



# Whole genome sequencing of Indonesian dengue virus isolates using next-generation sequencing

Benediktus Yohan<sup>1</sup>, Rama Dhenni<sup>1</sup>, Rahma F. Hayati<sup>1</sup>, Frilasita Aisyah Yudhaputri<sup>1</sup>, Dionisius Denis<sup>1</sup>, Yanuarni Wulandari Baji Pamai<sup>2</sup>, Anna Martiana Afida<sup>3</sup>, Ingrid A. Hutagalung<sup>4</sup>, Sotianingsih Haryanto<sup>5</sup>, Hidayat Trimarsanto<sup>1,6</sup>, Khin Saw Aye Myint<sup>1</sup>, and R. Tedjo Sasmono<sup>1,\*</sup>

<sup>1</sup>Eijkman Institute for Molecular Biology, Ministry of Research, Technology, and Higher Education, Jalan Diponegoro No. 69, Jakarta 10430, Indonesia

<sup>2</sup>Santa Elisabeth Hospital, Jalan Anggrek Blok II, Lubuk Baja Kota, Kota Batam, Kepulauan Riau 29444, Indonesia

<sup>3</sup>Dr. H.M. Anshari Saleh Hospital, Jalan Brig. Jend. H. Hasan Basri No. 1, Alalak Utara, Banjarmasin Utara, Banjarmasin, South Kalimantan 70125, Indonesia

<sup>4</sup>Dr. M. Haulussy Hospital, Jalan Dr. Kayadoe, Benteng, Nusaniwe, Ambon, Maluku 97116, Indonesia

<sup>5</sup>Faculty of Medicine and Health Science, Jambi University, Jalan Letjen Suprpto No. 33, Telanaipura, Jambi 36361, Indonesia

<sup>6</sup>Agency for the Assessment and Application of Technology, Jalan MH Thamrin 8, Jakarta 10340, Indonesia

\*Corresponding author: [sasmono@eijkman.go.id](mailto:sasmono@eijkman.go.id)

SUBMITTED 14 Sept 2018 REVISED 19 Nov 2018 ACCEPTED 11 Dec 2018

**ABSTRACT** Indonesia is a tropical country and hyperendemic for dengue. The disease prevalently affected Indonesian and it caused high morbidity and substantial economic burden. This vector-borne viral disease is caused by infection of dengue viruses (DENVs), which are the member of Flaviviridae family. While most of dengue studies in Indonesia focused on the epidemiology, the clinical aspects, the vectors, and to certain extent the virology, there were still gaps in the DENVs genomic aspects. Considering their high mutation rate, the DENVs were known for their high genetic diversity and it might affect the characteristics of the viruses. Comprehensive DENV genomic data were thus important for many aspects of disease management, including virus surveillance, pathogenesis, diagnostics, antiviral drug design, and vaccine development. We established in this study a method for DENV whole genome sequencing using the advanced Next-Generation Sequencing (NGS) and Nextera XT DNA library preparation kit, coupled with simplified bioinformatic analysis methods. The Indonesian DENVs from four serotypes were isolated from patients' sera, while library was prepared from enriched templates and sequenced using Illumina NGS. Our study highlighted the potential of a robust NGS method in producing whole genome sequence of DENVs, which would be important for future dengue studies.

**KEYWORDS** dengue; genome; Indonesia; next-generation sequencing

## 1. Introduction

Dengue disease is endemic in Indonesia and currently remains the most important arboviral disease in the country. Indonesia continually reports the highest number of dengue cases in the WHO Southeast Asia Region (WHO 2012). Dengue fever, an acute febrile disease, is caused by dengue virus (DENV) infection. Four DENV serotypes (DENV-1, -2, -3, and -4) circulate in tropical and subtropical regions of the world, including Indonesia. DENV is transmitted through the aid of *Aedes* mosquitoes as the vector, mainly *Aedes aegypti* and *Ae. albopictus* (Lambrechts et al. 2010). The bite of infected mosquito to a susceptible human host may lead to dengue clinical manifestations, ranging from asymptomatic or a mild flu-like syndrome known as classic Dengue fever (DF), to the more severe form known as Dengue hemorrhagic fever (DHF) and the potentially lethal Dengue shock syndrome (DSS)

(Martina et al. 2009). The DENV genome is a ~10.7 kb single-stranded positive-sense RNA encoding 3 structural (C, prM/M, E) and 7 non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) proteins (Guzman et al. 2010). DENV has very diverse genetic characteristics, as shown by the presence of four serotypes. The diversity is continued to the level of within serotype, in which there are several clusters of variants termed as genotypes (Holmes 2009).

Despite the endemicity of the disease, genomic data of DENV in Indonesia is scarce. The number of DENV complete genome sequences from Indonesia searched on GenBank public database revealed approximately only 2.7% of the total number of DENV genomes available from all over the world (119/4,297 genomes, accessed mid 2018), suggesting that Indonesia is under-sampled for DENV genomic studies. Genomic data can provide many benefits such as powerful insights into disease transmission and dy-

namics, especially for rapidly evolving RNA viruses such as DENV (Costa et al. 2012). Comparative analysis of DENV genomes has been used widely to study the genetic diversity of the dengue viruses (Holmes and Burch 2000) and previous studies have also suggested a correlation between viral pathogenicity with its genetic structures (Leitmeyer et al. 1999). The in-depth analysis of DENV genome is also important to supplement epidemiological data with information that can be used to reconstruct the history of epidemics in time and space (Yohan et al. 2018).

Given the endemicity of dengue in Indonesia and the large gap on DENV genomic data that is currently present, efforts to generate DENV genome data from all regions in Indonesia should be sought. Using the conventional Sanger micro-capillary sequencing, we have been able to sequence the complete genomes of 80 DENV isolates from Makassar (Sasmono et al. 2015). This genomic study was the first and only report on the Indonesia DENV genomes in which the sequencing was performed solely in Indonesia using the available sequencing facility in our Institute.

The advent of new technologies for rapid genome sequencing, in which many millions of nucleotides can be obtained in a single run, has greatly enhanced the endeavors in genomic studies (Capobianchi et al. 2013; van Dijk et al. 2014). The next generation sequencing (NGS) technologies allow rapid and cost-effective acquisition of full-length genomes and enable significant contributions to multiple areas in virology, including virus discovery and metagenomics (Quiñones-Mateu et al. 2014).

In this study, we utilize the NGS technology to sequence DENV genomes. We presented NGS method that is relatively simple and can be used by scientists with limited resources and knowledge in bioinformatics. DENV isolated from archived serum samples were sequenced, and phylogenetic analyses were conducted to understand their genetic characteristics.

