




RESEARCH ARTICLE

REVISED Complete assembly of a dengue virus type 3 genome from a recent genotype III clade by metagenomic sequencing of serum [version 2; referees: 2 approved]

Mary Dias^{1*}, Chitra Pattabiraman ^{2*}, Shilpa Siddappa³, Malali Gowda⁴, Anita Shet⁵, Derek Smith^{6,7}, Barbara Muehleemann^{6,7}, Krishnapriya Tamma⁸, Tom Solomon⁹, Terry Jones^{6,7}, Sudhir Krishna¹⁰

¹St. John's Medical College and Hospital, Bangalore, 560034, India

²National Institute of Mental Health and Neurosciences, India, Bangalore, 560029, India

³Centre for Cellular and Molecular Platforms, Bangalore, 560065, India

⁴Trans-Disciplinary University, Foundation for Revitalization of Local Health Traditions, Bangalore, 560064, India

⁵Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, 21205, USA

⁶Center for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK

⁷World Health Organization Collaborating Center for Modeling, Evolution, and Control of Emerging Infectious Diseases, Cambridge, CB2 3EJ, UK

⁸Indian Institute of Science, Bangalore, 560012, India

⁹Institute of Infection and Global Health, and National Institute for Health Research, Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, Liverpool, L69 7BE, UK

¹⁰National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, 560065, India

* Equal contributors

v2 First published: 23 Apr 2018, 3:44 (<https://doi.org/10.12688/wellcomeopenres.14438.1>)

Latest published: 10 Jan 2019, 3:44 (<https://doi.org/10.12688/wellcomeopenres.14438.2>)

Abstract

Background: Mosquito-borne flaviviruses, such as dengue and Japanese encephalitis virus (JEV), cause life-threatening diseases, particularly in the tropics.

Methods: Here we performed unbiased metagenomic sequencing of RNA extracted from the serum of four patients and the plasma of one patient, all hospitalized at a tertiary care centre in South India with severe or prolonged febrile illness, together with the serum from one healthy control, in 2014.

Results: We identified and assembled a complete dengue virus type 3 sequence from a case of severe dengue fever. We also identified a small number of JEV sequences in the serum of two adults with febrile illness, including one with severe dengue. Phylogenetic analysis revealed that the dengue sequence belonged to genotype III. It has an estimated divergence time of 13.86 years from the most highly related Indian strains. In total, 11 amino acid substitutions were predicted for this strain in the antigenic envelope protein, when compared to the parent strain used for development of the first commercial dengue vaccine.

Conclusions: We demonstrate that both genome assembly and detection of a low number of viral sequences are possible through the unbiased sequencing

Open Peer Review

Referee Status:  

Invited Referees

1 2

REVISED

version 2

published
10 Jan 2019

version 1

published
23 Apr 2018

 report

 report

- 1 **Shijulal Nelson-Sathi**, Rajiv Gandhi
Centre for Biotechnology-BIC Campus,
India
- Easwaran Sreekumar**, Rajiv Gandhi
Centre for Biotechnology (RGCB), India

of clinical material. These methods may help ascertain causal agents for febrile illnesses with no known cause.

Keywords

DENV3, metagenomics, febrile illness



This article is included in the [Wellcome Trust/DBT India Alliance](#) gateway.

2 **John A Lednicky**, University of Florida, USA

Discuss this article

[Comments](#) (0)

Corresponding author: Chitra Pattabiraman (pattabiraman.chitra@gmail.com)

Author roles: **Dias M:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Pattabiraman C:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Siddappa S:** Investigation, Methodology; **Gowda M:** Methodology; **Shet A:** Methodology, Resources; **Smith D:** Funding Acquisition, Methodology, Software, Supervision, Visualization; **Muehleman B:** Methodology, Software, Visualization, Writing – Original Draft Preparation; **Tamma K:** Software, Writing – Original Draft Preparation; **Solomon T:** Conceptualization, Funding Acquisition, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Jones T:** Methodology, Software, Visualization, Writing – Original Draft Preparation; **Krishna S:** Conceptualization, Funding Acquisition, Supervision, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This study was supported by the Department of Biotechnology, Glue Grant to SK. PC was supported by the Department of Biotechnology Research Associate fellowship, Royal Society, SERB Newton International fellowship and the India Alliance (DBT-Wellcome Trust) Early Career Fellowship. TS is supported by the NIH Health Protection Research Unit in Emerging and Zoonotic Infections and the European Union's Horizon 2020 research and innovation program ZikaPLAN (Preparedness Latin America Network), grant agreement No. 734584. DS, BM, and TJ were supported by the European Union FP7 programme ANTIGONE (grant agreement No. 278976). The funders had no role in the in any steps of experimental design, data collection or analysis or the decision to publish.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Dias M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Dias M, Pattabiraman C, Siddappa S *et al.* **Complete assembly of a dengue virus type 3 genome from a recent genotype III clade by metagenomic sequencing of serum [version 2; referees: 2 approved]** Wellcome Open Research 2019, 3:44 (<https://doi.org/10.12688/wellcomeopenres.14438.2>)

First published: 23 Apr 2018, 3:44 (<https://doi.org/10.12688/wellcomeopenres.14438.1>)

REVISED Amendments from Version 1

Revisions have been made to the main manuscript and to [Figure 3](#) based on comments from the reviewers. Modifications have been made in – i) Page 7, discussion, paragraph 1 ii) page 3 – introduction, paragraph 3 iii) in the legend and results section for [Figure 3](#).

See referee reports

Introduction

Acute undifferentiated febrile illness refers to a sudden onset of high fever without localized organ-specific clinical features¹. Although the majority of patients recover over a few days, some can develop severe illnesses, resulting in high morbidity and even death in many parts of the world. Among the many causes of febrile illness, some of the most important across Asia are mosquito-borne viruses such as dengue virus^{1–6}. In addition, novel agents associated with acute febrile illness continue to be discovered^{7–9}.

Current molecular diagnostic techniques, such as polymerase chain reaction, are pathogen-specific and therefore pose limitations, as they may fail to detect co-infections and novel agents not commonly associated with the disease syndrome¹⁰. The unbiased metagenomic sequencing of clinical material from patients with acute fever can overcome these limitations^{3,11}.

