Journal of Bio-Molecule Research and Engineering
Vol. 1 | No. 2 | 40 – 47 | 2022
e-ISSN: 2962-343X
http://www.e-journal.unair.ac.id/biome

Phylogenetic Analysis of DENV-1 Isolated in Surabaya, Indonesia

Fistara Lesti Rahmafitria[1], Dzikra Nasyaya Mahfudhah[1], Teguh Hari Sucipto[2*], Nuniek Herdyastuti[1], Mirwa Adiprahara Anggarani[1], Shifa Fauziyah[3], Mamik Damayanti[4]

[1]Department of Chemistry, Universitas Negeri Surabaya, Jl. Ketintang Surabaya 60231, Indonesia
[2]Dengue Study Group, Institute of Tropical Disease, Universitas Airlangga, Campus C UNAIR Jl. Mulyorejo Surabaya 60115, Indonesia
[3]Delima Husada Gresik Health Analyst Academy, Jl. Arif Rahman Hakim Gresik 61111, Indonesia
[4]Proteomik Laboratory, University-CoE-Research Center for Bio-Molecule Engineering, Universitas Airlangga, Campus C UNAIR, Jl. Mulyorejo Surabaya 60115, Indonesia
*Corresponding author. Tel/Fax: +62 899 048 4365
E-mail address: teguhharisucipto@staf.unair.ac.id

ARTICLE INFO

Article history
Received 12th Sep 2022
Accepted 21st Nov 2022

Keywords:
phylogenetic analysis
dengue
DENV1 sequence
tropical disease
neighbor-joining
UPGMA

ABSTRACT

Dengue virus (DENV1-4) belongs to the Flaviviridae family, which is transmitted by the *Aedes* mosquito vector and is the leading cause of dengue fever and dengue hemorrhagic fever. Since one of the DENV serotypes, DENV1, has become an endemic known to be circulating worldwide, including in Indonesia, it becomes necessary to carry out molecular epidemiological research using phylogenetic analysis with two methods, n*eighbor-joining* (NJ) and UPGMA. This study aims to analyze the DENV-1 relationship and obtain information regarding the differences between those methods, including the level of accuracy. This study used one DENV-1 sequence isolated in Surabaya, aligned with similar sequences on the GenBank. The results showed two comparisons. First, in the NJ method, the DENV-1 sequence samples in Surabaya with branch lengths 0,000 were similar to the DENV-1 in Malaysia, and Singapore, with branch lengths 0,000; 0,002; which belong to Genotype 1. The UPGMA method resulted in the DENV-1 sequence in Surabaya with a branch length of 0,000 similar to the DENV-1 in Malaysia with a branch length of 0,000, which belong to Genotype 1. Second, their level of accuracy, which is in the NJ method, the construction of phylogenetic trees is based on periodic evolutionary times. In contrast, UPGMA assumes that each sequence is found at the same evolutionary time, which makes this method less accurate than the NJ method. We can conclude that using the NJ method, the construction and analysis of the phylogenetic tree of the DENV1 sequence isolated in Surabaya have higher similarity and accuracy.

1. Introduction

Dengue virus (DENV1-4) belongs to Flavivirus, which is transmitted through the *Aedes* mosquito vector that infects mammal hosts, including humans, it has become an endemic which impacts all types of ages, gender, and socioeconomic status[1]. Dengue virus spreads to more than 100 countries, with an average of 50 million cases yearly [2], with the most susceptible hosts being children (Halstead, 1988). Seven types of non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) [1] of dengue are what cause the clinical symptoms they cause to vary, ranging

from the least severe, such as dengue fever (DF) which can be cured naturally, to causing death, namely DHF and dengue shock syndrome (DSS) [3-5].

The four dengue serotypes (DENV1-4) are known to be circulating worldwide [6]. The geographical distribution of each of these serotypes forms a genotypic variation that depends on the nucleotide sequence it has. Globally, it can be divided into several different genotypes. DENV1, which is the focus of this study, consists of five genotypes: (I) Southeast Asia, China, and East Africa, (II) Thailand between the 1950s and 1960s, (III) Malaysia, (IV) Western Pacific Islands dan Australia, (V) America, West America, Asia [5, 7-9].

A high increase of dengue disease infections (94.355) and death cases (853) in Indonesia until 2022 still become a severe problem [10]. In Surabaya, one of the largest cities in East Java province, the latest cases of DF amount to 187 [11], which must be considered because it continues to spread throughout the archipelago [12]. Based on this phenomenon, epidemiological research and information about the previously isolated strains are needed to explain better how the virus can spread and develop. While in reality, it still becomes a limited understanding of the virus's spread. In contrast, knowledge about the various sequences needs to be known to characterize genotypes to monitor the circulation of the disease in an area and to show later that changes in the genetic composition of viruses significantly affect the shift in the probability of spreading dengue virus serotype pathogens [13].

