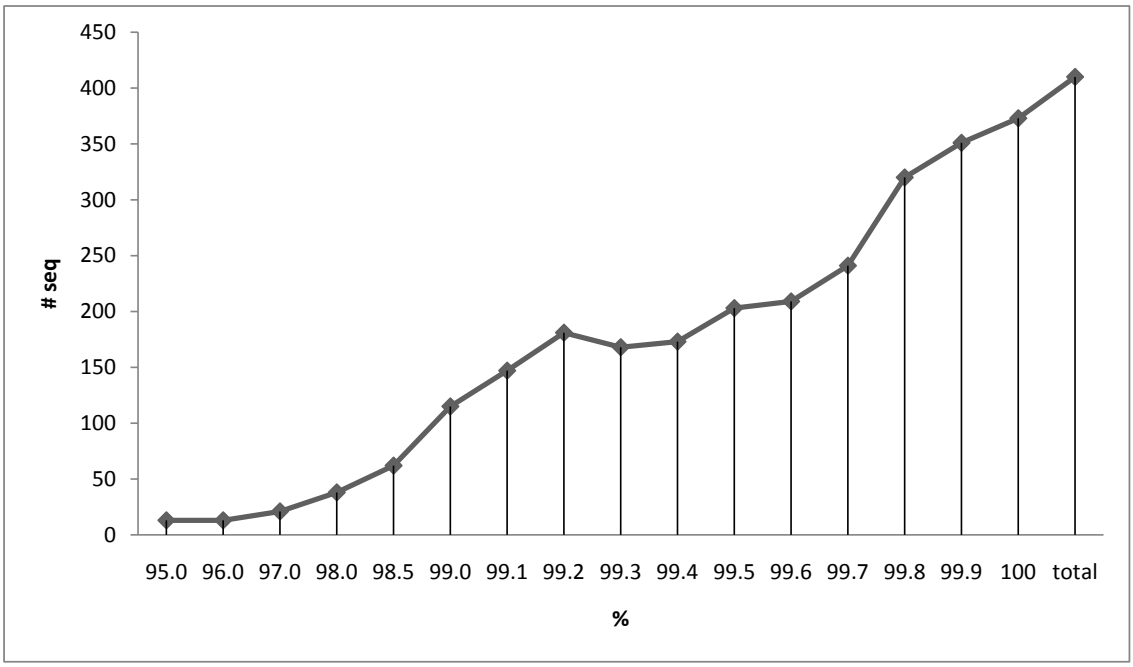
Figure 5. Lineages in Colombia The Colombian lineages represented as CO I-

III and the American Lineages by I-VIII. each clade represented as cartoons a showing the date and country of isolation. Each lineach showed the country and year of isolation.

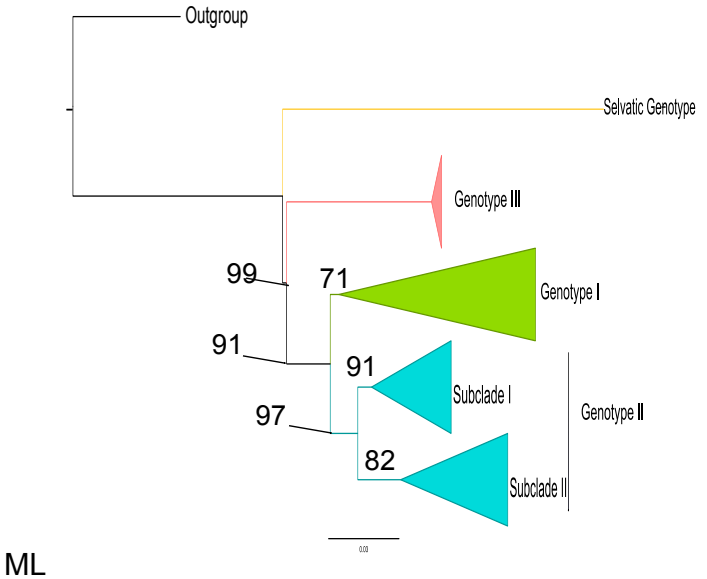
Figure 6. Resolution of topologies Vs size of the sequences in the gene's regions: pp(Parsimony), ML(Maximum-Likelihood). Correlation with the size of regions of the E gene; regardless of the method. The Black represent the regions of the E gene that recover the 4 genotypes.