Mosquito-borne viruses of the family *Flaviviridae*, which include dengue virus and Japanese encephalitis virus (JEV) are known to co-circulate in India and other parts of Asia¹². Dengue viruses are a major cause of acute febrile illness in Asia, with recurrent outbreaks having occurred¹³. JEV, on the other hand, is better known as a cause of acute encephalitis¹⁴. Although JEV has been noted as an agent that causes acute fever in Southeast Asia, it is not routinely tested as a cause of fevers in India^{5,6}. There are four distinct serotypes of dengue viruses (DENV1–DENV4), with their small RNA genomes (approximately 10.8 kbp) making them amenable for characterization by deep sequencing of infected mosquitoes or clinical material from infected individuals¹⁵. Sequencing dengue genomes is important for tracking virus evolution, given that they frequently mutate^{15,16}. Outbreaks of severe dengue disease associated with serotype switches or the introduction of a novel strain into the population have been reported from several different countries, including Sri Lanka, Pakistan and Singapore^{17–22}. Recent analysis suggests an influenza-virus-like pattern for dengue virus evolution, where strain-specific differences underlie antibody neutralization²³. Pre-existing antibodies to circulating dengue strains can therefore contribute to disease severity by inadequate neutralization of the virus or by antibody-mediated enhancement, which facilitates virus infection^{24–28}. This is supported by *in vitro* studies, which found that changes to the envelope (E) protein of DENV3 were sufficient to alter antibody binding²⁶. Multiple dengue vaccines are currently in various stages of development, and a tetravalent vaccine (CYD-TDV; Dengvaxia®, Sanofi Pasteur) has been approved for use in several countries^{29,30}. This vaccine has been shown to induce the expression of broadly neutralizing antibodies to multiple strains

and all serotypes of dengue viruses³¹. The results of a phase III trial of this vaccine suggest that both the immune state (with respect to dengue viruses) and circulating viruses may influence vaccine effectiveness²⁹. This underscores the need to characterize both the sequence evolution and antibody response of circulating dengue strains.

Here we used an unbiased sequencing/metagenomic approach, in order to determine both the identity and sequences of viruses associated with febrile illness. In particular, based on previous studies of sequencing data from the serum of febrile individuals, we expected that medium-depth sequencing (about 10–20 million sequence reads per sample) was necessary and sufficient to provide complete sequences of small viral genomes from clinical material^{2,9}. To test this, we sequenced RNA extracted from the serum of four individuals and the plasma of another presenting with febrile illness at a tertiary care hospital in Bangalore, India and one healthy control from the same hospital, during the dengue season of 2014. We recovered the complete coding sequence of DENV3 clustering into a recent genotype III clade.

Results

We sequenced RNA extracted from the serum of four patients hospitalized with severe febrile illness and from one plasma sample from a patient hospitalized with prolonged febrile illness ([Table 1](#)). We included serum from a healthy individual and water as controls. Approximately 10×10⁶ sequence reads were recovered from each sample, with the water control yielding a lower number of reads ([Figure 1A](#)).

A BLAST³² similarity search, mapping all sequenced reads to a database of NCBI reference viral sequences ([Table 1](#)), identified 19,120 DENV3 sequence reads and 14 JEV sequence reads in sample F2, and 12 JEV sequence reads in sample F5. A single DENV3 read was detected in sample F3. No animal viruses were confirmed by BLAST in the controls or in other samples ([Table 1](#) and [Figure 1B](#)).

On the basis of World Health Organization guidelines for the classification of dengue cases³³, F2 was classified as a case of severe dengue, as the presenting symptoms included respiratory distress (bilateral pleural effusions in chest X-ray) hypotension and elevated liver enzymes ([Table 1](#)).

The serum sample from this individual was positive for both the non-structural protein 1 antigen and dengue IgM, and we were able to obtain a complete DENV3 genome sequence from this sample. Genomes were assembled both by *de novo* (87.05% coverage) and mapping-based (99% coverage) assembly ([Table 2](#) and [Table 3](#), [Supplementary File 1](#)) and found to be identical ([Supplementary File 2](#)). Mapping revealed good coverage across the genome, with an average depth of 231.45 ([Figure 1](#), [Table 2](#)). The genome is missing 76 bp at the 5'-UTR and 28 bp at the 3'-UTR compared to the NCBI RefSeq ([NC_001475.2](#)) DENV3 genome.

The mapping-based assembly was used for phylogenetic analysis and submitted to GenBank, with accession number [KX855927](#).

Table 1. Clinical Profile of the sequenced cases. The clinical presentation, key diagnostics tests, provisional diagnosis, treatment followed and results from sequencing (SNAP alignment against viral databases) are shown.

Sample	Age/sex	Presentation	Investigations	Diagnosis	Management	Animal viruses (sequencing+ BLAST)
F1	34F	Fever, vomiting, loose stools, hypotension	<ul style="list-style-type: none"> dengue IgM + Serial platelet count: 57,000-12,000-37,000-60,000 cells/mm³ BP 106/72 mmHg 	Dengue	Platelet transfusion, antiemetics, IV fluid; patient recovered and was discharged after 5 days	None matched
F2	28F	Fever, severe myalgia for 4 days, hypotension	<ul style="list-style-type: none"> Dengue IgM + Dengue NS1 + LFT: AST 370 U/l; ALT 170 U/l; GGT 272 U/l Chest X ray: Bilateral pleural effusion Serial platelet count: -7000-16000-43000 cells/mm³ BP 80/60mmHg 	Dengue	Platelet transfusion IV fluids; patient improved and was discharged	Dengue virus 3 (19, 120 reads) Japanese encephalitis virus (14 reads)
F3	36F	Fever and severe myalgia for 15 days	Weil-Felix border line positive (OX K 1:80) for Rickettsial fever	Rickettsial fever	Doxycycline (200 mg for 7 days); patient recovered	None matched
F4	10M	Prolonged fever (>20 days)	No known cause	Provisional diagnosis Rickettsial or partially treated enteric/malaria		dengue virus 3 (1 read)
F5	42F	Fever for 13 days, chills and rigors, known diabetic	Weil-Felix suggestive of Rickettsial Fever (OX K 1:320)	Rickettsial fever	Doxycycline (200 mg for 5 days); patient improved	Japanese encephalitis virus sequences (12 reads)

M, Male; F, Female; IgM, dengue immunoglobulin M; NS1, dengue non-structural protein 1 test; LFT, liver function test; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma glutamyltransferase.

