



# Genomic analysis of dengue virus serotype 1 (DENV-1) genotypes from Surabaya, Indonesia

Benediktus Yohan<sup>1</sup> · Puspa Wardhani<sup>2,3</sup> · Hidayat Trimarsanto<sup>1,4</sup> · A. Aryati<sup>2,3</sup> · R. Tedjo Sasmono<sup>1</sup> 

Received: 14 February 2018 / Accepted: 29 March 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

Dengue has caused a significant public health impact globally. With the diverse genetic of the causative viruses, analysis of dengue virus (DENV) genomes is important to supplement epidemiological data with information that can be used to reconstruct the history of epidemics in time and space. We have reported the clinical and virological characteristics of dengue in Surabaya, Indonesia and revealed the presence of all four DENV serotypes and the predominance of DENV-1. The further classification of Surabaya DENV-1 into two different genotypes warrants in-depth genomic analysis to study the dynamics of both genotypes and their contribution to virus evolution, virus transmission, and disease. We performed full-length genome sequencing to nine isolates' representatives from DENV-1 Genotype I and Genotype IV. Phylogenetic and evolutionary analyses suggested the more recent introduction of Genotype I viruses compared to the more endemic Genotype IV. Comparative analysis of Surabaya DENV-1 genomes and other sequences available publicly revealed that the majority of the DENV-1 codons were under strong purifying selection, while seven codon sites identified to be under positive selection. We highlight a unique codon site under the positive pressure in the NS1 gene of DENV-1. Our results provide additional genomic data of DENV from Indonesia that may contribute to the better understanding of dengue disease dynamics.

**Keywords** Dengue · Indonesia · Surabaya · DENV-1 · Genome

Dengue is an acute febrile disease that is caused by dengue virus (DENV), a member of the *Flaviviridae* family. The disease has been established globally in both endemic and epidemic transmission cycles and caused significant human health problem [1]. Global spread of dengue, facilitated by urbanization and international travel, has been marked by worldwide expansion of the DENV serotypes and disease hyperendemicity [2].

A diverse genetic characteristics has been shown by DENVs as reflected by the presence of four serotypes, namely DENV-1, -2, -3, and -4 [3]. Each serotype of DENV further harbors extensive genetic diversity in the form of phylogenetically distinct clusters termed genotypes. These genotypes differ in both their geographical distributions, fitness, and virulence [4, 5]. The ~10.7 kb single-stranded positive-sense RNA genome encodes 3 structural (C, prM/M, E) and 7 non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) proteins within a single open-reading frame (ORF) [6]. Genetic analysis of DENV is important to supplement epidemiological data with information that can be used to reconstruct the history of epidemics in time and space [7].

Since virus evolution has been known as a significant force that drives epidemiologic changes, analysis of the entire viral genome sequence is thus an important tool both for studies of the mechanism of virus virulence as well as surveillance for advanced dengue epidemics warning [8]. Phylogenetic analysis of viral genomic sequences can be used to understand DENV evolution and its effects on virus transmission and disease. Moreover, phylogenetic analysis of

---

Edited by Lorena Passarelli.