## 2. Materials and methods

### 2.1. Sample preparation and DENV isolation in tissue culture

The study utilized archived specimens collected from dengue surveillance studies in Jambi (Haryanto et al. 2016) and East Kalimantan (manuscript submitted) in 2014–2016. The representatives of each DENV serotype were tested for this method and are listed in Table 1. Ethical clearance for using archived specimens was granted

from Eijkman Institute Research Ethics Commission document No. 113. Serum samples were separated from 3–5 mL of blood using a blood-clot activator Vacutainer (BD Biosciences, NJ, USA) and centrifugation at  $3000 \times g$ . The collected serum was stored in aliquots at  $-80^{\circ}\text{C}$ .

The DENV-positive serum sample was proceed to virus inoculation attempt, using protocol described elsewhere (Fahri et al. 2013). In brief, an amount of 200  $\mu\text{L}$  of serum was prepared in RPMI medium supplemented with 2% of Fetal Bovine Serum (FBS) (Gibco-Thermo Fisher Scientific, Carlsbad, CA) and inoculated into a monolayer of C6/36 (*Aedes albopictus*, midgut) cells in T-25 flask. The inoculated cells were allowed for virus adsorption for 1 h at  $28^{\circ}\text{C}$  incubator, followed by medium change and incubation up to 14 d. Culture supernatant was harvested when cytopathic effect (CPE) detected or at the end of incubation period and tested for the presence of DENV.

### 2.2. RNA extraction, DENV detection, and serotyping

RNA was extracted from 140  $\mu\text{L}$  of serum sample or tissue culture supernatant using QIAamp Viral RNA mini kit (Qiagen, Hilden, Germany), according to protocol described by the manufacturer without the addition of carrier RNA. The extracted RNA was eluted in a total volume of 60  $\mu\text{L}$  and stored at  $-80^{\circ}\text{C}$  until use. The presence of DENV RNA and the corresponding serotype was determined by simultaneous real-time RT-PCR for DENV detection and serotyping assay using Simplexa Dengue kit (DiaSorin, Sallugia, Italy), as described elsewhere (Sasmono et al. 2014).

### 2.3. DENV cDNA preparation and PCR amplification of DENV whole genome fragments

The extracted RNA was subjected to cDNA preparation and whole-genome PCR amplification according to protocols previously described (Ong et al. 2008; Christenbury et al. 2010; Sasmono et al. 2015). The DENV cDNA was generated using extracted RNA as template and Superscript III Reverse Transcriptase (RT) enzyme kit (Invitrogen-Thermo Fisher Scientific), according to protocol recommended by the manufacturer. The RNase OUT recombinant ribonuclease inhibitor (Invitrogen) was added to maintain the integrity of initial RNA and a set of DENV serotype-specific antisense primers covering the whole genome of DENV was used. The RT reaction was allowed to run at  $50^{\circ}\text{C}$  for 1 h and products were stored at  $-20^{\circ}\text{C}$  until use.

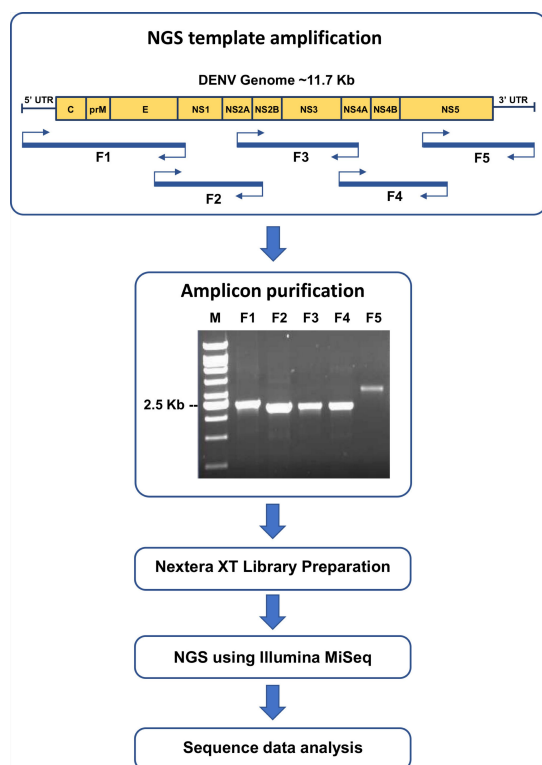
**TABLE 1** Representative DENV isolates used in NGS method development.

Sample ID	DENV Serotype	Isolation year	Location	GenBank accession no.	Total NGS reads	# Mapped to RefSeq (%)	Mean Depth (in X)	% RefSeq coverage
JMB-059	DENV-1	2015	Jambi	MH823207	1,072,312	885,756 (82.6)	11,879	100.0
JMB-010	DENV-2	2014	Jambi	MH823208	758,596	348,993 (46.0)	4,633	100.0
SMD-031	DENV-3	2016	Samarinda	MH823209	631,332	628,489 (99.5)	8,405	100.0
JMB-006	DENV-4	2014	Jambi	MH823210	394,654	391,828 (99.3)	5,292	100.0

Five overlapping 2–3 kb fragments were generated using primer sets for each DENV serotype (Ong et al. 2008; Christenbury et al. 2010; Sasmono et al. 2015) using *Pfu* Turbo DNA Polymerase enzyme kit (Agilent-Thermo Fisher Scientific) and generic PCR cycling condition of pre-denaturing at 95°C for 2 min; 40 cycles of denaturing at 95°C for 30 s, annealing at 55°C for 1 min, and extension at 72°C for 4 min 30 s; followed by final extension for 10 min and storage at 4°C. Primers used in RT-PCR reactions are listed in Table 2. The expected size of PCR fragments was sliced and purified from 0.8% agarose gel using QIAquick Gel Extraction kit (Qiagen), according to protocol from the manufacturer. The purified PCR amplicons' concentration were measured using Qubit DNA BR Assay kit (Thermo Fisher Scientific) and adjusted to 0.2 ng/μL using EB buffer (Qiagen). All five normalized PCR amplicons were pooled for each sample and stored at -20°C until use. The schematic representation of the steps performed is outlined in Figure 1.

#### 2.4. NGS library preparation and deep sequencing

The PCR amplicons were processed for a deep sequencing approach from an enriched template, as modified from protocol described elsewhere Aw et al. (2014). The NGS library preparation was prepared using Nextera XT DNA Library Prep Kit kit (Illumina, San Diego, CA) with the amounts of starting material per sample adjusted to 1 ng of pooled amplicons. Strict protocol was observed as recommended by the manufacturer.



