

RESEARCH ARTICLE

# Nature and Extent of Genetic Diversity of Dengue Viruses Determined by 454 Pyrosequencing

Md Abu Choudhury<sup>1,2\*</sup>, William B Lott<sup>2,3</sup>, Shahera Banu<sup>2</sup>, Anthony Youzhi Cheng<sup>4</sup>, Yik-Ying Teo<sup>4,5,6,7</sup>, Rick Twee-Hee Ong<sup>4</sup>, John Aaskov<sup>2</sup>

**1** Menzies Health Institute Queensland, Griffith University, Brisbane, Australia, **2** Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia, **3** School of Chemistry, Physics, and Mechanical Engineering, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia, **4** Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore, **5** Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore, **6** Life Sciences Institute, National University of Singapore, Singapore, Singapore, **7** Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore

\* [nahid.choudhury@griffith.edu.au](mailto:nahid.choudhury@griffith.edu.au)



**OPEN ACCESS**

**Citation:** Choudhury MA, Lott WB, Banu S, Cheng AY, Teo Y-Y, Ong RT-H, et al. (2015) Nature and Extent of Genetic Diversity of Dengue Viruses Determined by 454 Pyrosequencing. PLoS ONE 10 (11): e0142473. doi:10.1371/journal.pone.0142473

**Editor:** Matthias Johannes Schnell, Thomas Jefferson University, UNITED STATES

**Received:** May 22, 2013

**Accepted:** October 22, 2015

**Published:** November 13, 2015

**Copyright:** © 2015 Choudhury et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The work was supported by National Health and Medical Research Council of Australia, under grant 497203, Cook Estate and QUT Postgraduate Research Scholarships, Queensland University of Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Dengue virus (DENV) populations are characteristically highly diverse. Regular lineage extinction and replacement is an important dynamic DENV feature, and most DENV lineage turnover events are associated with increased incidence of disease. The role of genetic diversity in DENV lineage extinctions is not understood. We investigated the nature and extent of genetic diversity in the envelope (E) gene of DENV serotype 1 representing different lineages histories. A region of the DENV genome spanning the E gene was amplified and sequenced by Roche/454 pyrosequencing. The pyrosequencing results identified distinct sub-populations (haplotypes) for each DENV-1 E gene. A phylogenetic tree was constructed with the consensus DENV-1 E gene nucleotide sequences, and the sequences of each constructed haplotype showed that the haplotypes segregated with the Sanger consensus sequence of the population from which they were drawn. Haplotypes determined through pyrosequencing identified a recombinant DENV genome that could not be identified through Sanger sequencing. Nucleotide level sequence diversities of DENV-1 populations determined from SNP analysis were very low, estimated from 0.009–0.01. There were also no stop codon, frameshift or non-frameshift mutations observed in the E genes of any lineage. No significant correlations between the accumulation of deleterious mutations or increasing genetic diversity and lineage extinction were observed ( $p > 0.5$ ). Although our hypothesis that accumulation of deleterious mutations over time led to the extinction and replacement of DENV lineages was ultimately not supported by the data, our data does highlight the significant technical issues that must be resolved in the way in which population diversity is measured for DENV and other viruses. The results provide an insight into the within-population genetic structure and diversity of DENV-1 populations.

## Introduction

RNA virus populations are characterised by a high frequency of mutations introduced into their genomes; this ranges from  $10^{-3}$  to  $10^{-5}$  substitutions per nucleotide per round of replication [1, 2]. This effect is compounded by large population sizes and short generation times [3]. The observed high mutation rates are principally due to the error-prone nature of their virally encoded RNA-dependent RNA polymerases (RdRPs) [4]. The resulting RNA virus populations are genetically diverse, thus allowing these viruses to evolve rapidly to fill new ecological niches [5, 6]. For example, mutations in RNA genomes potentially allow RNA viruses to escape from vaccine strategies, antiviral drugs and host immune responses.

