CHAPTER 4

Phylogenetic Analysis of
Whole Genome
&
Individual Genes
PHYLOGENETIC ANALYSES OF FULL GENOME AND INDIVIDUAL GENE

Full genome sequencing and analysis are beneficial for surveys of viral biodiversity and molecular epidemiology. Full genome sequence analysis of DENV serotypes was undertaken to examine the variation in different genes and their contribution to the evolution of dengue. The downside of the full genome analysis in DENV was the scarcity of full genome sequences and lack of data for sylvatic isolates in Genbank. The nt sequence diversity analysis of full genome sequences and variation in aa sequence for each individual gene was done.

The phylogenetic trees were created for each individual gene to determine the effect of individual gene sequence variation on the genotyping DENV-1, 2, 3 and 4. For DENV-1 and 2 isolates were selected from each decade since 1960’s. There were no representative DENV-3 isolates for the few decades and those present in the repository could not be revived, therefore only one isolates from 1966 and two from 2005 (from different places) were sequenced and used for the analysis. In case of DENV-4 isolates of three decades, 1960’s, 1970’s and 2000 were used for sequencing and analysis. There were no DENV-4 outbreaks reported in India during the intervening decades.

DENV-1

The complete genome sequences of four Indian DENV-1 isolates (Table 6) were analyzed with 36 publicly available global isolates representing all genotypes. All genotypes established by E gene phylogenetic analysis in chapter 3 were reflected in the eleven trees derived from each gene and the complete genome. The Indian isolates fell into three lineages (India I, III, and Indo-Africa) of the AMAF genotype. India II lineage was not represented due to lack of complete genome sequences of the relevant isolates (Figure 26). Topology of the phylogenetic trees with complete genome and NS4A corroborated well with the E gene tree with good bootstrap values at major nodes. Trees observed with the sequences of the other genes differed in the position of isolates, NIV_715393_1973 and AF298807_Abidjan, of the Afro Indian lineage.
In the trees generated by NS1 and NS2A isolate NIV_715393_1971 was found to be at the root of AMAF genotype with good bootstrap values. In NS2B and NS5 trees the two isolates clustered closely but only the outer major node was supported by good bootstrap values. Very low bootstrap supports were observed in the C gene tree at all nodes wherein the Abidjan isolate clustered with India I lineage and NIV_715393_1971 isolate rooted at the base of America and India III lineages. Total change in the topology was observed with prM, NS3 and NS4B trees with Abidjan isolate at the base of all genotypes. However this tree was supported by very low bootstrap values and therefore results were considered not significant.
The complete genome sequences of five Indian DENV-2 isolates (Table 6) were analyzed with 40 publicly available global isolates representing all genotypes. All genotypes established by E gene phylogenetic analysis in chapter 3 were reflected in the eleven trees derived from each gene and the complete genome. The Indian Isolates NIV_P23085_1960, NIV_715394_1971, and NIV_803347_1980 fell into American genotypes and NIV_969201_1996 and NIV_053598_2005 into Cosmopolitan genotypes (Figure 27).

![ML trees derived from complete genome and 10 coding genes of DENV-2](image)

**Figure 27**: ML trees derived from complete genome and 10 coding genes of DENV-2

- **E/WG/NS1/2A/2B/3**
- **NS4A**
- **NS4B/5**

Topology of all phylogenetic trees except NS4A, NS4B and NS5 corroborated well with the E gene tree with good bootstrap values at major nodes. NS4A, NS4B and NS5 trees differed from other trees gene trees in the position of isolates in the American genotype however bootstrap values were not very good. In NS4A tree isolate
NIV_803347_1980 found at the root of American genotype and NIV_P23085_1960 shifted from American genotype to Cosmopolitan genotype with low bootstrap support at all nodes of Cosmopolitan genotype. In NS4B and NS5 gene trees isolate NIV_803347_1980 found to be at the root of American genotype with low bootstrap support at all internal nodes.

**DENV-3**

The complete genome sequences of five Indian DENV-3 isolates (Table 6) were analyzed with 29 publicly available global isolates representing all genotypes. All genotypes established by E gene phylogenetic analysis in chapter 3 were reflected in the eleven trees derived from each gene and the complete genome. The Indian isolates fell into two lineages, C and F of the GIII. Lineage B and E were not represented due to lack of complete genome sequences of representative isolates (Figure 28).