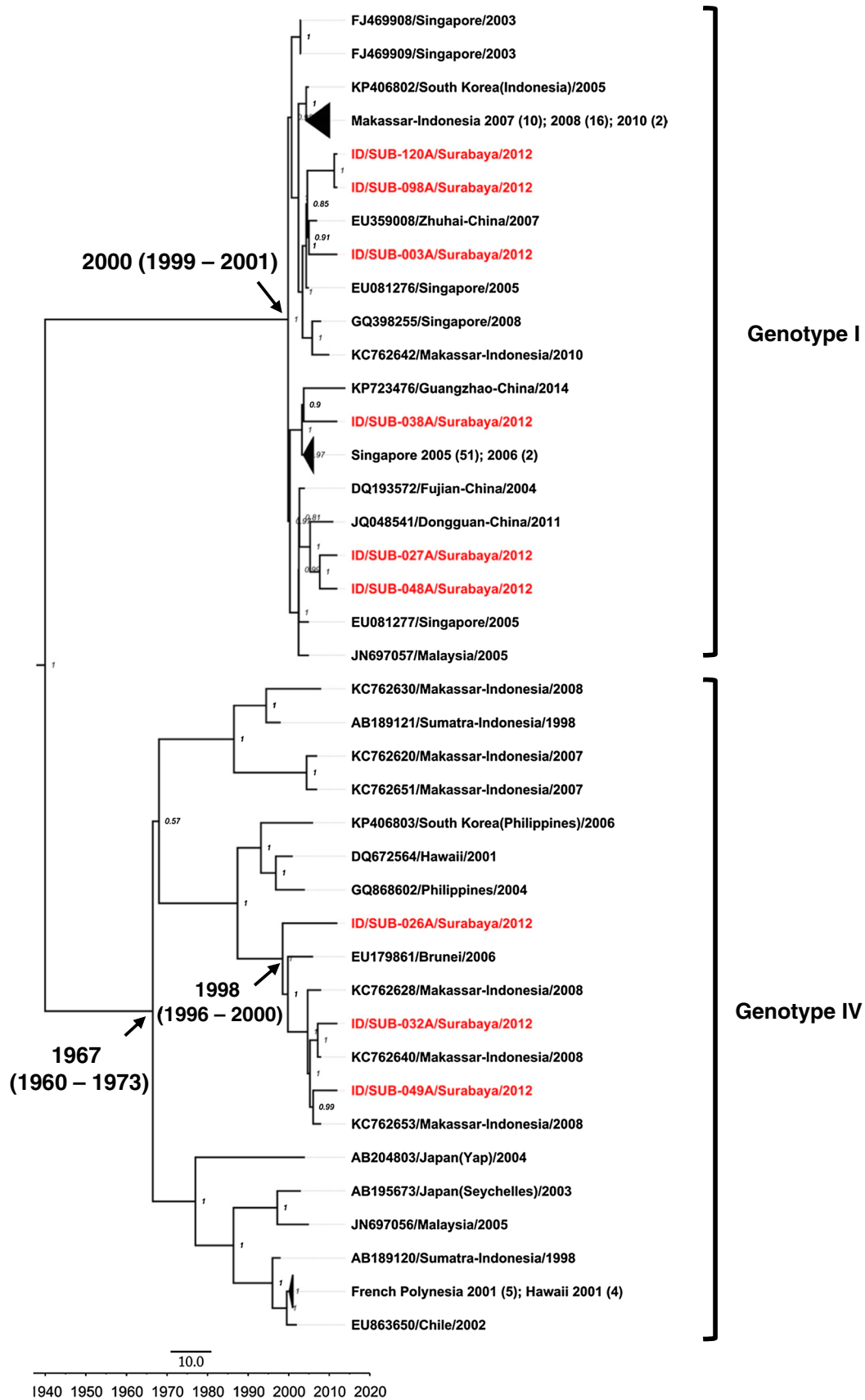
✉ R. Tedjo Sasmono  
sasmono@eijkman.go.id

<sup>1</sup> Eijkman Institute for Molecular Biology, Ministry of Research, Technology, and Higher Education, Jl. Diponegoro 69, Jakarta 10430, Indonesia

<sup>2</sup> Department of Clinical Pathology, School of Medicine, Universitas Airlangga, Surabaya, Indonesia

<sup>3</sup> Institute for Tropical Diseases, Universitas Airlangga, Surabaya, Indonesia

<sup>4</sup> Agency for the Assessment and Application of Technology, Jakarta, Indonesia



**Fig. 1** Phylogeny of the DENV-1 Genotype I and IV strains generated by Bayesian inference method as implemented in BEAST using GTR+G+I evolution model calculated using complete open-reading frame (ORF) sequences. The red labels indicated the isolates from Surabaya (this study). The number in the node indicated the posterior probability of that particular cluster

the full genome provides greater resolution on the dynamics of dengue epidemics than partial genome sequencing [9, 10].

We have reported the clinical and virological characteristics of dengue in Surabaya, Indonesia in a cross-sectional study performed in 2012 [11]. All four DENV serotypes were detected with the predominance of DENV-1. Furthermore, the diverse genetic of DENV in Surabaya was shown with the presence of two different genotypes of DENV-1, i.e., Genotype I and Genotype IV. In order to study the dynamics of these two genotypes, full-length genome sequencing was performed to nine isolates' representatives. The complete genome sequencing approach was performed using Sanger capillary method, as previously described [12].