**FIGURE 1** Schematic figure of steps performed in DENV whole-genome sequencing method using an enriched template and next-generation sequencing (NGS) platform.

The Nextera transposome was used to tagment DNA, in which the DNA was fragmented and then tagged with adapter sequences in a single step. Tagmented DNA was then subjected to an amplification step using a limited-cycle PCR program and Nextera XT index kit (24 indexes). The PCR reactions added the Index 1 (i7), Index 2 (i5), and full adapter sequences to the tagmented DNA from the previous step. The library DNA was purified and short library fragments were removed using Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN), utilizing the DNA-binding and magnetic properties of the beads.

Purified library was then measured for its concentration using Qubit high sensitivity dsDNA kit (ThermoFisher Scientific) and further quantified and validated using KAPA Library Quantification kit (KAPA Biosystem-Roche), according to protocol described by the manufacturer. The normalized library yield after amplification was expected to be 10–15 nM, or higher. The library from each sample was then pooled together, further diluted and loaded into the 500-cycle MiSeq Reagent Nano Kit v2 cartridge and run on Illumina MiSeq instrument.

#### 2.5. Sequencing reads processing and bioinformatic analysis

The sequence reads were retrieved as Fastq files. Individual sequence reads file quality was checked using FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and multiple quality control check for the generation of aggregated quality control reports were done using MultiQC software (Ewels et al. 2016), both available in the Galaxy public server at <https://usegalaxy.org> (Afgan et al. 2016). The reads paired-end processing and mapping to reference sequence were done using Geneious v.11.1.5 software and default settings (Kearse et al. 2012), allowing robust bioinformatics analyses in a single software format and graphical interface. The paired-end reads were joined and assembled using the map to reference algorithm according to each DENV-serotype reference sequence (Table 2). The Geneious mapper with highest sensitivity was used to assemble and map reads to the reference sequence with GenBank accession number NC\_001477 (GCF\_000862125.1), NC\_001474 (GCF\_000871845.1), NC\_001475 (GCF\_000866625.1), and NC\_002640 (GCF\_000865065.1) for DENV-1, -2, -3, and -4, respectively. Sequence coverage and depth were visualized as an interactive desktop display in the assembly report.

#### 2.6. DENV genetic distance and phylogenetic analyses

Contigs were generated and manual inspections were performed to ensure the correctness of final genome sequences. Sequence contigs were then queried to on-line BLASTn platform available at <https://blast.ncbi.nlm.nih.gov> (Zhang et al. 2000) for sequence identification and similarity analysis to all DENV genomes in public database using nucleotide collection database (nt). A total of 100 taxa hits was selected as the basis of phylogenetic tree reconstruction based on genetic distance and Neigh-

**TABLE 2** Primer sequences used in DENV whole-genome amplification using enriched NGS method, extracted from Christenbury et al. (2010) and Sasmono et al. (2015). Primer names printed in bold are the antisense primers used in RT reactions.

Serotype	PCR Fragment	Primer name	Amplicon size (Kb)	Primer sequences (5'-3')	Genome position (sense strand)
DENV-1	F1	d1s22	2.5	AGTTGTTAGTCTACGTGGAC	1-20
		<b>d1a17</b>		CCAATGGCYGCTGAYAGTCT	2540-2559
	F2	d1s6	2.3	GGYTCTATAGGAGGRGTGTTAC	2201-2223
		<b>d1a13</b>		TTCCACTTCYGGAGGGCT	4544-4561
	F3	d1s10	2.3	RGCYGGSCCACTAATAGCT	4213-4231
		<b>d1a9</b>		CCAGTYARCACAGCTATCAAAGC	6551-6573
	F4	d1s14	2.3	ATGGRGAAAGGAACAACCAG	6216-6235
		<b>d1a5</b>		AGRACACGTAACGTTCTWCCTTC	8045-8067
	F5	d1s18	2.5	CCACYCATGAAATGTAYGGGT	8211-8232
		<b>d1a5B</b>		TTTGTCGGTCTGGGGGGGTATAGAACCTGTTGATTCAACRGC	10716-10735*
DENV-2	F1	d2s23	2.4	AGTWGTTAGTCTACGTGGAC	1-20
		<b>d2a18</b>		CCACTGCCACATTTTCAGTTC	2455-2474
	F2	d2s5	2.4	GGTGACACAGCCTGGGATTT	2182-2201
		<b>d2a14</b>		GCCGTGATTGGTATTGATACAGGA	4461-4484
	F3	d2s9	2.4	GCATTTTRGCCAGTTCTCTCCTA	4175-4197
		<b>d2a10</b>		TACGCCCTTCCRCTGCTTCA	6477-6497
	F4	d2s13	2.4	GCAGACAGAAGGTGGTGTTTT	6193-6213
		<b>d2a6</b>		CATGGTAWGCCAYGTTTTGT	8468-8488
	F5	d2s16	3.0	CAGGAAGTGATAGAACCTTAGCA	7669-7692
		<b>d2a5B</b>		TTTGTCGGTCTGGGGGGGTATAGAACCTGTTGATTCAACAG	10704-10723*
DENV-3	F1	d3s23	2.3	AGTTGTTAGTCTACGTGGAC	1 to 20
		<b>d3a18</b>		GATTCCTATCGCAATGCATG	2361-2380
	F2	d3s5	2.3	TGAACCTCCTTTGGGGAA	2035-2053
		<b>d3a14</b>		ACTGTGATCATTARTTGTGGGA	4334-4356
	F3	d3s9	2.3	GAAAACAGATTGGCTCCCAA	4030-4049
		<b>d3a10</b>		GCYGCAAARTCCTTGAATTCCT	6339-6360
	F4	d3s13	2.3	CCAGCTCTCTTTGAACCAGAAA	6032-6053
		<b>d3a6</b>		GCATTRACATGTCGRGTTC	8342-8361
	F5	d3s17	2.7	CAACAGTGAAGAAAGCAGAAC	8025-8046
		<b>d3a5B</b>		TTTGTCGGTCTGGGGGGGTATAGAACCTGTTGATTCAACAGC	10688-10707*
DENV-4	F1	d4f1	2.6	AGTTGTTAGTCTGTGTGGACCGAC	1 to 24
		<b>d4r18</b>		GGGCATTYAATATTGCAGACGCTA	2548-2571
	F2	d4f5	3.0	CTCCGTGTAAGTCCCCATAGAGA	1953-1930
		<b>d4r13</b>		TCYTGTDDATGATRGGAGARCCAGA	4926-4950
	F3	d4f11	2.5	CCTMGCYATAGAACCAGGAAAAAATC	4814-4839
		<b>d4r8</b>		ACYTGCCCTAATTGCTTTTCAA	7308-7330
	F4	d4f16	2.5	AGRACAGCTGCTGGGATCATGA	7224-7245
		<b>d4r3</b>		TGTGRAARTGGTGGGAGCAAAA	9708-9687
	F5	d4f21	1.0	GAAAGACATYCCGCAGTGGGAA	9629-9650
		<b>d4a5B</b>		TTTGTCGGTCTGGGGGGGTATAGAACCTGTTGGATCAACAAC	10643-10666*

\*Additional flanking sequences added to primer sequences..