The literature suggests that increased genetic diversity facilitates virus adaptation and transmission, its dissemination in new tissues and organs, and infection in new hosts [7–10]. However, most phenotypic mutations in RNA viruses during replication are deleterious [11], and this increases the possibility of extinction during repeated population bottlenecks. RNA viruses, therefore, live on the edge of the “error catastrophe” [12], striking a delicate balance between the genetic diversity required to ensure survival in mutable and hostile host environments and the accumulation of deleterious mutations that would ultimately lead to extinction.

RNA viruses such as the dengue viruses (DENVs) show extensive intra- and inter-host genetic diversity [13]. Regular lineage extinction and replacement is one of the dynamic features of DENV [10, 14], and although most DENV lineage extinction and replacement events are associated with increased incidence of disease, the mechanisms responsible are not well understood. Recent reports of increased diversity in populations of RNA viruses associated with increased fitness [7, 9] appear to argue against most mutations being deleterious. The emergence of defective viruses into a circulating functional DENV population, including non-sense mutations in the envelope (E) gene and defective interfering (DI) viral particles, resulted in increased transmission in nature [15, 16].

DI DENV particles contain RNA genomes with large internal deletions relative to the parental virus. The residual sub-genomic fragment must be replicated in the host cell by complementation with co-infecting functional viruses. The mode of transmission of these co-infecting functional viruses causes DENV populations to frequently experience population bottlenecks. A high proportion of defective genomes within a circulating DENV population would be expected to make that population vulnerable to extinction during bottleneck events. Thus, the accumulation of defective viruses could cause DENV lineages to become extinct. To date, however, no study has determined the impact of DENVs genetic diversity on lineage turnover.

Cloning and Sanger sequencing is the gold standard technique to measure DENV intra- and inter-host genetic diversity [15, 17, 18]. However, this technique is limited by the amount of the genome that can be cloned into a single plasmid for subsequent sequencing, and by the logistics of sequencing large numbers of cloned genome fragments. Previous experiments exploring DENV intra-host genetic diversity were limited to a small set of clones ( $n = 10$  to 30) containing short, amplified segments of the viral genome. Recent advances in sequencing techniques, such as next-generation sequencing (NGS), can potentially overcome these limitations [19–21] through clonally amplifying thousands of DENV template fragments without traditional cloning. NGS should more accurately identify DENVs containing short defective genomes than analogous cloning techniques. However, to our knowledge, no previously reported studies have comprehensively measured the genetic diversity of DENVs from different evolutionary histories by NGS.

The DENVs are single stranded, positive sense RNA viruses consisting of four antigenically distinct serotypes (DENV-1, 2, 3 and 4) belonging to the *Flavivirus* genus within the family *Flaviviridae*. DENV RNA genome is approximately 11Kb and contains a single long open reading

frame (ORF) flanked by the two untranslated regions (5' and 3' UTR). The ORF encodes a poly-protein that is co- and post-translationally cleaved by host and viral proteases into three structural proteins [the capsid protein (C), a membrane associated protein (M), and an envelope protein (E)], and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [22, 23]. The E protein contains domains I, II, and III (which form the ectodomain) and the C-terminal hydrophobic domain (C-term), which has distinct structural functionalities in the folding and assembly of the protein on the surface of the virion and potentiate different antibody responses in humans [21].

To quantify and characterise genetic variations between extinct and circulating lineages, we sequenced the E genes of 13 DENV-1 from four lineages by Roche/454 pyrosequencing. Myanmar DENV-1 isolates were used in this study because the laboratory has had access to a large collection of DENV isolates since 1998. At the time of analysis, lineages A and B had become extinct, and lineages C and D continued to circulate. Lineage A contained the first DENV-1 isolate; this became extinct in 1998—about the same time that lineages B and C appeared. No examples of lineage B have been recovered since 2002; however, lineage C was still circulating in 2008. Lineage D, first detected in 2006, was also still circulating in 2008 [24].