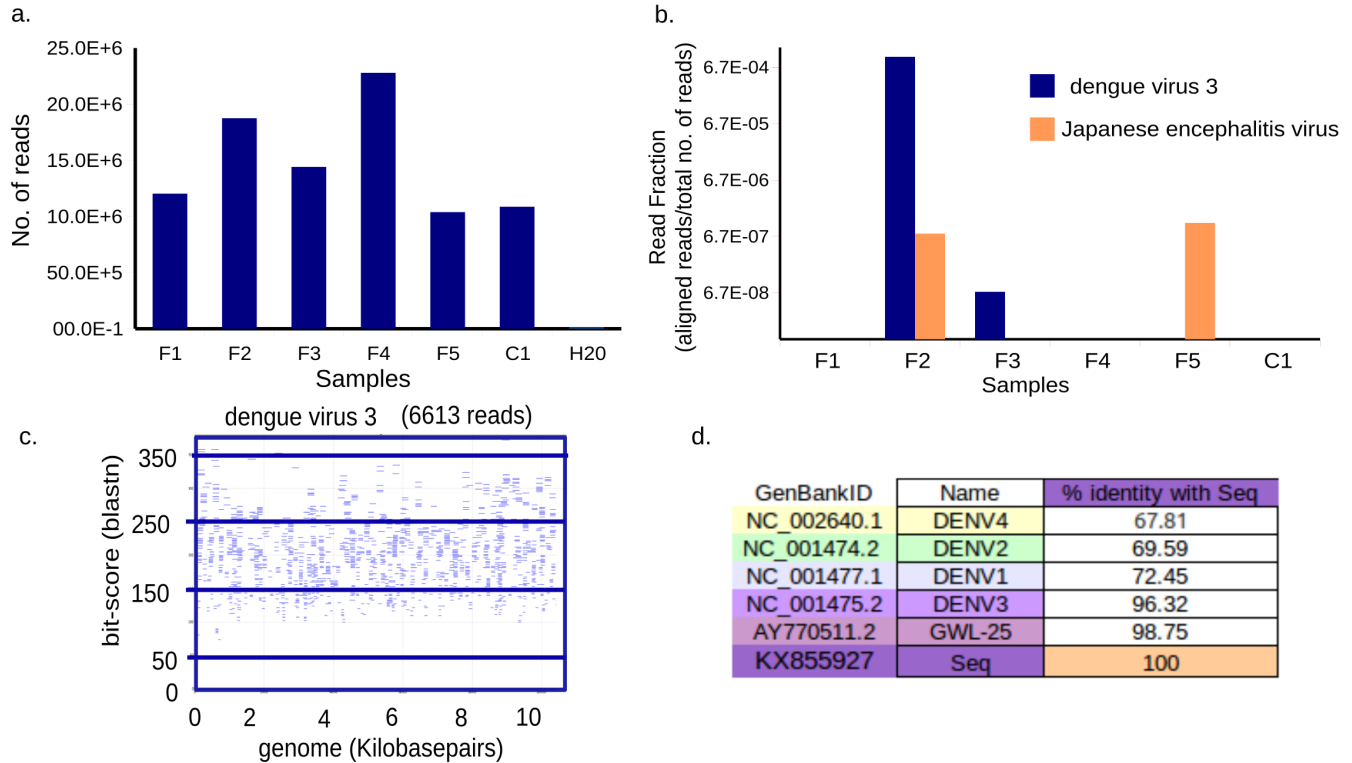


Figure 1. Dengue virus type 3 (DENV3) and Japanese encephalitis virus sequences identified from febrile serum. (a) Number of sequence reads generated per sample. (b) Bar graph showing number of reads that aligned to a particular virus as a fraction of the total number of reads (*y*-axis, log scale) from that sample (*x*-axis) using the SNAP alignment. (c) Alignment of sequences mapping only to DENV3 by nucleotide BLAST. Each rectangle shows sequencing reads (blue lines), their alignment to the genome (*x*-axis) and their blast bit-score (*y*-axis). Numbers below the title represents number of reads that mapped to the title. (d) Percentage identity of KX855927 with all four dengue viruses and the closest Indian strain.

Table 2. Assembly characteristics for mapping based assembly. The quality, coverage and percentage nucleotide identity of the assembled DENV3 genome using different back bones and sequences for mapping using MIRA assembler are shown.

Criteria	Backbone	av.qual	#-reads	mx.cov.	av.cov	GC%	CnNoCov
All Reads from F2 against all 4 Refseq of dengue viruses	DENV3	41	2009	96	26.27	46.67	145
	DENV1	30	2	3	1.01	46.67	10587
	DENV2	30	1	1	1	45.82	10723
	DENV4	30	1	1	1	47.12	10649
"virus reads" from F2 against all 4 Refseq of dengue viruses	DENV3	42	18180	788	231.53	46.66	104
	DENV1	30	3	4	1.02	46.67	10587
	DENV2	30	1	1	1	45.82	10723
	DENV4	30	1	1	1	47.12	10649
"virus reads" from F2 against DENV3 and an Indian strain	DENV3 (RefSeq)	42	18178	793	231.51	46.66	104
	AY770511.2	43	18696	792	236.58	46.65	104

Table shows the quality, coverage and percentage nucleotide identity of the assembled DENV3 genome using different back bones and sequences for mapping using MIRA assembler. Backbone, reference genome used for assembly; av.qual, average quality of assembly; mx.cov, maximum coverage of assembled genome by reads; av.cov, average coverage of assembled genome by reads; No cov, number of nucleotides of reference not covered in assembly; DENV3, dengue virus type 3.

The degree of nucleotide identity between this strain and the reference DENV3 genome (NC_001475.2) was 96.32%, and with the closest DENV3 strain from India, 98.75%.

Phylogenetic analysis was carried out with BEAST2 using the coding sequence of KX855927 and 79 sequences selected as being similar to KX855927, using the BLAST search against dengue genomes in the Virus Pathogen Database and Analysis Resource³⁴ (Supplementary File 3). The strain clusters with recent DENV3 sequences from India, China and Singapore (Figure 2). This clade split from other DENV3 and other DENV3 genotype III strains around 15 years ago. The branch length of KX855927 is longer than most others in the tree, with an estimated divergence time of 13.86 years (with the 95%

highest posterior densities between 12.94 and 14.83 years) from the closest Indian strain (Figure 2). A maximum likelihood tree showed the same topology as the consensus tree from BEAST, although many clades had low bootstrap support (Supplementary File 4).

Both synonymous and non-synonymous substitutions were predicted throughout the genome, as compared to the DENV3 reference sequence (Supplementary File 5). We aligned the E protein of all the complete genomes from Indian strains against the parent strain used to derive the tetravalent dengue vaccine (CYD-TDV; Dengvaxia®, Sanofi Pasteur) (Figure 3). Multiple amino acid substitutions were predicted throughout the envelope protein and two additional stop codons (at amino

Table 3. Assembly characteristics for *de novo* assembly. The assembly characteristics by *de novo* assembly of sequences from sample F2 after quality assessment was performed using the QUASt tool.

Fraction of genome covered	Largest alignment	Total aligned length	% nucleotide identity with mapping assembly	Reference for quality
87.046	3127	9403	100.00%	Refseq DENV3

DENV3, dengue virus type 3.