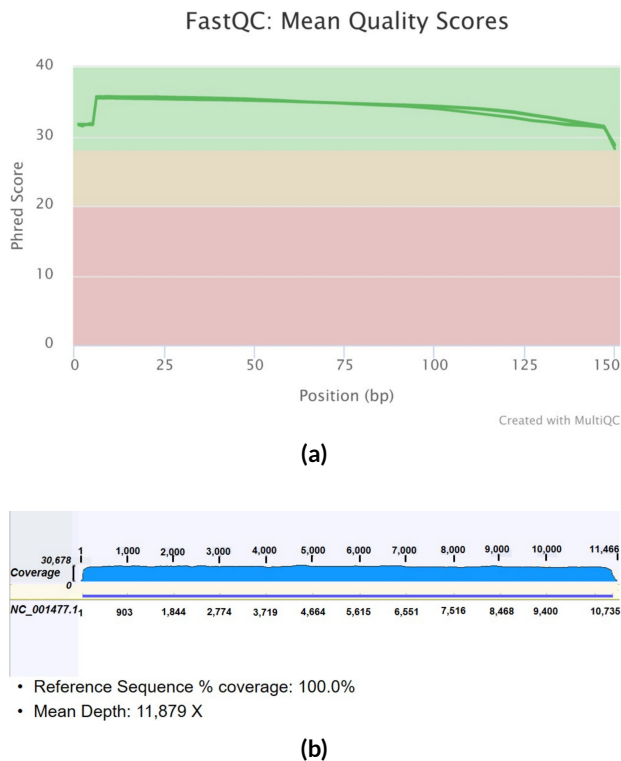


bor Joining algorithm. Tree files were downloaded and edited in FigTree v.1.4.3. The DENV genotypes classification in each serotype was based on classifications by Goncalvez et al. (2002), Twiddy et al. (2002), Lanciotti et al. (1994), and Lanciotti et al. (1997) for DENV-1, -2, -3, and -4, respectively.

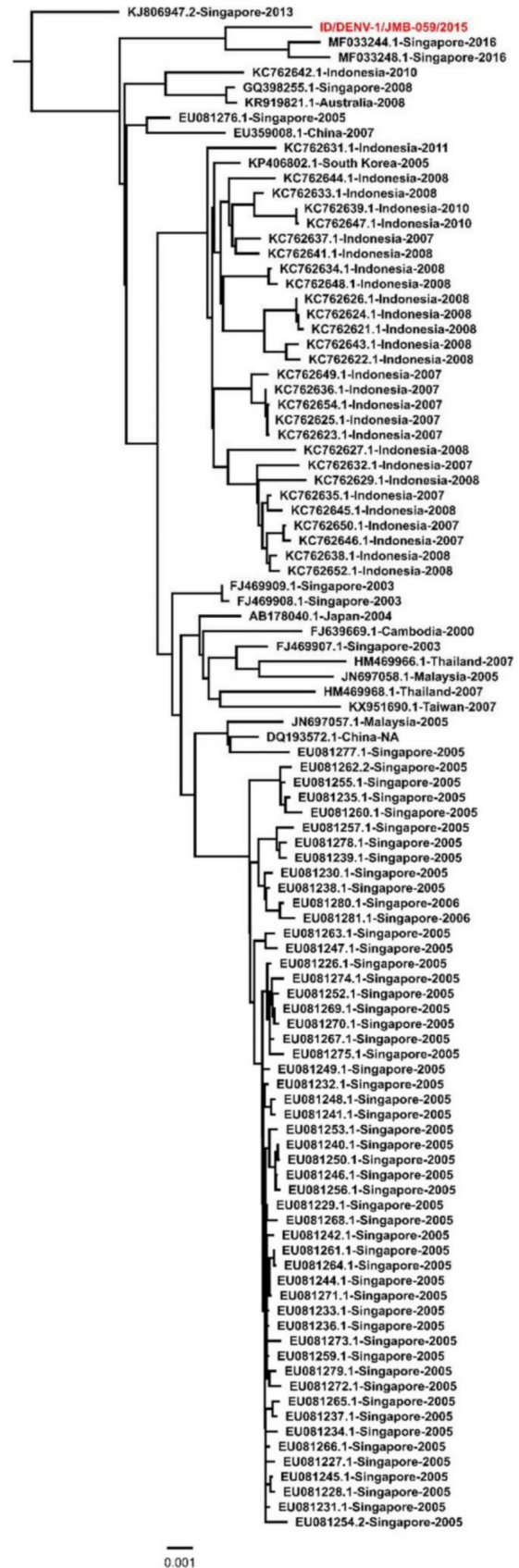
### 3. Results

#### 3.1. NGS run of enriched template using Illumina MiSeq

The amplification approach yielded five overlapping DNA fragments covering the whole genome of DENV (Figure 1, shown as an example for DENV-2, please refer to Table 2 for fragment sizes of other serotypes). Using the starting amount of 1 ng per sample pool, the enriched template was loaded into the MiSeq NGS instrumentation. The generated sequence reads were sorted by their respective library adapters and compiled into the respective Fastq files. Quality controls on reads as aggregated using MultiQC showed optimum quality for all four DENV isolates, as shown by Phred quality value score of higher than 30 across each base position (green lines in green zone of the histogram) (Figure 2a). The mapped reads to DENV reference sequences as computed using Geneious mapper also showed a very high genome coverage (reaching 100%) and optimum high depth for each isolate (Figure 2b, shown only for DENV-1 as figure representative



**FIGURE 2** Snapshots of quality control parameters assessed on the generated sequence reads for 4 DENV serotypes' representatives. (a) Sequence quality histogram depicted as the mean quality value across each base position in the read. (b) Sequence reads coverage and mean depth view snapshot generated using Geneious map-to-reference algorithm.