This study was conducted to do the phylogenetic analysis of the DENV-1 relationship and obtain information regarding the differences between *neighbor-joining* (NJ) and *unweighted pair group methods with arithmetic mean* (UPGMA) methods as a step to identify evolution, genetic diversity, and origin as well as to increase the ability to understand and predict the emergence of DENV1 in Indonesia, including the level of accuracy. We used one DENV-1 sequence isolated in Surabaya, where the establishment of a phylogenetic tree with data sequences on the NCBI GenBank website.

2. Materials and methods

2.1 Method

This study utilized secondary data from the NCBI (National Center for Biotechnology Information) GenBank website with BLAST, software for determining homology or similarities of a sequence of DNA or amino acids sequenced with the existing database at NCBI. In the molecular world, this method is significant for analyzing a sequence, one of which is determining kinship or phylogenetic analysis.

The target for the DENV1 sequence, which was collected in Surabaya, is the envelope region. Where according to Dang et al. (2020), DENV has a genomic structure consisting of the envelope (E), capsid (C), membrane (M), and ten types of structural and non-structural proteins [14]. This sequence is then matched with similar DNA on the NCBI website for phylogenetic analysis to determine the relationship of the sequences owned.

2.2 Data Analysis

Sequences pre-owned at the overlapping peaks based on the electropherogram at the 3' and 5' ends using the Bioedit Sequence Alignment Editor ver. 7.2.5 software. The sequences were then aligned with 50 sequences from the GenBank database obtained using the MEGA11 software. The next step is to carry out a phylogenetic analysis using two methods. The first method is the *neighbor-*

joining model evolution P distance with a bootstrap of 1000, while the second is the UPGMA with the same evolution model, including analyzing the comparison of these two methods from any literature.

3. Results and discussion

Phylogenetic analysis is of great value not only for studying the organism that harbors the gene but also for the evolutionary history of the gene, for example, whether a sequence is descended from a common ancestor or represents one or more ancient duplications [15]. In addition, Aryati (2012) also explained that genotyping as a determining step for the genetic diversity of a viral genome is very useful in molecular epidemiology so that it can unify the distribution or circulation of the genotype in an area. In Sasmono's study (2019), the DENV1 sequence isolated from Balikpapan was very close to strains from Surabaya, East Java, Indonesia, and Singapore that appeared in 2002-2003 and included in genotype one, which, based on analysis, is known to have occurred in 2000, while genotype four which appeared in 1991 was not found in East Kalimantan [12]. In another study by Dieng et al., (2021), the appearance of DENV1 in Senegal and Mali in 2015 originated in Asia and was then introduced to Medina Gounass, Senegal. It spread to other countries, namely Mali [2]. Arenas (2013) also explained that the DENV1 genotype IV virus found in Asia and the Pacific then spread to mainland India and America from the Philippines, which Indonesia is also a possible source of spread to Indian Ocean countries [7].

Sequences with overlapping peaks on the electropherogram are cut to avoid reading errors. Similar sequences of the cut results were then matched to similar DNA sequences at NCBI (National Center for Biotechnology Information) with BLAST. In this study, 50 similar sequences were used, which were then aligned. This alignment aimed to arrange deoxyribonucleic acid (DNA) sequences to identify regions of similarity that might result from functional, structural, or evolutionary relationships between sequences [16]. Aligning multiple sequences at once can indicate being in the same family. In addition, it can show all relationships or relations between families of existing sequences. After alignment, phylogenetic tree construction was carried out using the *neighbor-joining* and UPGMA methods.

The *neighbor-joining* method is very suitable when the evolutionary mean of the lineage separation is under different considerations. When the length of a branch of a tree whose topology is known changes by stimulating yang levels varies from evolutionary changes, the *neighbor-joining* method is the most suitable for predicting trees correctly [17]. *Neighbor-joining* selects a sequence that, if combined, will give the best estimate of the length of the closest branch reflecting the distance real between sequences [18].

The distance method gives a reasonable estimate of an evolutionary tree and is unaffected by inner variation, the average change along the tree branches. The UPGMA method is a simple method for tree construction that assumes an average change along the tree is constant, and distance is approximately ultrametric (ultrametric usually expressed as a molecular clock tree). Method UPGMA begins with a branch length calculation among the most closely related sequences, then average the distance between these sequences or sequence groups and the following sequence or sequence group and continues through all sequences included in the tree. This method predicts the root position of the tree [18].