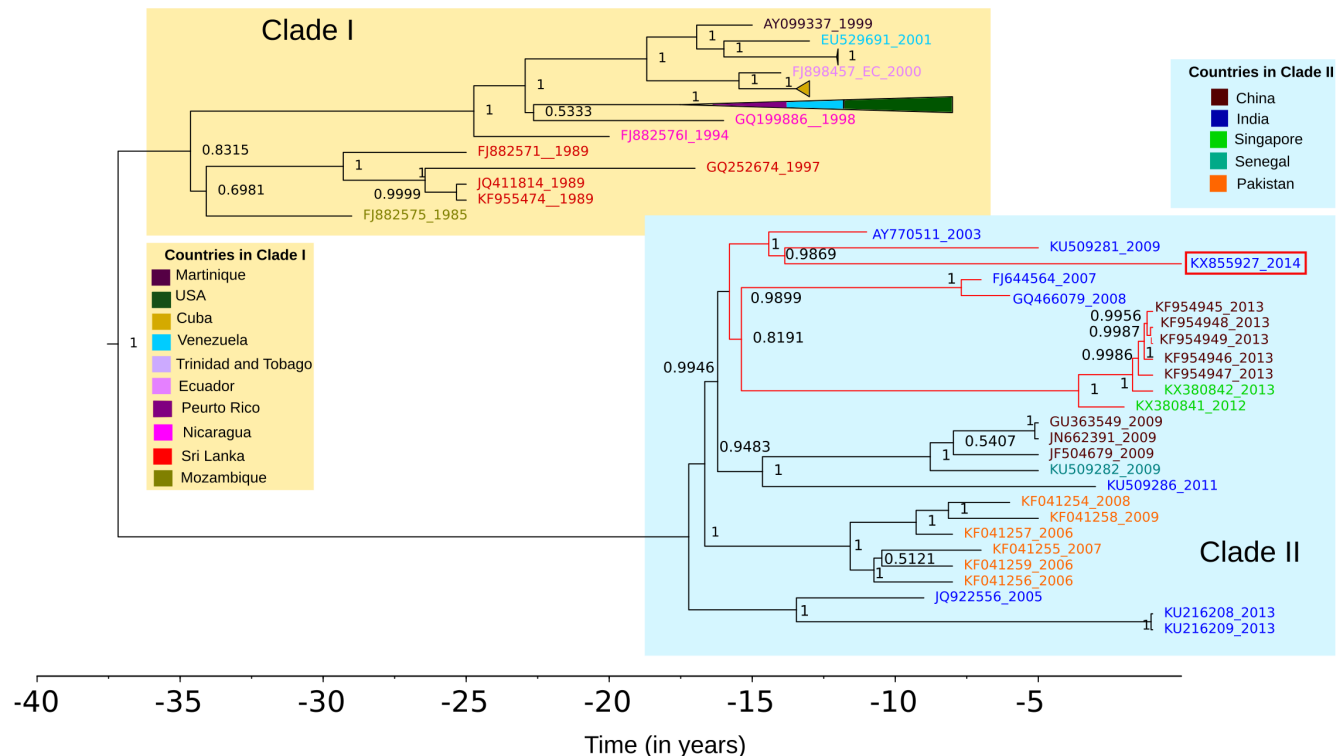


Figure 2. The sequenced strain KX855927 (2014) belongs to a recent Asian clade within genotype III. Figure shows the BEAST maximum clade credibility tree of the top 79 BLAST matches to KX855927. The Indo–China–Singapore strain to which KX855927 (2014) is shown in red. All strains are represented by their GenBank IDs and coloured by country. For ease of visualization, a clade containing viruses from the USA, Venezuela and Puerto Rico in Clade I has been collapsed (pyramids colored by country). The x-axis represents time in years.

sequence-based -enrichment of viral sequences using conserved sequences, will enable the recovery of complete genomes from routine clinical samples even with by lower-depth sequencing³⁵.

We identified a low number of reads mapping specifically to JEV. JEV is known to cause fevers^{5,6,36}. Further systematic analysis using a combination of polymerase chain reaction and IgM testing is required to ascertain how much JEV contributes to the acute fever burden in India. The low number of JEV reads obtained in both samples in which reads mapping to JEV were found suggests there was not much active viral replication occurring. There are previous reports of the detection of JEV sequences many months after infection³⁷. The sequences we found could therefore be remnants of a previous infection or may be the result of an infection from a mosquito bite that was checked by the immune system. The low number of reads in these cases mapped to the same gene (non-structural protein 5) (Supplementary File 6). This could reflect the higher stability of some parts of the JEV RNA genome.

The results of metagenomic sequencing, however, do need to be interpreted with caution owing to issues related to contamination^{10,11}. Contamination can occur in every step of the procedure, starting from sample collection, processing, sequencing and, when multiple indexed samples are sequenced together, de-multiplexing (the process in which reads get assigned to a sample). This needs to be taken into consideration, particularly when the number of sequences supporting the presence of a pathogen are low, when there is incomplete genome information, or when the same sequence is present in all the samples, including the controls. We have tried to mitigate this partially by the use of controls—serum from a healthy individual collected at the same time and place and a water sample processed in the same way as the clinical samples. However, we believe that independent methods are required to confirm novel/unexpected findings by this method.

DENV3 has been shown to be re-emerging in India, and has been responsible for severe outbreaks in other geographic regions, including in South America and Cuba^{27,38,39}. The full-length DENV3 (KX855927) we describe here clusters into a clade containing DENV3 viruses from India and is related to an Indo–China–Singapore clade. We observed a longer branch length for this particular strain, which could be the result of incomplete sampling of this clade or could indicate that this lineage is showing accelerated rates of molecular evolution⁴⁰. This can be resolved in future studies by the addition of more sequence information, as more full-length dengue sequences from India become available in the databases.

While both synonymous and non-synonymous changes were observed throughout the DENV3 (KX855927) genome compared to the DENV3 reference sequence (NC_001475.2), the changes in the antigenic E protein are of particular interest. Neutralizing antibodies have been described against the envelope protein that target particular epitopes^{26,41}. Critical amino acid residues that change antibody binding have also been described by others²⁶. The results from our phylogenetic

analyses are consistent with previous work tracing the emergence of new clade of DENV3 genotype III strains in India³⁹. The ability of a dengue vaccine to elicit neutralizing antibodies against locally circulating DENV3 strains therefore needs to be evaluated in this light.

Methods

Description of samples

In total, samples from five patients (two diagnosed with dengue fever (serum; F1 and F2), two with Rickettsial fever (serum; F3 and F5) and one with unknown fever (plasma; F4) presenting with febrile illness, and one healthy control (serum; C1) at St. John's Medical College and Hospital (SJMCH), Bangalore, were assessed in this study. Table 1 provides clinical characterization, treatment and outcomes of patients. The study was done after obtaining approval from the Institutional Ethics Committee of SJMCH, Bangalore, India (IEC Study Ref. No. 5/2016). A waiver of consent was sought and obtained for the analysis as it was done on samples remaining after routine diagnostic testing, which were de-linked from the identity of the patients. We have been granted a waiver of consent by the Institutional Ethics Committee of St. John's Medical College and Hospital, which does not permit the use of the generated data for human genetic studies.

Isolation of RNA

RNA was extracted using the Qiagen All-Prep kit, using 300–500 µl of serum/plasma and lysed using 1 ml of lysis buffer. The remaining protocol was performed as recommended by the manufacturer. Eluted RNA was concentrated and used for sequencing reactions.

Sequencing

Sequencing libraries were prepared using the Ion Proton library preparation protocol. Indexing was performed using the IonXpress RNA Seq Barcode kit (Thermo Fisher Scientific, Inc.). Samples F1–4 and C1 were run on the same chip; sample F5 was run on a separate chip. Libraries were pooled to give equimolar concentrations of 10 pM. This was used in template-preparation steps and RNA sequencing was performed using the Ion PI sequencing kit on the Ion Proton platform using the Ion PI™ ChipV2 and Ion PI™ Sequencing Kit V3 (Thermo Fisher Scientific, Inc.).