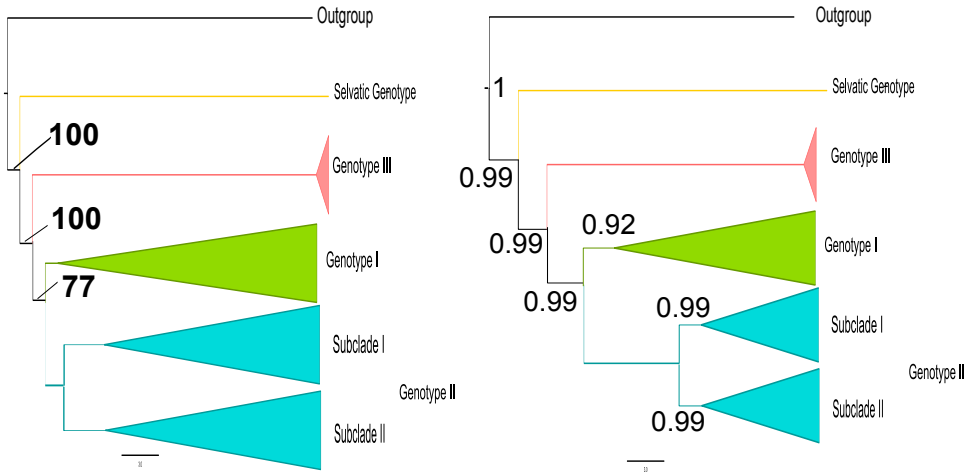
Figures 1. Identity thresholds evaluated using Uclust



Figures 2. Identity thresholds VS numbers of seeds.