## Materials and Methods

### Study population

Viruses were recovered from acute phase sera from dengue patients admitted to the Yangon Children's Hospital [25]. Strains of DENV-1 used in the study (and described in Table 1) were passaged once in C6/36 before sequencing. A DENV-1 infectious clone (I.C) was engineered and kindly provided by Dr. Wen Lu (The Army Malaria Institute, Australia). The study was approved by the Queensland University of Technology Research Ethics Unit (Ethics No. 0700000910). As no patient tissue was employed in this study, the University Ethics Unit did not require informed patient consent. All patient identifiers were removed from the dengue virus samples before their use in the research.

**Table 1. Strains of DENV-1 used.**

Strain	Country	Date of Isolation	Accession number	Source	Roche/454 pyrosequencing accession number <sup>a</sup>	Source	Passage number
31459	Myanmar	1998	AY588272	[25]	ERS901694	This study	P1 in C6/36
31987	Myanmar	1998	AY588273	[25]	ERS901695	This study	P1 in C6/36
32514	Myanmar	1998	AY600860	[25]	ERS901696	This study	P1 in C6/36
36957	Myanmar	2000	AY620951	[25]	ERS901697	This study	P1 in C6/36
43826	Myanmar	2001	DQ264966	[15]	ERS901698	This study	P1 in C6/36
44988	Myanmar	2002	AY726552	[24]	ERS901699	This study	P1 in C6/36
47317	Myanmar	2002	KF559253	[24]	ERS901700	This study	P1 in C6/36
47662	Myanmar	2002	DQ265041	[15]	ERS901701	This study	P1 in C6/36
49440	Myanmar	2002	DQ265137	[15]	ERS901702	This study	P1 in C6/36
62690	Myanmar	2005	KF559255	[24]	ERS901703	This study	P1 in C6/36
68417	Myanmar	2007	KF559256	[24]	ERS901704	This study	P1 in C6/36
80579	Myanmar	2009	KF559257	[24]	ERS901705	This study	P1 in C6/36
49440 I.C	Myanmar	2002	KF559254	[24]	ERS901706	This study	1 X BHK, 1 X C6/36

I.C: Infectious clone

<sup>a</sup>: The European Nucleotide Archive

doi:10.1371/journal.pone.0142473.t001

## RNA extraction, RT-PCR and Sanger sequencing

Dengue virus RNA was extracted from 140  $\mu$ l samples of virus using the QIAamp Viral RNA mini kit (Qiagen), according to the manufacturer's instructions. RNA was quantified by spectrophotometer. Complementary DNA (cDNA) was produced from the RNA of DENV-1, using random hexanucleotide primers (Boehringer Mannheim) and Expand reverse transcriptase (Expand RT; Roche). Briefly, 1  $\mu$ l random hexamer primers (200ng/ $\mu$ l) were added to 11  $\mu$ l RNA in a 0.5 ml tube, and the mixture was incubated at 65°C for 5 minutes in a heating block, before being placed on ice for 2 minutes. Four microliters of 5x RT buffer (Roche), 1  $\mu$ l 100mM DTT (Roche), 1  $\mu$ l 10mM dNTPs (Roche), 40 units RNase inhibitor (Roche), and 50 units 1  $\mu$ l Expand RT were added to the tube and the volume made up to 20  $\mu$ l with nuclease free water. RT reactions were incubated at 55°C for 1.5 hours. The primers used for PCR amplification corresponded to a region of the E of DENV-1. These were: D1 843F, 5' ATGCCATAGGAACA TCC 3' and D1 2465R, 5' TTGGTGACAAAAATGCC 3'. Five microliters of 10x Expand high fidelity PCR buffer with 15mM MgCl<sub>2</sub> (Roche), 1  $\mu$ l 10mM dNTP, 2  $\mu$ l forward primer (0.1 $\mu$ mol), 2  $\mu$ l reverse primer (0.1 $\mu$ mol), 3 units Expand high fidelity enzyme mix (Roche), 5  $\mu$ l cDNA, and 34.25  $\mu$ l nuclease free water were mixed to make the total volume of 50  $\mu$ l. PCR was performed using cycling conditions of 94°C for 2 minutes for one cycle followed by 92°C for 30 seconds, 58°C for 40 seconds and 68°C for 2.30 minutes for 10 cycles, 92°C for 30 seconds, 58°C for 30 seconds and 68°C for 3 minutes for 10 cycles, 92°C for 30 seconds, 58°C for 30 seconds and 68°C for 3.30 minutes for 18 cycles run for 39 cycles followed by 68°C for 10 minutes for final extension. PCR products were electrophoresed on 1.0% agarose in 1x TBE buffer, and products of the correct size were gel purified with the MinElute PCR purification kit (Qiagen), according to the manufacturer's instructions. The purified DNA (100 ng per 300 bp of product) was added to 3.2 pmol of oligonucleotide primers (forward and reverse) in a final volume of 12  $\mu$ L. The remaining sequencing reaction was performed by the Australian Genome Research Facility Ltd, Brisbane.