**FIGURE 3** Rapid DENV-1 phylogenetic tree based on genetic distance as inferred using Neighbor joining (NJ) algorithm and implemented in BLASTn phylogenetic tree online module. The Indonesian DENV-1 isolate is written in red font and classified into the Genotype I of DENV-1.

and Table 1). Lower depth was observed for 5' and 3'-ends of the genome. As a comparative analysis, we also performed reads alignment using BWA-MEM alignment and SPAdes assembler within the Linux-based environment. The results from Geneious mapper were comparable to the results generated in Linux (data not shown).

**3.2. Sequence contig assembly, curation, and deposition to public database**

Assembled sequences were exported as contigs of 10,728; 10,723; 10,707; and 10,648 bp for DENV-1, -2, -3, and -4, respectively. The sequences were curated for ambiguity, deposited to GenBank and granted accession numbers MH823207 - MH823210.

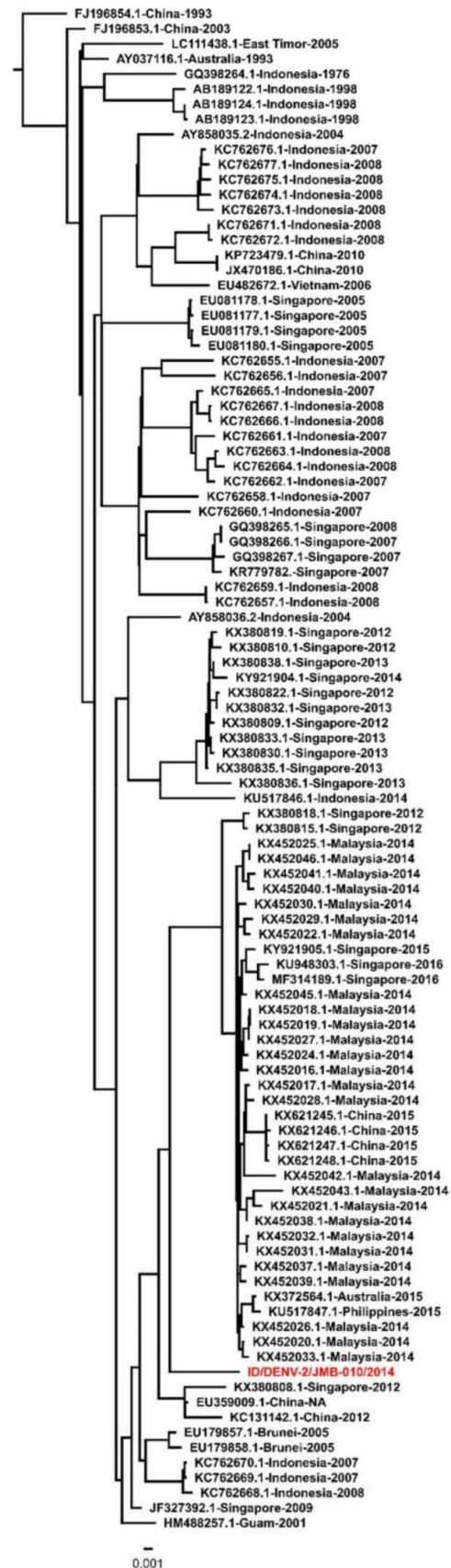
**3.3. DENV genetic distance and phylogenetic analyses**

Using the generated nearly-complete whole-genome sequences, BLASTn queries were employed to determine DENV sequence identity and similarity with other DENV complete genome sequences available in GenBank, according to its serotype. One hundred hits were downloaded as sequences with the highest sequence identity and maximum score with lowest expect (E) value in BLASTn. The Neighbor Joining phylogenetic analysis using these 100 reference sequences revealed the close-relatedness of Indonesia isolates with other strains from Indonesia and from the surrounding regions. The DENV-1 Indonesia isolate was closely related to sequences from Singapore (Figure 3). Analysis of generated whole-genome sequences of DENV-2 Indonesia isolate revealed the close genetic distance with strains from China and Singapore (Figure 4). The close genetic distance was observed between Indonesia DENV-3 isolate and strains from Jakarta and Makassar, Indonesia; Singapore; and Cairns, Australia. The Indonesia DENV-4 isolate shared close genetic relationship with strains from China and The Philippines (Figure 5).

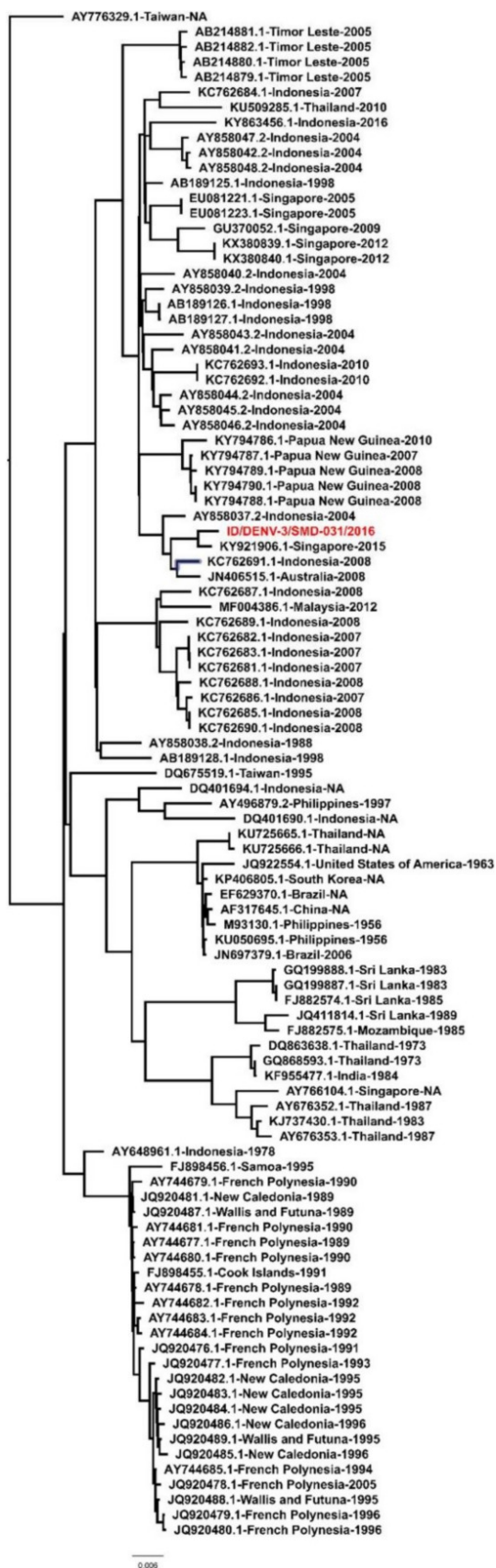
**4. Discussion**

Indonesia is hyperendemic for dengue in which all four DENV serotypes circulate. Although a common disease in Indonesia, unfortunately until now there has not been much DENV genome data available from Indonesia. This is one of the challenges that are faced by scientists in Indonesia and around the world in obtaining DENV genetic information to be used for both basic and applied research. A large gap on the genome data between Indonesia and neighboring countries needs to be filled.