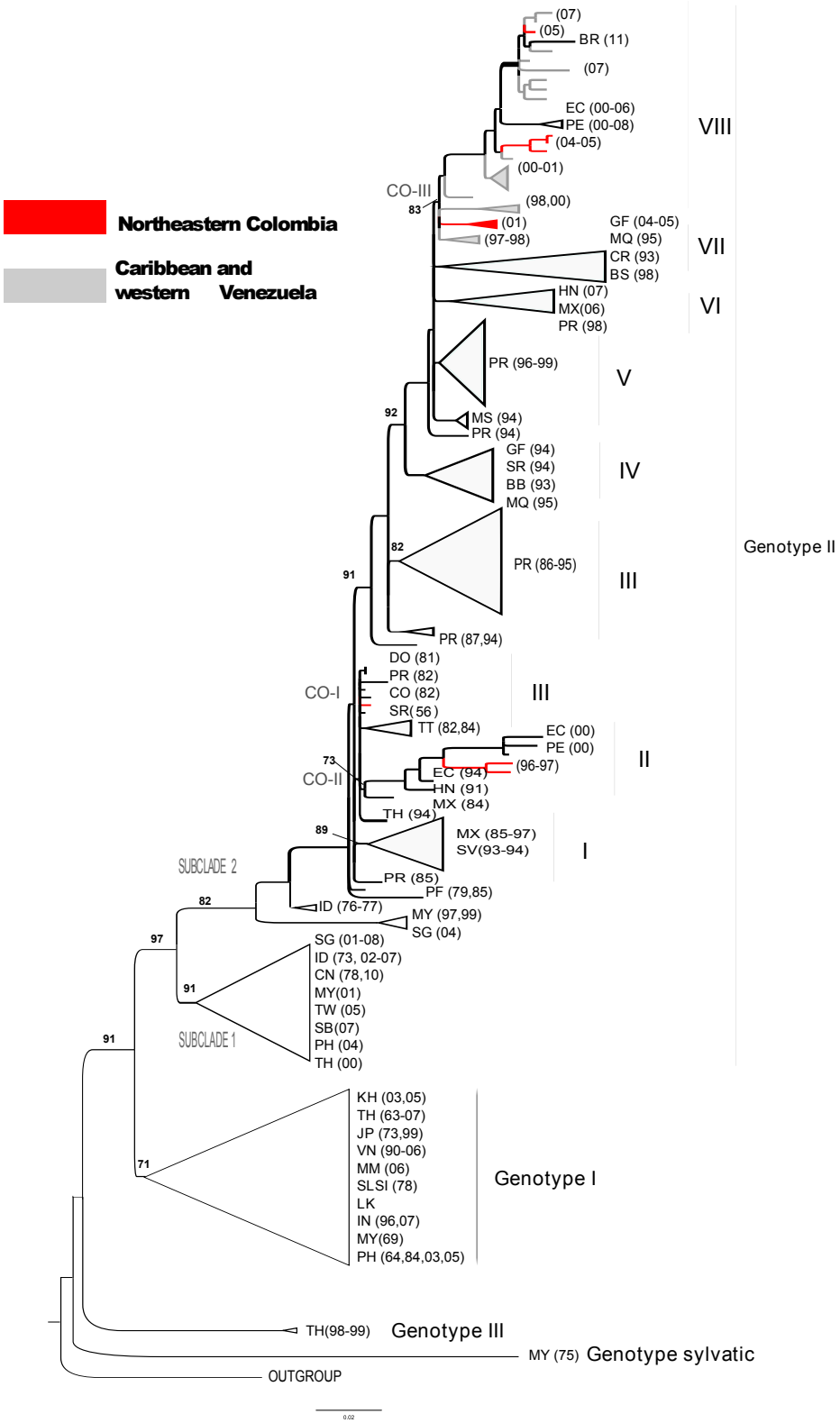
ML



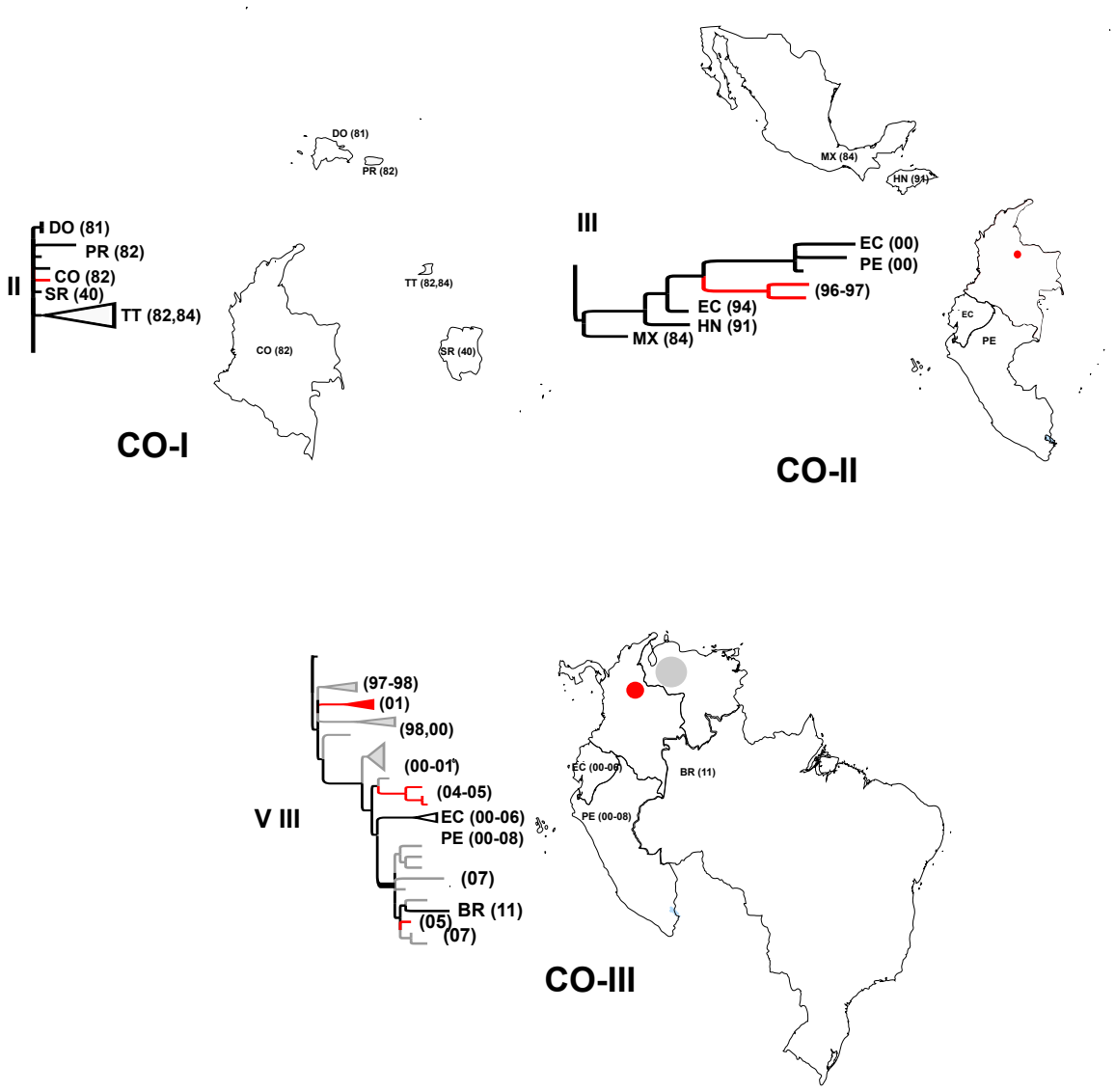
P

BB

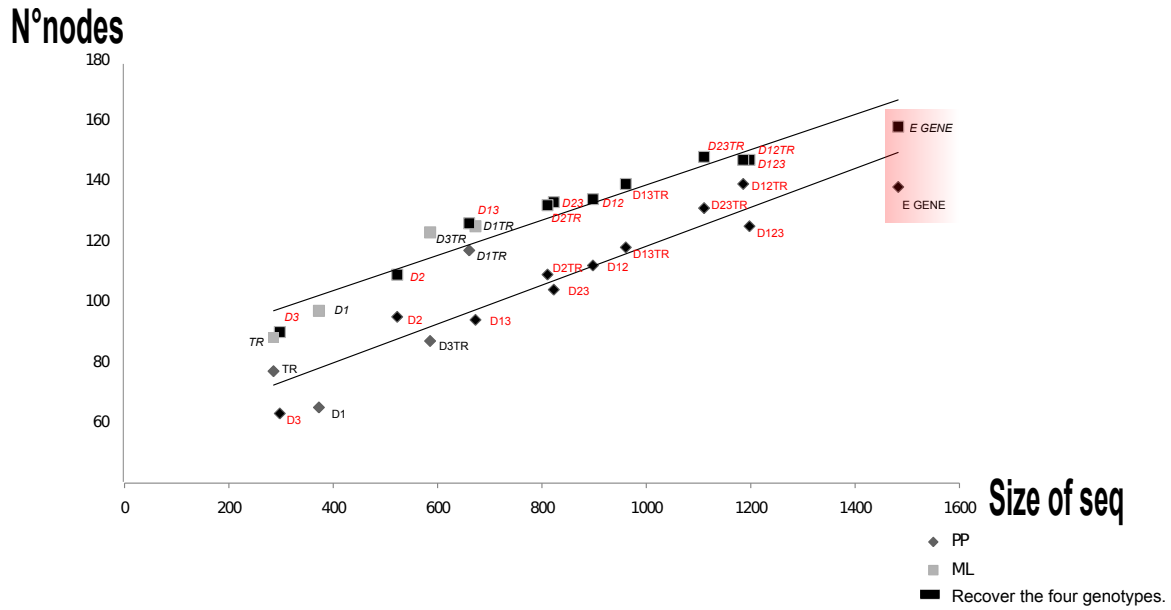
Figures 3. Four genotypes recuperated with the Parsimony (PP), Maximum Likelihood (ML) and Bayesian Inference (BB) analyses.



Figures 4. Phylogenetic relationships of Dengue virus type 4.



Figures 5. Lineages in Colombia.



Figures 6. Resolution of topologies Vs size of the sequences in the gene's regions: pp(Parsimony), ML(Maximum-Likelihood).