![Phylogenetic trees of DENV-3](image)

**Figure 28:** ML trees derived from complete genome and 10 coding genes of DENV-3

Topology of the phylogenetic trees with complete genome and C, NS1, NS3, NS4A, NS4B and NS5 corroborated well with the E gene tree with good bootstrap values at major nodes. The topology of NS2A and 2B gene trees was similar were supported by different bootstrap values. The NS2A tree was supported by good bootstrap values than NS2B tree. In the prM gene tree change in the topology observed because lineage D shifted from GIII.
to GIV. However this tree was supported by very low bootstrap values and therefore results were considered not significant.

**DENV-4**

The complete genome sequences of three Indian DENV-4 isolates (Table 6) were analyzed with nine publicly available global isolates representing all genotypes. All genotypes established by E gene phylogenetic analysis in chapter 3 were reflected in the eleven trees derived from each gene and the complete genome.

The Indian isolates fell into two genotypes GI (lineage C) and GV. Topology of all the phylogenetic trees differed from the E gene tree in the position of isolate, NIV_793679_1979. In the E gene tree the isolate clustered with isolate NIV_624000_1962 of GV whilst in the other trees it clustered with GI (Figure 29). Both the topologies were supported by good bootstrap values. This genotype shift may be due to recombination breakpoint present in the E gene of the isolate. The 1960’s isolate of GV probably contributed a larger chunk of E gene. On the other hand the rest of the genome was evidently contributed by the Sri Lankan isolate of GI.
Figure 29: ML trees derived from complete genome and 10 coding genes of DENV-4

All trees differed in the position of GV. Topology of the trees generated from C, NS1, NS2A, NS3 and NS5 were similar with GV at the root of GI. The NS1, NS3 and NS5 trees were supported by good bootstrap values whereas C and NS2A trees had low bootstrap. Very low bootstrap supports were observed in NS2B and 4A gene trees at all nodes wherein the GV rooted at the base of GI/GIII and GI/GII, therefore results were considered not significant. Total change in the topology was observed with prM and NS4B gene trees which also did not have good support value.
DISCUSSION

Genetic diversity plays an important role in DENV evolution which has major impact on the virulence and dissemination of viruses. To determine the effect of gene variation on DENV evolution phylogenetic analysis of each individual gene and WG sequences was carried out. Indian and global isolates which were geographically divergent and represented all known genotypes for all serotypes were included in the analysis.

For DENV-1 the trees derived from the complete genome and each individual gene were found to be different due to the position of isolates belonging to the Afro India lineage. This may have been due to the transient presence of the Afro India lineage in India, which had been imported from Africa as shown in the E gene analysis in Chapter 3. Low bootstrap supports were observed when isolate NIV_715393_1971 clustered with lineage, India III (C, Prm, NS3, NS4B gene trees) as compared to the bootstrap values observed when the isolate clustered with India I lineage, revealed that Afro India lineage isolates were closer to currently circulating isolates than the extinct lineage.

In all the trees isolate NIV_631288_1963 representing India III lineage was found to be at the base of the American lineage supported by good bootstrap values. This confirmed the finding that virus was exported from India to America as shown earlier by E gene molecular clock analysis in Chapter 3. Based on bootstrap values it can be conclude that E and NS4A genes were suitable for the phylogenetic analysis but not C, prM, NS3 and NS4B genes.

In DENV-2 isolate export of virus from India to America is supported by the good bootstrap values observed whenever the isolate, NIV_P23085_1960 rooted the American genotype (C, prM, NS1, NS2B, NS3) as observed in molecular clock analysis using E gene. Origin of the DENV-2 observed in E gene molecular clock was American genotype which is also observed in the tree generated by the NS5 gene and supported by the 99% bootstrap values but in the same tree NIV_p23085_1960 at the root of all genotypes rather at the root of American genotype supported by very low bootstrap values 47%. Analysis also revealed that C, prM, E, NS1, NS2B and NS3 can be used for genotyping of DENV-2.

In DENV-3 it was observed that lineage F rooted the Gill with good bootstrap support indicating that the Gill evolved from lineage F consisting only old Indian isolates as observed in the E gene analysis. Analysis also revealed that all genes except prM, NS2A and NS2B genes can be used for the DENV-3 genotyping.
In DENV-4 good bootstrap values observed with the GV at the base of GI and low values with GV at the root of GII and GIII indicating that GV is the origin of only GI and not GII and GIII which was similar to the results observed E gene molecular clock analysis. Analysis also supported that the C, E, NS1, NS2A, NS3 and NS5 and be used for DENV-4 genotyping.