Sequence reads of DENV-1 isolates from Surabaya were assembled to generate contigs based on DENV-1 reference sequence (GenBank accession no. NC\_001477). Genome analysis was performed to compare six isolates of Genotype I and three isolates of Genotype IV. Nucleotide sequences and deduced amino acid analysis were analyzed using MEGA 6.0 software. The sequences consisted of 10,735 nt, with ORF of 10,179 nt (3392 aa). The complete genome sequences have been submitted and granted GenBank accession numbers of KY057365–73. Multiple sequence alignment of the ORFs revealed 124 aa differences (3.7%) where most of the aa differences were specific of the two Genotype groups (data not shown). Amino acid diversity mostly occurred in Envelope gene (4.04% of the aa), followed by NS5 gene (2.78%). No recombination event occurred in the dataset, as analyzed using RDP4 v.4.56 software.

In order to analyze phylogenetic and evolutionary information and the relatedness of Surabaya DENV-1 isolates with other isolates worldwide, complete genome sequences were aligned together with all publicly available DENV-1 complete genome sequences in GenBank. The initial screening yielded dataset of 1640 DENV-1 taxa with complete ORF. Multiple sequence alignment was performed using MAFFT rapid alignment software. Initial phylogenetic tree analysis was done to generate UPGMA tree based on genetic distance of the ORF sequences. For clarity of tree view, strains with known isolation year which are most closely related to our Surabaya isolates were selected, generating a set of 121 taxons from Genotypes I and IV. The tree trimming and pruning steps were done using Jalview desktop 2.9 software. Phylogenetic analysis of ORF sequences revealed the grouping of DENV-1 Surabaya into Genotype I and IV, based on the DENV-1 classifications by Goncalvez et al.

[13] (Fig. 1). Six isolates were grouped into Genotype I and further grouped into three different clades. The isolates were closely related to strains circulating in China. The other three isolates were grouped into Genotype IV and closely related to strains from Brunei and Makassar, Indonesia.

The phylogenetic and evolutionary analyses were inferred using Bayesian Markov Chain Monte Carlo (MCMC) method as implemented in BEAST v.1.8.4. The selection of best-fitted model was done using jModelTest v.2.1.4 [14], in which GTR + G+I and relaxed uncorrelated lognormal clock molecular clock analysis were applied in BEAST analysis with the tip of each taxon calibrated using the year of isolation. Tree analysis was performed using Bayesian skyline prior, with 100 million chains sampled for every 1000<sup>th</sup> iteration and 10% burn-in. The initial estimated evolutionary rate was set at  $7.6 \times 10^{-4}$  substitutions per site per year, as previously described [7]. The resulting DENV-1 tree age was estimated to be approximately 76 years with the DENV-1 mean evolutionary rate of  $8.44 \times 10^{-4}$  subs/site/year (95% HPD:  $7.37\text{--}9.54 \times 10^{-4}$ ), as calculated by BEAST. The calculated coefficient of variation was 0.19 (95% HPD: 0.10–0.28). The time to the most recent common ancestors (TMRCA) analysis estimated the Genotype I of DENV-1 to have emerged circa the year 2000 (95% HPD: 1999–2001), while the Genotype IV has TMRCA since the year 1967 (95% HPD: 1960–1973). The Surabaya Genotype IV isolates, however, have been estimated to have common ancestors in the year 1998 (1996–2000). Phylogenetic and evolutionary analyses suggested the more recent introduction of DENV-1 Genotype I viruses compared to the more endemic Genotype IV (Fig. 1). We and others have described the occurrence of lineage replacement in DENV-1 in the Southeast Asian region [11, 12, 15–17]. It is argued that the replacement of older Genotype IV was related to the better viral fitness of the newly introduced Genotype I [12]. The other possible explanations for this phenomena were the occurrence of purifying selection or virus population bottlenecks related to vector population and density [15].

The codons of the DENV-1 ORF were subjected to selection pressure analysis. The dataset was investigated using web-based HyPhy application within the Datamonkey webserver. Initial analysis of episodic diversifying selection was performed using BUSTED (Branch-site Unrestricted Statistical Test for Episodic Diversification) and further analyses under the general reversible (REV) nucleotide substitution model utilized in four different approaches, including SLAC (Single-Likelihood Ancestor Counting), FEL (Fixed Effects Likelihood), MEME (Mixed Effects Model of Evolution), and FUBAR (Fast, Unconstrained Bayesian Approximation). The BUSTED analysis revealed the evidence of episodic diversifying selection with LRT  $p$  value of  $1.94 \times 10^{-8}$ . Further selection pressure analysis revealed that the majority of the

**Table 1** Selection pressure analysis of the open-reading frame (ORF) sequence dataset (3392 aa) of DENV-1 Genotype I and IV ( $N=121$ ) using four different methods