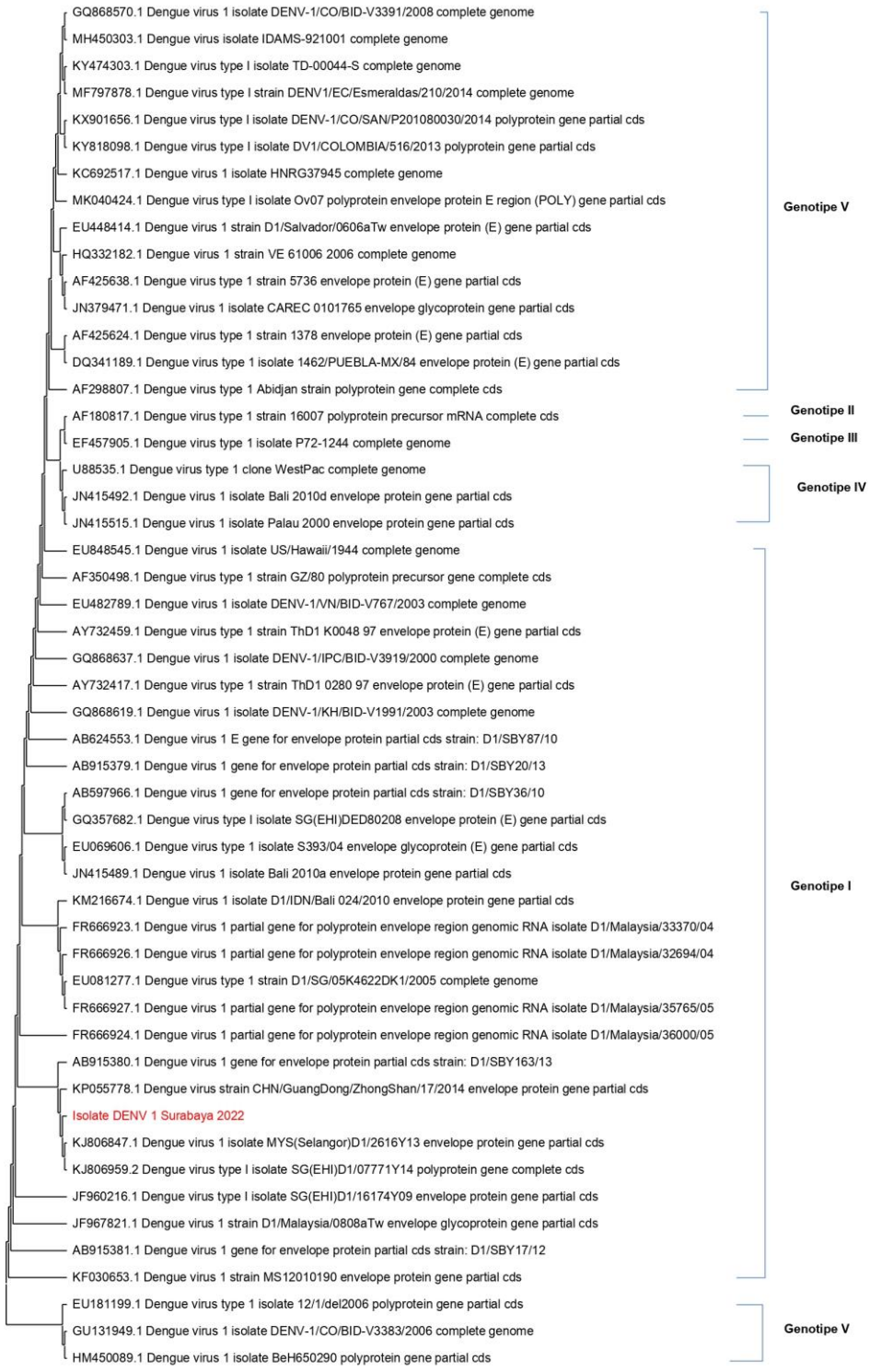

Figure 1. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree is shown. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. This analysis involved 51 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 11694 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [19].

Figure 2. Phylogeny analysis was inferred using the UPGMA method. The optimal tree is shown. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. This analysis involved 51 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 11694 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [19].