Complete understanding of the evolution and epidemiology of DENV, particularly with respect to the etiology of severe disease, will require DENV genomic studies and the comparative analysis of complete genome sequences (Holmes and Twiddy 2003). DENV genome data can be linked with epidemiological, clinical, and immunological data. The data can also be used to track the transmission of virulent DENV and eventually predict possible epidemics. Furthermore, DENV genomic data may also benefit the antiviral drug development, protein/antigen immunogenicity



**FIGURE 4** Rapid DENV-2 phylogenetic tree based on genetic distance as inferred using Neighbor joining (NJ) algorithm and implemented in BLASTn phylogenetic tree online module. The Indonesia DENV-2 isolate is written in red font and classified into the Cosmopolitan Genotype of DENV-2.



**FIGURE 5** Rapid DENV-3 phylogenetic tree based on genetic distance as inferred using Neighbor joining (NJ) algorithm and implemented in BLASTn phylogenetic tree online module. The Indonesian DENV-3 isolate is written in red font and classified into the Genotype I of DENV-3.

studies, dengue vaccine development and vaccine introduction. In the vaccine field, vaccine efficacy could depend on the similarity of the dengue virus genotypes used in vaccine production and those circulating in the area of vaccine introduction (Usme-Ciro et al. 2014). In the other hand, dengue diagnostics sensitivity was shown to be influenced by the geographical regions (Guzman et al. 2010; Aryati et al. 2013). This has been linked with the distinct distribution of DENV in different regions, one of the characteristics of DENV genotypes. Thus, information on the DENV genetic diversity and identity in a particular region will be important to dengue diagnostics development and application.

Because of its importance, there is now a concerted effort to obtain the complete genome sequences of DENV isolates from all over the world, an effort directed by the Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard University (Holmes 2009). In Indonesia, efforts to obtain DENV genome data have been initiated. Using the conventional Sanger capillary sequencing technique, we sequenced the whole genomes of Makassar DENV to understand their genetic diversity. Using genomic data, we determined the DENV genotypes and revealed the Makassar DENV genotype fitness which may underlie the better virus transmission in the population (Sasmono et al. 2015). Using the same approach, the comparative analysis of Surabaya DENV-1 genomes revealed the virus evolutionary selection patterns in which the majority of the DENV-1 codons were under strong purifying selection, while seven codon sites were under positive selection (Yohan et al. 2018).

Among the available methods for NGS, the Nextera XT library preparation kit was designed to sequence small genomes including PCR amplicons greater than 300 bp, plasmids, microbial genomes, concatenated amplicons, and double-stranded cDNA, and thus suitable for DENV genome sequencing. The other advantage of this kit is the low DNA quantity requirement, as low as 1 ng for each sample. In regards to quality control of results, we observed that Phred scores of the four isolates' sequencing reads per base pair position were above 30, shown by the green zone in the Figure 2a. The coverage and depth of the reads were also optimal to ensure a reliable contigs generation. However, the use of transposons enzymatic cleavage of double-stranded DNA may reduce the effectiveness of the 5' and 3' end region sequencing of the genome, as shown by a relatively low sequence depth on those regions (Figure 2b). To generate complete genome sequences, additional conventional capillary sequencing may solve the problem of low reads in those genome parts. Nevertheless, the NGS method has proven to be more efficient in time and laboratory cost compared to Sanger sequencing in term of the multiplication effect where in one MiSeq run, 24 different isolates can be sequenced simultaneously with optimum depth. The NGS run using the enriched amplicon template and NexTera library preparation kit was proven to yield good result as shown by good quality control value and up to 100% of nearly complete genome coverage.