Codon site (ORF)	Codon position (protein)	SLAC		FEL		MEME		FUBAR	
		DN/dS ( $\omega$ )	$P$	DN/dS ( $\omega$ )	$P$	DN/dS ( $\omega$ )	$P$	DN/dS ( $\omega$ )	PP
<i>prM</i>									
232	118	0.293	1.472	8.404	<b>0.074</b>	1.928	<b>0.088</b>	1.448	0.892
252	138	0.331	1.511	3.450	<b>0.059</b>	409.335	<b>0.001</b>	0.285	0.771
<i>NS1</i>									
869	94	0.326	1.125	3.971	<b>0.056</b>	4.517	<b>0.097</b>	0.335	0.761
1053	278	0.308	2.192	11.295	<b>0.003</b>	12.985	<b>0.030</b>	3.082	<b>0.988</b>
<i>NS2A</i>									
1286	159	0.328	1.882	4.389	<b>0.040</b>	5.919	<b>0.048</b>	0.549	0.852
<i>NS3</i>									
1594	119	0.333	1.515	3.194	<b>0.066</b>	2.427	<b>0.088</b>	0.275	0.767
<i>NS5</i>									
2778	285	0.308	1.459	6.233	<b>0.022</b>	23.782	<b>0.022</b>	0.706	0.859

Significant evidence of positive selection analysis was assessed using criteria:  $p$  value  $<0.1$  in SLAC, FEL, and MEME, and posterior probability (PP)  $>0.9$  in FUBAR, written in bold font. The sites with evidence of positive selection by at least two different methods are listed

DENV-1 codons were under strong negative (purifying) selection. The dominant negative selection pressure was also observed in other studies involving DENV-1 [16, 18]. This event may be introduced by the elimination of deleterious mutation strains by purifying selection [19]. Nevertheless, the analysis detected a total number of 1, 11, and 55 positively selected sites using FUBAR, FEL, and MEME methods, respectively. There was no positively selected sites detected using SLAC method. We observed seven codon sites identified to be under positive selection by two or more methods in the *prM*, *NS1*, *NS2A*, *NS3*, and *NS5* genes (Table 1). One codon site number 1053 of ORF or codon number 278 of the *NS1* gene was detected under positive selection by three different methods. It has been proposed that DENV evolution that drives lineage turnover can be attributed in part to positive selection, in particular on the *NS2A* gene [20]. Our analysis confirmed the presence of positively selected codon in this gene. The identification of positive pressure in *prM*, *NS3*, and *NS5* genes of DENV-1 has also been reported [18]. Here, we report the detection of a unique site of selection pressure in *NS1* gene that is crucial for viral replication and viability. The *NS1* gene has been considered as the major target of positive selection during flavivirus speciation [21]. The positively selected codon was mapped into the epitope region 3 of DENV *NS1* gene and thus may be related to immune recognition [22].

In conclusion, we report here the genetic diversity of DENV-1 from Surabaya, Indonesia. The evolutionary and selection pressure data generated may add to the better understanding on dengue epidemics and surveillance purposes for the advanced warning of outbreaks.

**Acknowledgements** This study was funded by grant from the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia.

**Authors' contributions** BY, PW, AA, and RTS conceived and designed the study. PW and AA collected the samples. BY and PW performed the experiments. BY and HT performed the data analysis. BY wrote the first draft. HT and RTS reviewed and edited the manuscript. All authors read and approved the final manuscript.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were approved by the Ethic Commission of the Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

1. S. Bhatt, P.W. Gething, O.J. Brady, J.P. Messina, A.W. Farlow, C.L. Moyes, J.M. Drake, J.S. Brownstein, A.G. Hoen, O. Sankoh, M.F. Myers, D.B. George, T. Jaenisch, G.R.W. Wint, C.P. Simmons, T.W. Scott, J.J. Farrar, S.I. Hay, *Nature* **496**, 504 (2013)
2. J.P. Messina, O.J. Brady, T.W. Scott, C. Zou, D.M. Pigott, K.A. Duda, S. Bhatt, L. Katzelnick, R.E. Howes, K.E. Battle, C.P. Simmons, S.I. Hay, *Trends Microbiol.* **22**, 138 (2014)
3. E.C. Holmes, *J. Clin. Investig.* **119**, 2488 (2009)
4. E.C. Holmes, S.S. Burch, *Trends Microbiol.* **8**, 74 (2000)
5. E.C. Holmes, S.S. Twiddy, *Infect. Genet. Evol.* **3**, 19 (2003)
6. M.G. Guzman, S.B. Halstead, H. Artsob, P. Buchy, J. Farrar, D.J. Gubler, E. Hunsperger, A. Kroeger, H.S. Margolis, E. Martínez,