Based on Figure 1. the phylogenetic tree using the UPGMA method resulted in a sample of DENV-1 sequences in Surabaya having similarities with 2 DENV-1 sequences, namely MYS(Selangor)D1/2616Y13 (Malaysia) and SG(EHI)D1/07771Y14 (Singapore) which belong to Genotype 1. Then, when viewed in terms of form, the *neighbor-joining* method produces a form not rooted in the phylogenetic tree. Meanwhile, based on Figure 2. the phylogenetic tree using the UPGMA method resulted in the DENV-1 sequence sample in Surabaya being similar to the MYS(Selangor)D1/2616Y13 (Malaysia) DENV-1 sequence, which belongs to Genotype 1. Then, when viewed in terms of form, the UPGMA method produces a form rooted in a phylogenetic tree.

This analysis can also be seen from the length of the branches of the phylogenetic tree. The length of the branches in the tree represents a split in the evolutionary line of genes into two different species. The length of each branch at the next point indicates the number of altered sequences that occurred before its level of splitting. The branch length between sequences indicates that the species have the same evolutionary average [18]. The lengths of the DENV1 sequence branches that have similarities can be seen in the following table:

Table 1. Branch lengths Isolate DENV-1

Isolate DENV-1	Branch lengths
Sample isolate DENV-1 Surabaya	0,000
KJ806847.1 Dengue virus 1 isolate MYS(Selangor)D1/2616Y13 envelope protein gene partial cds	0,000
KJ806959.2 Dengue virus type I isolate SG(EHI)D1/07771Y14 polyprotein gene complete cds	0,002

Based on the branch length data in the table, the UPGMA method produced a DENV-1 sample with a branch length of 0.000, similar to the MYS(Selangor)D1/2616Y13 (Malaysia) sequence with the same branch length of 0.000. UPGMA adopts the same evolutionary rate in all lines, i.e., the mutation rate is constant over time and for all lineages in the tree. This is called the 'molecular clock hypothesis'. This means all leaves (terminal nodes) are equidistant from the roots [20]. Phylogenetic analysis using the UPGMA method does not pay attention to rate evolution, where all isolates are at the same evolution rate, which causes this method to be less accurate.

Meanwhile, in the *neighbor-joining* method, the DENV-1 sequence with a branch length of 0.000 is similar to the DENV-1 MYS(Selangor)D1/2616Y13 (Malaysia) with a branch length of 0.000 and SG(EHI)D1/07771Y14 (Singapore) with a branch length of 0.002. This is because the *neighbor-joining* method considers the evolution rate between isolates, which is based on the principle of minimum evolution with estimated branch lengths so that the resulting branch lengths

are not the same but will be proportional to the changes. In the *neighbor-joining* method, the branch length of the tree is stimulated with varying degrees of evolutionary change, so this method is the most suitable for predicting trees correctly [17].

The UPGMA and Neighbor-Joining methods are two methods for constructing distance-based phylogenetic trees. Thus, the complexity and reliability of the phylogenetic tree generated from the *neighbor-joining* tree method are high. But not as fast as the UPGMA method. For more details, Neighbor Joining is a repeated clustering method based on minimum evolution criteria; topology with the least total branch length is preferred at each step. Meanwhile, UPGMA is an agglomerative hierarchical grouping based on the average linkage method [21].

4. Conclusions

The phylogenetic tree using the UPGMA method resulted in the DENV-1 sequence sample in Surabaya being similar to the MYS(Selangor)D1/2616Y13 (Malaysia) DENV-1 sequence, which belongs to Genotype 1. The phylogenetic tree using the *neighbor-joining* method resulted in a sample of DENV-1 sequences in Surabaya, with branch length 0,000 having similarities with 2 DENV-1 sequences, namely MYS(Selangor)D1/2616Y13 (Malaysia) and SG(EHI)D1/07771Y14 (Singapore) with branch lengths 0,000; 0,002; which belong to Genotype 1.

The UPGMA and *neighbor-joining* methods are two methods for constructing distance-based phylogenetic trees. UPGMA results in a rooted phylogenetic tree, while the *neighbor-joining* tree method produces an unrooted phylogenetic tree. Their level of accuracy, which is in the *neighbor-joining* method, the construction of phylogenetic trees is based on periodic evolutionary times. In contrast, UPGMA assumes that each sequence is found at the same evolutionary time, which makes this method less accurate than the *neighbor-joining* method. We can conclude that using the *neighbor-joining* method, the construction and analysis of the phylogenetic tree of the DENV1 sequence isolated in Surabaya have higher similarity and accuracy.

Acknowledgments

The authors are indebted to the Institute of Tropical Disease Universitas Airlangga for financial and technical support to this research.