Analysis of sequences

We aligned the sequencing reads to a database of all known viruses using the SNAP alignment tool (snap-1.0beta.16-linux)⁴². All hits were verified using nucleotide BLAST sequence search and visualized using tools from the Dark Matter project. Reads aligning to the human genome, human mRNA, rRNA large subunit and rRNA small subunit from the SILVA database were removed⁴³. The aligned sequences were used as the input for assembly. *De novo* assembly was performed using the SPAdes (v3.10.1) tool⁴⁴. Quality assessment of the assembly was performed using the QUAST tool⁴⁵. MIRA 4.0.2 was used for mapping based assembly, with the GenBank sequence NC_001475.2 for DENV3 as the backbone for assembly and NC_001437 as the backbone for JEV⁴⁶. Contigs were

subjected to nucleotide BLAST using the online BLAST tool. The mapping based assembly of DENV3 obtained using MIRA was manually checked for regions with low confidence using [Gap5](#) (staden-2.0.0b11-2016-linux-x86_64)⁴⁷. Fewer than 30 nucleotides were found to have low confidence, of which 22 were in the 3'-UTR end region. The files from the MIRA assembly, together with the contributing reads, are provided as [Supplementary File 1](#). This sequence was submitted to GenBank with the accession number [KX855927](#).

Phylogenetic analysis

Phylogenetic analysis was performed with BLAST search hits to [KX855927](#) in the [VipR dengue virus](#) database³⁴. Only the coding sequence was used for the analyses. The alignment was visualized using [AliView](#) software (v1.18)⁴⁸. Nucleotide distances of [KX855927](#) from other dengue viruses, using the reference sequence and the closest BLAST hit from India, were estimated using the [MUSCLE](#) alignment tool to create a percentage identity matrix⁴⁹. The Generalized Time Reversible Model, namely GTR+I+G, GTR+I+G, GTR+G, were found to be the best evolutionary models for codon positions 1, 2, and 3, respectively, using [PartitionFinder](#) (v2.1.1)⁵⁰, where I represents invariant and G represents gamma, a shape parameter for the model. A previously estimated rate of substitution for DENV3 = 7.48×10^{-4} subs/site/year (4.47×10^{-4} ; 10.72×10^{-4}) was used to set a strict molecular clock⁵¹. The input XML file to [BEAST](#) (v2.4.6)⁵² is provided in [Supplementary File 3](#). Tracer (v1.6) was used to confirm sufficient sampling (effective sample size > 200 for all parameters). [TreeAnnotator](#) (v2.4.6) was used to generate the maximum clade credibility tree, where the node heights represent median height. Posterior probabilities for both the split of Clade I and II and the clade containing [KX855927](#) were >95%. The tree was visualized using [FigTree](#) (v1.4.3). The Maximum Likelihood tree was generated using thorough search and 1000 bootstraps in [RaXML](#) (RAxML -NG v0.4.1 BETA) ([Supplementary File 4](#))⁵³.

Analysis of E protein

E protein alignments for the DENV3 complete genomes from India were performed in [AliView](#) and amino acid differences were highlighted compared to PaH881/8 (AF349753) the parent strain used in the development of Dengvaxia (CYD-TDV; a tetravalent, live attenuated, chimeric dengue vaccine with a yellow fever 17D backbone). Homology modeling was performed for the E protein of [KX855927](#) using [SWISS-MODELSWISS-MODEL](#) and the best model was chosen for showing the substitutions. The protein surfaces, as visualized using [PyMOL](#) (version 1.8; PyMOL Molecular Graphics System, Schrödinger, LLC), are shown in light brown; the amino acids found to be different in the [KX855927](#) strain are colored by the CHNOS elements. The datasets supporting the conclusions of this article are included within the article and in [Supplementary File 1–Supplementary File 6](#).

An earlier version of this work can be found on bioRxiv (<https://doi.org/10.1101/204503>).

Data availability

The raw files from sequencing are not provided in their entirety as these are metagenomic datasets that contain identifying host information. Therefore we have used only sequences not aligning to the human genome for our research. This data has been uploaded in fastq format on OSF (see below). As our experiments were designed to identify pathogens, we do expect the accompanying human data to be free from biases involving sampling, storage and handling. However, under the conditions that the samples remain de-identified, and the work is not directly on human genetics, approval for data sharing of the complete data from the RNA sequencing experiment, which includes any human sequences, can be sought with the Institutional Ethics Committee, St. John's Medical College and Hospital, Bangalore. A request for use of this data for a research proposal must be submitted to the ethics committee via the lead author (soniamarydias@hotmail.com).

Fastq files have been made available from OSF, <http://doi.org/10.17605/OSF.IO/RMQDF54>.

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Grant information

This study was supported by the Department of Biotechnology, Glue Grant to SK. PC was supported by the Department of Biotechnology Research Associate fellowship, Royal Society -SERB Newton International fellowship and the India Alliance (DBT-Wellcome Trust) Early Career Fellowship. TS is supported by the NIHR Health Protection Research Unit in Emerging and Zoonotic Infections and the European Union's Horizon 2020 research and innovation program ZikaPLAN (Preparedness Latin America Network), grant agreement No. 734584. DS, BM, and TJ were supported by the European Union FP7 programme ANTIGONE (grant agreement No. 278976).

The funders had no role in the in any steps of experimental design, data collection or analysis or the decision to publish.

Acknowledgements

Dr. Lisa Ng and Dr. Julian Hiscox for critical input and helpful discussions. Sreejayan Nambiar, Field application specialist, Thermo Fisher Scientific, Inc. for technical support during the sequencing.

Supplementary material

Supplementary File 1: Dengue virus type 3 assembly with contributing reads. Consensus sequence and reads contributing to the assembly, it contains regions/bases flagged by MIRA including 30 base positions which have low confidence. The file format (.maf) can be converted to compatible file formats for viewing with Gap5, Consed and other genome editors.

[Click here to access the data.](#)

Supplementary File 2: Percentage identity of dengue virus type 3 sequence assembled by different methods. Table shows the percentage similarity between the *de novo* and mapping assemblies compared to NCBI reference sequences of dengue virus types 1-4 and the closest Indian strain.

[Click here to access the data.](#)

Supplementary File 3: Template for BEAST. The input file used for phylogenetic analysis using the BEAST program.

[Click here to access the data.](#)

Supplementary File 4: Maximum Likelihood Tree from RaxML. Figure shows the maximum likelihood trees with bootstrap values, GenBank IDs and year of sequencing are shown on the tips. The data is coloured by country and some of the clades have been collapsed for ease of viewing.

[Click here to access the data.](#)

Supplementary File 5: Single nucleotide polymorphisms in KX855927 with respect to NC_001475. Description of the single nucleotide polymorphisms in KX855927 with respect to NC_001475, as detected by the MIRA assembly program.

[Click here to access the data.](#)

Supplementary File 6: BLAST results of Japanese encephalitis virus contigs from sample F2 and F5. Contains the Top 5 nucleotide BLAST lists for the Japanese encephalitis virus contigs assembled from samples F2 and F5, against the nucleotide database, flavivirus database and dengue virus database.