- M.B. Nathan, J.L. Pelegrino, C. Simmons, S. Yoksan, R.W. Peeling, *Nat. Rev. Microbiol.* **8**, S7 (2010)
7. R.L. Costa, C.M. Voloch, C.G. Schrago, *Infect. Genet. Evol.* **12**, 309 (2012)
  8. J.G. Christenbury, P.P.K. Aw, S.H. Ong, M.J. Schreiber, A. Chow, D.J. Gubler, S.G. Vasudevan, E.E. Ooi, M.L. Hibberd, *J. Virol. Methods* **169**, 202 (2010)
  9. S.H. Ong, J.T. Yip, Y.L. Chen, W. Liu, S. Harun, E. Lystiyaningsih, B. Heriyanto, C.G. Beckett, W.P. Mitchell, M.L. Hibberd, A. Suwandono, S.G. Vasudevan, M.J. Schreiber, *Infect. Genet. Evol.* **8**, 191 (2008)
  10. M.J. Schreiber, E.C. Holmes, S.H. Ong, H.S.H. Soh, W. Liu, L. Tanner, P.P.K. Aw, H.C. Tan, L.C. Ng, Y.S. Leo, J.G.H. Low, A. Ong, E.E. Ooi, S.G. Vasudevan, M.L. Hibberd, *J. Virol.* **83**, 4163 (2009)
  11. P. Wardhani, A. Aryati, B. Yohan, H. Trimarsanto, T.Y. Setianingsih, D. Puspitasari, M.V. Arfijanto, B. Bramantono, S. Suharto, R.T. Sasmono, *PLoS ONE* **12**, e0178443 (2017)
  12. R.T. Sasmono, I. Wahid, H. Trimarsanto, B. Yohan, S. Wahyuni, M. Hertanto, I. Yusuf, H. Mubin, I.J. Ganda, R. Latief, P.J. Bifani, P.-Y. Shi, M.J. Schreiber, *Infect. Genet. Evol.* **32**, 165 (2015)
  13. A.P. Goncalvez, A.A. Escalante, F.H. Pujol, J.E. Ludert, D. Tovar, R.A. Salas, F. Liprandi, *Virology* **303**, 110 (2002)
  14. D. Darriba, G.L. Taboada, R. Doallo, D. Posada, *Nat. Methods* **9**, 772 (2012)
  15. C. Zhang, M.P. Mammen, P. Chinnawirotpisan, C. Klungthong, P. Rodpradit, P. Monkongdee, S. Nimmanitya, S. Kalayanaroj, E.C. Holmes, *J. Virol.* **79**, 15123 (2005)
  16. V. Duong, C. Simmons, L. Gavotte, A. Viari, S. Ong, N. Chantha, N.J. Lennon, B.W. Birren, S. Vong, J.J. Farrar, M.R. Henn, V. Deubel, R. Frutos, P. Buchy, *Infect. Genet. Evol.* **15**, 59 (2013)
  17. A. Yamanaka, K.C. Mulyatno, H. Susilowati, E. Hendrianto, A.P. Ginting, D.D. Sary, F.A. Rantam, S. Soegijanto, E. Konishi, *PLoS ONE* **6**, e27322 (2011)
  18. P.K. Dash, S. Sharma, M. Soni, A. Agarwal, A.K. Sahni, M. Parida, *Virus Res.* **195**, 124 (2015)
  19. E.C. Holmes, *J. Virol.* **77**, 11296 (2003)
  20. S.N. Bennett, E.C. Holmes, M. Chirivella, D.M. Rodriguez, M. Beltran, V. Vorndam, D.J. Gubler, W.O. McMillan, *Mol. Biol. Evol.* **20**, 1650 (2003)
  21. M. Sironi, D. Forni, M. Clerici, R. Cagliani, *PLoS Negl. Trop. Dis.* **10**, e0004978 (2016)
  22. P. Masrinoul, M.O. Diata, S. Pambudi, K. Limkittikul, K. Ikuta, T. Kurosu, *Jpn. J. Infect. Dis.* **64**, 109 (2011)