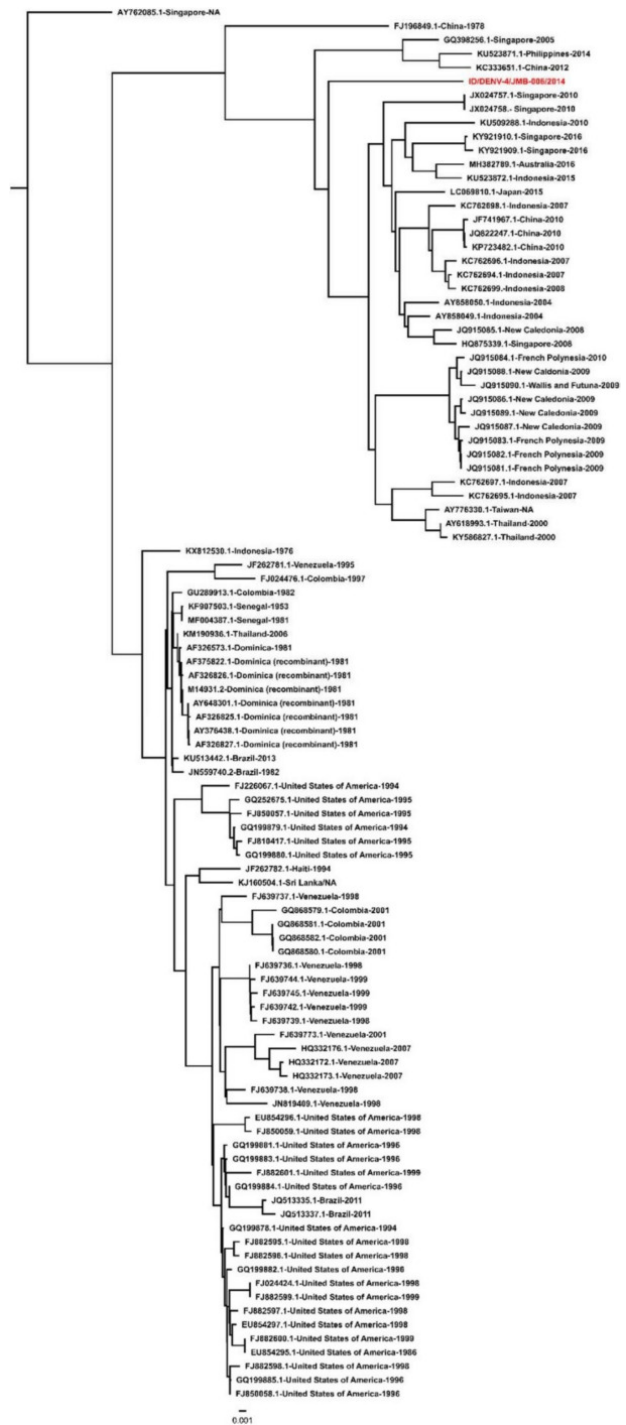


The nearly complete genomes of Indonesia DENV were submitted to GenBank and made accessible to public. Using fast online algorithm at the BLASTn web page, the sequence identity was confirmed to be DENV according to its own serotype. One of the benefit of the genetic distance tree analysis was to observe the genetic close-relatedness of our sequence compared to all available genomes in public database. From this point, we can proceed to the more in-depth bioinformatic analyses, such as selection pressure and virus evolutionary analyses (Yohan et al. 2018). The Indonesian DENV isolates were classified into the Genotype I, Cosmopolitan Genotype, Genotype I, and Genotype II for DENV-1, -2, -3, and -4, respectively, as observed from the genotyping approach using Envelope gene (Haryanto et al. 2016). Using the nearly complete genome, the DENV genomes sequenced in this study showed close-relatedness with strains from local Indonesia, neighboring countries, and countries in the region. The DENV-1 isolate showed close-relatedness to sequence from Singapore isolated in 2016 (Koo et al. 2018). The DENV-2 isolate was closely related to strains from Singapore and China and furthermore located in the basal position of a clade consisted of strains from Singapore and Malaysia (isolated in Johor in 2014, unpublished). The DENV-3 isolate also showed close genetic relationship with strains from cities in Indonesia, Singapore, and Australia. Meanwhile, the DENV-4 isolate showed close relatedness with strains from China and Philippines. Together, the findings confirms our previous data of complete genomes from Makassar (Sasmono et al. 2015) where Indonesian isolates showed close genetic relationships with strains from other cities in Indonesia, Singapore, and other regions, such as China and Australia. Moreover, results from this study may confirm our previous findings that highlighted the endemicity of Indonesian DENV isolates with the possibility of becoming the potential source of DENV importation in other countries regionally (Lestari et al. 2017; Megawati et al. 2017; Masyeni et al. 2018). These findings may support the use of rapid analysis using online tools in preliminary analysis of DENV whole genome sequences.

One of the bottlenecks of NGS often encountered by researcher is the downstream processing of the generated sequence reads which requires many computation and knowledge in bioinformatics. Here, we present a simplified method to analyze NGS output using Windows-based or online tools/software offering relatively user-friendly graphical interface. With this approach, scientists can perform basic and simple analyses of NGS output before continuing to the more in-depth bioinformatic processing.

### 5. Conclusions

The NGS method using enriched template approach is proven to be powerful and robust to generate DENV whole-genome sequences from Indonesian isolates and aided by the use of simplified bioinformatic analysis methods. The methods presented in this study will be useful to generate DENV genomes and build the genomic database



**FIGURE 6** Rapid DENV-4 phylogenetic tree based on genetic distance as inferred using Neighbor joining (NJ) algorithm and implemented in BLASTn phylogenetic tree online module. The Indonesian DENV-4 isolate is written in red font and classified into the Genotype II of DENV-4.

needed to study and combat this virus and contribute to the better understanding of the dengue disease in Indonesia.

### Acknowledgments

This study was funded by the DIPI-MRC Joint Call 2017, LPDP, Ministry of Finance of the Republic of Indonesia in collaboration with the UK Medical Research Council.



## Authors' contributions

BY, HT, KSAM, RTS designed the study. YWBP, AMA, IAH, SH, KSAM contributed in sample collection and reagents. BY, RD, RFH, FAY, DD carried out the laboratory works. BY, HT, DD analyzed the data. BY, RTS wrote the manuscript. All authors read and approved the final version of the manuscript.

## Competing interests

RTS is on the editorial board of the *Indonesian Journal of Biotechnology*, and was recused from this article's review and decision. The authors declare no other competing interests.