[Click here to access the data.](#)

References

- Susilawati TN, McBride WJ: **Acute undifferentiated fever in Asia: a review of the literature.** *Southeast Asian J Trop Med Public Health.* 2014; **45**(3): 719–26. [PubMed Abstract](#)
- Mueller TC, Siv S, Khim N, *et al.*: **Acute undifferentiated febrile illness in rural Cambodia: a 3-year prospective observational study.** *PLoS One.* 2014; **9**(4): e95868. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wylie KM, Mihindukulasuriya KA, Sodergren E, *et al.*: **Sequence analysis of the human virome in febrile and afebrile children.** *PLoS One.* 2012; **7**(6): e27735. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Capeding MR, Chua MN, Hadinegoro SR, *et al.*: **Dengue and other common causes of acute febrile illness in asia: an active surveillance study in children.** *PLoS Negl Trop Dis.* 2013; **7**(7): e2331. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mayxay M, Castonguay-Vanier J, Chansamouth V, *et al.*: **Causes of non-malarial fever in Laos: a prospective study.** *Lancet Glob Heal.* 2013; **1**(1): e46–54. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chheng K, Carter MJ, Emary K, *et al.*: **A prospective study of the causes of febrile illness requiring hospitalization in children in Cambodia.** *PLoS One.* 2013; **8**(4): e60634. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- McMullan LK, Folk SM, Kelly AJ, *et al.*: **A new phlebovirus associated with severe febrile illness in Missouri.** *N Engl J Med.* 2012; **367**(9): 834–41. [PubMed Abstract](#) | [Publisher Full Text](#)
- Yu XJ, Liang MF, Zhang SY, *et al.*: **Fever with thrombocytopenia associated with a novel bunyavirus in China.** *N Engl J Med.* 2011; **364**(16): 1523–32. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Yozwiak NL, Skewes-Cox P, Stenglein MD, *et al.*: **Virus identification in unknown tropical febrile illness cases using deep sequencing.** *PLoS Negl Trop Dis.* 2012; **6**(2): e1485. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Firth C, Lipkin WI: **The genomics of emerging pathogens.** *Annu Rev Genomics Hum Genet.* 2013; **14**: 281–300. [PubMed Abstract](#) | [Publisher Full Text](#)
- Barzon L, Lavezzo E, Costanzi G, *et al.*: **Next-generation sequencing technologies in diagnostic virology.** *J Clin Virol.* 2013; **58**(2): 346–50. [PubMed Abstract](#) | [Publisher Full Text](#)
- Dash AP, Bhatia R, Sunyoto T, *et al.*: **Emerging and re-emerging arboviral diseases in Southeast Asia.** *J Vector Borne Dis.* 2013; **50**(2): 77–84. [PubMed Abstract](#)
- Shepard DS, Halasa YA, Tyagi BK, *et al.*: **Economic and disease burden of dengue illness in India.** *Am J Trop Med Hyg.* 2014; **91**(6): 1235–42. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Turtle L, Griffiths MJ, Solomon T: **Encephalitis caused by flaviviruses.** *QJM.* 2012; **105**(3): 219–23. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sim S, Hibberd ML: **Genomic approaches for understanding dengue: insights from the virus, vector, and host.** *Genome Biol.* 2016; **17**: 38. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Drake JW, Holland JJ: **Mutation rates among RNA viruses.** *Proc Natl Acad Sci U S A.* 1999; **96**(24): 13910–3. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Akram M, Fatima Z, Purdy MA, *et al.*: **Introduction and evolution of dengue virus type 2 in Pakistan: a phylogeographic analysis.** *Virology.* 2015; **12**: 148. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tissera HA, Ooi EE, Gubler DJ, *et al.*: **New dengue virus type 1 genotype in Colombo, Sri Lanka.** *Emerg Infect Dis.* 2011; **17**(11): 2053–5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

19. Kanakarathne N, Wahala WM, Messer WB, *et al.*: **Severe dengue epidemics in Sri Lanka, 2003–2006.** *Emerg Infect Dis.* 2009; 15(2): 192–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Kotaki T, Yamanaka A, Mulyatno KC, *et al.*: **Divergence of the dengue virus type 2 Cosmopolitan genotype associated with two predominant serotype shifts between 1 and 2 in Surabaya, Indonesia, 2008–2014.** *Infect Genet Evol.* 2016; 37: 88–93.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Manokaran G, Finol E, Wang C, *et al.*: **Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness.** *Science.* 2015; 350(6257): 217–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Manakkadan A, Joseph I, Prasanna RR, *et al.*: **Lineage shift in Indian strains of Dengue virus serotype-3 (Genotype III), evidenced by detection of lineage IV strains in clinical cases from Kerala.** *Virology.* 2013; 453: 37–43.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Katzelnick LC, Fonville JM, Gromowski GD, *et al.*: **Dengue viruses cluster antigenically but not as discrete serotypes.** *Science.* 2015; 349(6254): 1338–43.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Yacoub S, Mongkolsapaya J, Screaton G: **Recent advances in understanding dengue [version 1; referees: 3 approved].** *F1000Res.* 2016; 5: pii: F1000 Faculty Rev-78.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Dejnirattisai W, Jumnainsong A, Onsririsakul N, *et al.*: **Cross-reacting antibodies enhance dengue virus infection in humans.** *Science.* 2010; 328(5979): 745–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Wahala WM, Donaldson EF, de Alwis R, *et al.*: **Natural strain variation and antibody neutralization of dengue serotype 3 viruses.** *PLoS Pathog.* 2010; 6(3): e1000821.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Beltramello M, Williams KL, Simmons CP, *et al.*: **The human immune response to dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity.** *Cell Host Microbe.* 2010; 8(3): 271–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Guzman MG, Alvarez M, Halstead SB: **Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: an historical perspective and role of antibody-dependent enhancement of infection.** *Arch Virol.* 2013; 158(7): 1445–59.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Guy B, Lang J, Saville M, *et al.*: **Vaccination Against Dengue: Challenges and Current Developments.** *Annu Rev Med.* 2016; 67: 387–404.
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Guy B, Briand O, Lang J, *et al.*: **Development of the Sanofi Pasteur tetravalent dengue vaccine: One more step forward.** *Vaccine.* 2015; 33(50): 7100–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Barban V, Munoz-Jordan JL, Santiago GA, *et al.*: **Broad neutralization of wild-type dengue virus isolates following immunization in monkeys with a tetravalent dengue vaccine based on chimeric yellow fever 17D/dengue viruses.** *Virology.* 2012; 429(2): 91–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Altschul SF, Gish W, Miller W, *et al.*: **Basic local alignment search tool.** *J Mol Biol.* 1990; 215(3): 403–10.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. World Health Organization and the Special Programme for Research and Training in Tropical Diseases: **Dengue Guidelines for diagnosis, treatment, prevention and control.** 2009.
[Reference Source](#)
34. Pickett BE, Sadat EL, Zhang Y, *et al.*: **ViPR: an open bioinformatics database and analysis resource for virology research.** *Nucleic Acids Res.* 2012; 40(Database issue): D593–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Briese T, Kapoor A, Mishra N, *et al.*: **Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis.** *mBio.* 2015; 6(5): e01491–15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Solomon T, Dung NM, Kneen R, *et al.*: **Japanese encephalitis.** *J Neurol Neurosurg Psychiatry.* 2000; 68(4): 405–15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Ravi V, Desai AS, Shenoy PK, *et al.*: **Persistence of Japanese encephalitis virus in the human nervous system.** *J Med Virol.* 1993; 40(4): 326–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Rodriguez-Roche R, Blanc H, Borderia AV, *et al.*: **Increasing Clinical Severity during a Dengue Virus Type 3 Cuban Epidemic: Deep Sequencing of Evolving Viral Populations.** *J Virol.* 2016; 90(9): 4320–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Dash PK, Parida MM, Saxena P, *et al.*: **Reemergence of dengue virus type-3 (subtype-III) in India: implications for increased incidence of DHF & DSS.** *Virology.* 2006; 353: 55–60.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Schuettelpelz E, Pryer KM: **Reconciling extreme branch length differences: decoupling time and rate through the evolutionary history of filmy ferns.** *Syst Biol.* 2006; 55(3): 485–502.
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Wahala WM, de Silva AM: **The human antibody response to dengue virus infection.** *Viruses.* 2011; 3(12): 2374–95.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Zaharia M, Bolosky WJ, Curtis K, *et al.*: **Faster and More Accurate Sequence Alignment with SNAP.** arXiv: 11115572v1. 2011.
[Reference Source](#)
43. Quast C, Pruesse E, Yilmaz P, *et al.*: **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.** *Nucleic Acids Res.* 2013; 41(Database issue): D590–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Nurk S, Bankevich A, Antipov D, *et al.*: **Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads.** Springer, Berlin, Heidelberg; 2013; 158–70.
[Publisher Full Text](#)
45. Gurevich A, Saveliev V, Vyahhi N, *et al.*: **QUAST: quality assessment tool for genome assemblies.** *Bioinformatics.* 2013; 29(8): 1072–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Chevreaux B, Pfisterer T, Drescher B, *et al.*: **Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs.** *Genome Res.* 2004; 14(6): 1147–59.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Bonfield JK, Whitwham A: **Gap5—editing the billion fragment sequence assembly.** *Bioinformatics.* 2010; 26(14): 1699–703.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Larsson A: **AliView: a fast and lightweight alignment viewer and editor for large datasets.** *Bioinformatics.* 2014; 30(22): 3276–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res.* 2004; 32(5): 1792–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Lanfear R, Calcott B, Ho SY, *et al.*: **Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses.** *Mol Biol Evol.* 2012; 29(6): 1695–701.
[PubMed Abstract](#) | [Publisher Full Text](#)
51. Twiddy SS, Holmes EC, Rambaut A: **Inferring the rate and time-scale of dengue virus evolution.** *Mol Biol Evol.* 2003; 20(1): 122–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
52. Bouckaert R, Heled J, Kühnert D, *et al.*: **BEAST 2: a software platform for Bayesian evolutionary analysis.** *PLoS Comput Biol.* 2014; 10(4): e1003537.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics.* 2014; 30(9): 1312–3.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
54. Pattabiraman C, Mary D: **Febrile_illness_sequencing.** *Open Science Framework.* 2018.
<http://www.doi.org/10.17605/OSF.IO/RMQDF>