References

[1] M. Y. Carrillo-Hernandez, J. Ruiz-Saenz, L. Jaimes-Villamizar, S. M. Robledo-Restrepo, and M. J. P. O. Martinez-Gutierrez, "Phylogenetic and evolutionary analysis of dengue virus serotypes circulating at the Colombian–Venezuelan border during 2015–2016 and 2018–2019," vol. 16, no. 5, p. e0252379, 2021.

[2] I. Dieng *et al.*, "Origin and spread of the dengue virus type 1, genotype V in Senegal, 2015–2019," vol. 13, no. 1, p. 57, 2021.

[3] K. H. Eckels and R. J. A. i. v. r. Putnak, "Formalin-inactivated whole virus and recombinant subunit flavivirus vaccines," vol. 61, pp. 395-418, 2003.

[4] S. Miller, S. Kastner, J. Krijnse-Locker, S. Buhler, and R. J. J. o. B. C. Bartenschlager, "The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner," vol. 282, no. 12, pp. 8873-8882, 2007.

[5] F. Sjatha, Y. Takizawa, A. Yamanaka, E. J. I. Konishi, Genetics, and Evolution, "Phylogenetic analysis of dengue virus types 1 and 3 isolated in Jakarta, Indonesia in 1988," vol. 12, no. 8, pp. 1938-1943, 2012.

[6] W. J. G. W. H. O. Searo, "Comprehensive guidelines for prevention and control of dengue and dengue hemorrhagic fever," pp. 1-212, 2011.
[7] C. J. Villabona-Arenas and P. M. d. A. J. P. o. Zanotto, "Worldwide spread of dengue virus type 1," vol. 8, no. 5, p. e62649, 2013.
[8] S. C. Weaver, N. J. I. Vasilakis, genetics, and evolution, "Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease," vol. 9, no. 4, pp. 523-540, 2009.
[9] H. Harapan, A. Michie, R. T. Sasmono, and A. J. V. Imrie, "Dengue: a minireview," vol. 12, no. 8, p. 829, 2020.
[10] N. A. Loasana. (2022). *Dengue outbreak bites as COVID-19 cases rise*. Available: https://www.thejakartapost.com/indonesia/2022/07/22/dengue-outbreak-bites-as-covid-19-cases-rise.html
[11] P. Surabaya. (2022). *ANTISIPASI DBD, WALI KOTA ERI CAHYADI AJAK SEMUA ELEMEN MASYARAKAT GENCARKAN PSN*. Available: https://surabaya.go.id/id/berita/70223/antisipasi-dbd-wali-kota-eri-cahyadi-ajak-semua-elemen-masyarakat-gencarkan-psn
[12] R. T. Sasmono *et al.*, "Multiple introductions of dengue virus strains contribute to dengue outbreaks in East Kalimantan, Indonesia, in 2015–2016," vol. 16, pp. 1-15, 2019.
[13] A. J. I. J. O. C. P. Aryati and M. LABORATORY, "Analisis Filogenetik Dengue Di Indonesia," vol. 18, no. 2, pp. 111-116, 2012.
[14] T. T. Dang, M. H. Pham, H. V. Bui, and D. J. V. j. Van Le, "Whole genome sequencing and genetic variations in several dengue virus type 1 strains from unusual dengue epidemic of 2017 in Vietnam," vol. 17, no. 1, pp. 1-10, 2020.
[15] D. E. Soltis and P. S. J. P. p. Soltis, "The role of phylogenetics in comparative genetics," vol. 132, no. 4, pp. 1790-1800, 2003.
[16] V. Singh, A. Singh, R. Chand, and C. J. I. J. B. R. Kushwaha, "Role of bioinformatics in agriculture and sustainable development," vol. 3, no. 2, pp. 221-226, 2011.
[17] N. Saitou, M. J. M. b. Nei, and evolution, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," vol. 4, no. 4, pp. 406-425, 1987.
[18] N. J. W. Dharmayanti, "Filogenetika molekuler: metode taksonomi organisme berdasarkan sejarah evolusi," vol. 21, no. 1, pp. 1-10, 2011.
[19] K. Tamura, G. Stecher, S. J. M. b. Kumar, and evolution, "MEGA11: molecular evolutionary genetics analysis version 11," vol. 38, no. 7, pp. 3022-3027, 2021.
[20] R. R. Sokal and C. J. E. Michener, "A quantitative approach to a problem in classification," vol. 11, no. 2, pp. 130-162, 1957.
[21] G. A. Pavlopoulos, T. G. Soldatos, A. Barbosa-Silva, and R. J. B. m. Schneider, "A reference guide for tree analysis and visualization," vol. 3, pp. 1-24, 2010.