## References

- Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44(W1):W3–W10. doi:10.1093/nar/gkw343.
- Aryati A, Trimarsanto H, Yohan B, Wardhani P, Fahri S, Sasmono RT. 2013. Performance of commercial dengue NS1 ELISA and molecular analysis of NS1 gene of dengue viruses obtained during surveillance in Indonesia. *BMC Infect Dis.* 13(1):611. doi:10.1186/1471-2334-13-611. 00000.
- Aw PPK, de Sessions PF, Wilm A, Hoang LT, Nagaranjan N, Sessions OM, Hibberd ML. 2014. Next-generation whole genome sequencing of dengue virus. *Methods Mol Biol.* 1138:175–195. doi:10.1007/978-1-4939-0348-1\_12. 00008.
- Capobianchi MR, Giombini E, Rozera G. 2013. Next-generation sequencing technology in clinical virology. *Clin Microbiol Infect.* 19(1):15–22. doi:10.1111/1469-0691.12056.
- Christenbury JG, Aw PPK, Ong SH, Schreiber MJ, Chow A, Gubler DJ, Vasudevan SG, Ooi EE, Hibberd ML. 2010. A method for full genome sequencing of all four serotypes of the dengue virus. *J Virol Methods.* 169(1):202–206. doi:10.1016/j.jviromet.2010.06.013. 00012 PMID: 20600330.
- Costa RL, Voloch CM, Schrago CG. 2012. Comparative evolutionary epidemiology of dengue virus serotypes. *Infect Genet Evol.* 12(2):309–314. doi:10.1016/j.meegid.2011.12.011. 00035.
- Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 32(19):3047–3048. doi:10.1093/bioinformatics/btw354.
- Fahri S, Yohan B, Trimarsanto H, Sayono S, Hadisaputro S, Dharmana E, Syafruddin D, Sasmono RT. 2013. Molecular surveillance of dengue in Semarang, Indonesia revealed the circulation of an old genotype of dengue virus serotype-1. *PLoS Negl Trop Dis.* 7(8):e2354. doi:10.1371/journal.pntd.0002354.
- Goncalvez A, Escalante A, Pujol F, Ludert J, Tovar D, Salas R, Liprandi F. 2002. Diversity and evolution of the envelope gene of dengue virus type 1. *Virology.* 303(1):110–119. doi:10.1006/viro.2002.1686.
- Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, Hunsperger E, Kroeger A, Margolis HS, Martínez E, et al. 2010. Dengue: a continuing global threat. *Nat Rev Microbiol.* 8(12 Suppl):S7–16. doi:10.1038/nrmicro2460.
- Haryanto S, Hayati RF, Yohan B, Sijabat L, Sihite IF, Fahri S, Meutiawati F, Halim JAN, Halim SN, Soebandrio A, Sasmono RT. 2016. The molecular and clinical features of dengue during outbreak in Jambi, Indonesia in 2015. *Pathog Glob Health.* 110(3):119–129. doi:10.1080/20477724.2016.1184864.
- Holmes E, Burch S. 2000. The causes and consequences of genetic variation in dengue virus. *Trends Microbiol.* 8(2):74–77. doi:10.1016/s0966-842x(99)01669-8.
- Holmes E, Twiddy S. 2003. The origin, emergence and evolutionary genetics of dengue virus. *Infect Genet Evol.* 3(1):19–28. doi:10.1016/s1567-1348(03)00004-2.
- Holmes EC. 2009. RNA virus genomics: a world of possibilities. *J Clin Invest.* 119(9):2488–2495. doi:10.1172/JCI38050.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28(12):1647–1649. doi:10.1093/bioinformatics/bts199.
- Koo C, Ping T Wei, Helen X, Janet O, Rajarethinam J, Lai, Yee Ling, Ng, Lee-Ching, Chanditha H Hapuarachchige. 2018. Highly selective transmission success of dengue virus type 1 lineages in a dynamic virus population: An evolutionary and fitness perspective. *iScience.* 6:38–51. doi:10.1016/j.isci.2018.07.008.
- Lambrechts L, Scott TW, Gubler DJ. 2010. Consequences of the expanding global distribution of *Aedes albopictus* for dengue virus transmission. *PLoS Negl Trop Dis.* 4(5):e646. doi:10.1371/journal.pntd.0000646. 00193.
- Lanciotti RS, Gubler DJ, Trent DW. 1997. Molecular evolution and phylogeny of dengue-4 viruses. *J Gen Virol.* 78(9):2279–2284. doi:10.1099/0022-1317-78-9-2279.
- Lanciotti RS, Lewis JG, Gubler DJ, Trent DW. 1994. Molecular evolution and epidemiology of dengue-3 viruses. *J Gen Virol.* 75 (Pt 1)(1):65–75. doi:10.1099/0022-1317-75-1-65.
- Leitmeyer K, Vaughn D, Watts D, Salas R, Villalobos I, de C, Ramos C, Rico-Hesse R. 1999. Dengue virus structural differences that correlate with pathogenesis. *J Virol.* 73(6):4738–47.
- Lestari CSW, Yohan B, Yunita A, Meutiawati F, Hayati RF, Trimarsanto H, Sasmono RT. 2017. Phylogenetic

- and evolutionary analyses of dengue viruses isolated in Jakarta, Indonesia. *Virus Genes*. 53(6):778–788. doi:10.1007/s11262-017-1474-7.
- Martina BEE, Koraka P, Osterhaus ADME. 2009. Dengue virus pathogenesis: an integrated view. *Clin Microbiol Rev*. 22(4):564–581. doi:10.1128/CMR.00035-09.
- Masyeni S, Yohan B, Somia IKA, Myint KSA, Sasmono RT. 2018. Dengue infection in international travellers visiting Bali, Indonesia. *J Travel Med*. 25(1):1–7. doi:10.1093/jtm/tay061.
- Megawati D, Masyeni S, Yohan B, Lestari A, Hayati RF, Meutiawati F, Suryana K, Widarsa T, Budiayasa DG, et al. 2017. Dengue in Bali: clinical characteristics and genetic diversity of circulating dengue viruses. *PLoS Negl Trop Dis*. 11(5):e0005483. doi:10.1371/journal.pntd.0005483.
- Ong S, Yip J, Chen Y, Liu W, Harun S, Lystiyaningsih E, Heriyanto B, Beckett C, Mitchell W, Hibberd M, et al. 2008. Periodic re-emergence of endemic strains with strong epidemic potential—a proposed explanation for the 2004 Indonesian dengue epidemic. *Infect Genet Evol*. 8(2):191–204. doi:10.1016/j.meegid.2007.12.005.
- Organization WH. 2012. *Global strategy for dengue prevention and control, 2012-2020*. Geneva: WHO Press.
- Quiñones-Mateu ME, Avila S, Reyes-Teran G, Martinez MA. 2014. Deep sequencing: becoming a critical tool in clinical virology. *J Clin Virol*. 61(1):9–19. doi:10.1016/j.jcv.2014.06.013.
- Sasmono RT, Aryati A, Wardhani P, Yohan B, Trimarsanto H, Fahri S, Setianingsih TY, Meutiawati F. 2014. Performance of Simplexa dengue molecular assay compared to conventional and SYBR green RT-PCR for detection of dengue infection in Indonesia. *PLoS ONE*. 9(8):e103815. doi:10.1371/journal.pone.0103815.
- Sasmono RT, Wahid I, Trimarsanto H, Yohan B, Wahyuni S, Hertanto M, Yusuf I, Mubin H, Ganda IJ, Latief R, Bifani PJ, Shi PY, Schreiber MJ. 2015. Genomic analysis and growth characteristic of dengue viruses from Makassar, Indonesia. *Infect Genet Evol*. 32:165–177. doi:10.1016/j.meegid.2015.03.006.
- Twiddy SS, Farrar JJ, Vinh Chau N, Wills B, Gould EA, Gritsun T, Lloyd G, Holmes EC. 2002. Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology*. 298(1):63–72. doi:10.1006/viro.2002.1447.
- Usme-Ciro JA, Méndez JA, Laiton KD, Páez A. 2014. The relevance of dengue virus genotypes surveillance at country level before vaccine approval. *Hum Vaccin Immunother*. 10(9):2674–2678. doi:10.4161/hv.29563.
- van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends Genet*. 30(9):418–426. doi:10.1016/j.tig.2014.07.001.
- Yohan B, Wardhani P, Trimarsanto H, Aryati A, Sasmono RT. 2018. Genomic analysis of dengue virus serotype 1 (DENV-1) genotypes from Surabaya, Indonesia. *Virus Genes*. 54(3):461–465. doi:10.1007/s11262-018-1558-z.
- Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol*. 7(1-2):203–214. doi:10.1089/10665270050081478.