Open Peer Review

Current Referee Status:  

Version 1

Referee Report 08 August 2018

<https://doi.org/10.21956/wellcomeopenres.15716.r33558>



John A Lednicky

Department of Environmental and Global Health, College of Public Health and Health Professions, Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

The report is straight-forward and presents a successful application of metagenomics sequencing to help diagnose febrile illnesses with no known causes. The assembled DENV3 sequence will be very useful for epidemiology studies and to track the evolution of DENV3.

Specific items for the authors to consider:

1. OPTIONAL: On page 3, the authors mention Dengvaxia. It will actually strengthen the authors' projected use of metagenomics sequencing if they mentioned the problems in the Philippines and elsewhere with this vaccine. Will it help to know before vaccines are used which virus strains are in circulation? And will that sequence knowledge be useful in the future for evaluating the possible side-effects of vaccines before the vaccines are administered?
2. Figure 3: As viewed on my computer and a printed copy of the manuscript, neither blue nor violet colors were discernable for FIG. 3b.
3. OPTIONAL: The authors might clarify: On page 7, DISCUSSION, par. 1: It is not clear if the authors agree with the reference cited (35), or whether they have experienced this first hand.
4. Page 8: Isolation of RNA: Why was the All-Prep kit chosen for serum/plasma instead of a QIAGEN viral RNA kit? The former is typically used for whole blood. How was the eluted RNA concentrated? Why?
5. OPTIONAL: The authors make no mention of the costs/time commitments/ personnel/training/facilities to accomplish the type of work they performed. Though the approach is very good (and clearly the wave of the future), it is at present very costly and not practical for most medical laboratories, even in developed countries. How can this be remedied?
6. OPTIONAL: The authors might comment further on this: It will be curious to many readers why JEV sequences were also detected. Is this a finding of major importance? Has this been observed in other studies (persistence of JEV sequences)? Could the severity of illness in patient F2 have been affected by a "smoldering" JEV infection?
7. More explanation would be useful regarding how the sequencing libraries were made. For example, flavivirus RNA is not polyadenylated and is capped. Do the authors have a specific approach to capture these types of RNA?

*OPTIONAL: Some possible clarifications/comments from the authors:

(1) Next generation sequencing definitely has a role diagnostics, phylogeographic studies, etc. On page 3, the authors state "We hypothesized that ...". Some readers will interpret that to mean the authors claim they conceived this concept, and this is the first time the approach has been used. The authors might reword that statement.

(2) The authors did not comment on the specificity of the DENV IgM test.

(3) Many NGS-derived virus sequences deposited in data banks have errors. Thus, whereas the NGS-generated data is useful for the diagnosis, indels etc. are not always verified in the submitted sequences, the general understanding among users being that errors may be present. Also, often, the 5' and 3' UTRs are not included. The authors might discuss these issues. In particular, there are artefacts associated with Ion Proton sequencing associated with repeats of the same nucleotide (example: TTTTT).

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 03 Jan 2019

Chitra Pattabiraman, National Institute of Mental Health and Neurosciences, India, India

We thank the reviewer for the comments. They are addressed point-wise below.

1. New Generation sequencing approaches are attractive options to identify viral quasispecies in patients. In the present article, the authors have presented a common consensus final sequence of DENV-3 and compared it with reference sequence and the closely related GWL-25 genome. Did the authors had a chance to look at the raw read clusters, and identify possible variations that could represent quasi species by looking at sites/ sequence regions showing consistent variability? A comparison table showing the non-synonymous sites could have been included.

We have not performed a systematic analysis for viral variants within our sample, however, given that we got greater than 100X coverage of many regions of the genome, we believe this analysis is possible. A comparison table showing the changes between the DENV3 strain reported here and the RefSeq DENV3 strain are included in Supplementary File 5.

2. *Considering that DENV virus protein translation starts as a polyprotein formation from a single ORF, identifying two internal stop codons at positions 58 & 168 seems to be curious. Have the authors done a manual verification of the corresponding raw read alignments? What could be the implications of this finding?*

We have realigned the sequence and manually verified this. Both changes are well supported. Our experiment cannot distinguish between defective viral particles and viable ones. Also, we are potentially putting together sequences from different viruses to make the consensus. These two factors limit our ability to interpret these findings.

3. *The authors have identified two missing regions-one each in the 5' and 3' UTR. Are they internal gaps or terminal truncation in the sequence? Could this be a failure to sequence these regions by the technique or is it possible that the DENV-3 strain itself is lacking these regions? Since the UTRs contain major replication regulatory elements, discussing this observation seems to be important.*

Given the importance of the UTRs in viral replication, we strongly suspect that our approach has limitations in capturing the ends of the genome.

4. *In the result section, the unique substitution is mentioned as D361E, where as in the fig.3A it is shown as E361D. The reader gets confused on the order of the sequences that are being compared.*

The text has been corrected in line with the figure.

Most of the amino acid substitutions were shared among all the Indian strains, while a D361E substitution was unique to the DENV3 strain reported here (Figure 3A).

to

Most of the amino acid substitutions were shared among all the Indian strains, while a E361D substitution was unique to the DENV3 strain reported here (Figure 3A).

5. *In Fig.3b(ii), the authors have highlighted the amino acid residues that are important in antibody binding and antigenicity. But how many of these changes are functionally important, based on the chemical nature of the amino acid substitutions observed? Also, the authors should indicate the wild type as well as the mutant amino acids along with the position number in the figure.*

The figure and accompanying legend have been modified.

6. *Was there any specific reason in using plasma sample from one patient and serum from others? It would be good to indicate, if available, the duration of fever in F1, and treatment and outcome in F4, for reasons of consistency.*

In this study, we used samples remaining after routine testing and plasma sample was available for that patient.

Competing Interests: No competing interests were disclosed.

<https://doi.org/10.21956/wellcomeopenres.15716.r33079>



Shijulal Nelson-Sathi¹, **Easwaran Sreekumar**²

¹ Computational Biology & Bioinformatics, Rajiv Gandhi Centre for Biotechnology-BIC Campus, Kinfra, Kazhakoottam, Thiruvananthapuram, Kerala, India

² Molecular Virology Laboratory, Rajiv Gandhi Centre for Biotechnology (RGCB), Thiruvananthapuram, Kerala, India

The manuscript by Dias *et al* describes the use of metagenomic sequencing of RNA from serum/plasma samples of febrile patients using an Ion Proton system to identify viral sequences. Using this approach they could identify presence and complete sequence of DENV-3 in one of the samples tested. They could also detect presence of Japanese Encephalitis virus sequences in the sample from the same patient and also in another sample. The work is interesting and has shown the possibility of using this approach in clinical samples for detection of pathogen sequence signatures.

This reviewer had the following queries:

1. New Generation sequencing approaches are attractive options to identify viral quasispecies in patients. In the present article, the authors have presented a common consensus final sequence of DENV-3 and compared it with reference sequence and the closely related GWL-25 genome. Did the authors had a chance to look at the raw read clusters, and identify possible variations that could represent quasi species by looking at sites/ sequence regions showing consistent variability?
A comparison table showing the non-synonymous sites could have been included.
2. Considering that DENV virus protein translation starts as a polyprotein formation from a single ORF, identifying two internal stop codons at positions 58 & 168 seems to be curious. Have the authors done a manual verification of the corresponding raw read alignments? What could be the implications of this finding?
3. The authors have identified two missing regions-one each in the 5' and 3' UTR. Are they internal gaps or terminal truncation in the sequence? Could this be a failure to sequence these regions by the technique or is it possible that the DENV-3 strain itself is lacking these regions? Since the UTRs contain major replication regulatory elements, discussing this observation seems to be important.
4. In the result section, the unique substitution is mentioned as D361E, where as in the fig.3A it is shown as E361D. The reader gets confused on the order of the sequences that are being compared.
5. In Fig.3b(ii), the authors have highlighted the amino acid residues that are important in antibody binding and antigenicity. But how many of these changes are functionally important, based on the chemical nature of the amino acid substitutions observed? Also, the authors should indicate the wild type as well as the mutant amino acids along with the position number in the figure.
6. Was there any specific reason in using plasma sample from one patient and serum from others? It would be good to indicate, if available, the duration of fever in F1, and treatment and outcome in F4, for reasons of consistency.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Molecular Virology, Dengue, Chikungunya, Host-pathogen interaction, Virus evolution

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 03 Jan 2019

Chitra Pattabiraman, National Institute of Mental Health and Neurosciences, India, India

We thank the reviewers for the comments. They are addressed point wise below

1. New Generation sequencing approaches are attractive options to identify viral quasispecies in patients. In the present article, the authors have presented a common consensus final sequence of DENV-3 and compared it with reference sequence and the closely related GWL-25 genome. Did the authors had a chance to look at the raw read clusters, and identify possible variations that could represent quasi species by looking at sites/ sequence regions showing consistent variability? A comparison table showing the non-synonymous sites could have been included.

We have not performed a systematic analysis for viral variants within our sample, however, given that we got greater than 100X coverage of many regions of the genome, we believe this analysis is possible. A comparison table showing the changes between the DENV3 strain reported here and the RefSeq DENV3 strain are included in Supplementary File 5.

2. Considering that DENV virus protein translation starts as a polyprotein formation from a single ORF, identifying two internal stop codons at positions 58 & 168 seems to be curious. Have the authors done a manual verification of the corresponding raw read alignments? What could be the implications of this finding?

We have realigned the sequence and manually verified this. Both changes are well supported. Our experiment cannot distinguish between defective viral particles and viable ones. Also, we are

potentially putting together sequences from different viruses to make the consensus. These two factors limit our ability to interpret these findings.

3. The authors have identified two missing regions—one each in the 5' and 3' UTR. Are they internal gaps or terminal truncation in the sequence? Could this be a failure to sequence these regions by the technique or is it possible that the DENV-3 strain itself is lacking these regions? Since the UTRs contain major replication regulatory elements, discussing this observation seems to be important.

Given the importance of the UTRs in viral replication, we strongly suspect that our approach has limitations in capturing the ends of the genome.

4. In the result section, the unique substitution is mentioned as D361E, where as in the fig.3A it is shown as E361D. The reader gets confused on the order of the sequences that are being compared.

The text has been corrected in line with the figure.

Most of the amino acid substitutions were shared among all the Indian strains, while a D361E substitution was unique to the DENV3 strain reported here (Figure 3A).

to

Most of the amino acid substitutions were shared among all the Indian strains, while a E361D substitution was unique to the DENV3 strain reported here (Figure 3A).

5. In Fig.3b(ii), the authors have highlighted the amino acid residues that are important in antibody binding and antigenicity. But how many of these changes are functionally important, based on the chemical nature of the amino acid substitutions observed? Also, the authors should indicate the wild type as well as the mutant amino acids along with the position number in the figure.

The figure and accompanying legend have been modified.

6. Was there any specific reason in using plasma sample from one patient and serum from others? It would be good to indicate, if available, the duration of fever in F1, and treatment and outcome in F4, for reasons of consistency.

In this study, we used samples remaining after routine testing and plasma sample was available for that patient.

Competing Interests: No competing interests were disclosed.