## RT- PCR for NGS

Equal amounts of RNA were used for RT. cDNA was produced from the RNA of DENV-1, as in the protocol described above. The primers were designed and used for PCR amplification corresponding to a region of the prM-E-NS1 of DENV-1. These were: D1-330F, 5' -AAGTGC TACGGGGTTTCAAG 3' (nt 330–350) and D1-3462R, 5' -GACCCTGCAGAGACCATTG 3' (nt 3462–3480). Five microliters of 10x Expand high fidelity PCR buffer with 15mM MgCl<sub>2</sub> (Roche), 1  $\mu$ l 10mM dNTP, 2  $\mu$ l forward primer (0.1 $\mu$ mol), 2  $\mu$ l reverse primer (0.1 $\mu$ mol), 4 units Expand high fidelity enzyme mix (Roche), 6  $\mu$ l cDNA, and 32.5  $\mu$ l nuclease free water were mixed in a 0.6 ml tube to make the total volume of 50  $\mu$ l. PCR cycling conditions were adopted to obtain maximum DNA products. These, was performed using 94°C for 2 minutes for one cycle and then 94°C for 30 seconds, 58°C for 45 seconds and 68°C for 3.50 minutes run up to 40 cycles and 68°C for 10 minutes for final extension. PCR was performed from the same reaction for all DENVs. PCR products were gel purified with the MinElute PCR purification kit (Qiagen), according to the manufacturer's instructions. The purified DNA (~1  $\mu$ g per sample) in a final volume of 10  $\mu$ L was sent for Roche/454 pyrosequencing. The remaining sequencing reaction was performed at Murdoch University, Perth, Australia.

## Roche/454 FLX Sequencing

Genome fragments of individual strains were amplified from the same PCR reaction from equimolar amounts of RNA and submitted for library preparation before subsequent next-generation sequencing (NGS) using 454 Roche FLX Titanium. Briefly, amplicons were sheared using

nebulization to average the fragment size to 800bps. The samples were then taken through library preparations using the rapid library multiplex identifier adaptors in order to enable sample identification following sequencing on the FLX genome sequencer. Samples were then taken through emulsion PCR and run on 1/8th region of a FLX titanium chemistry sequencing run.

## Roche/454 FLX Data Analysis

First, the Roche/454 sequenced reads (study accession number: PRJEB11280) were assessed for any sequencing issues by performing quality checks on the raw sequencing reads obtained with the quality control software, FastQC [26]. A full length phylogenetically relevant consensus sequence (M37726; AY726549, unpublished) generated from this laboratory was used as a reference sequence. The sequenced reads were then mapped to the DENV-1 reference genome M37726 (entire length of 11kbp), using the Burrows-Wheeler Aligner's Smith Waterman Alignment (BWA-SW) with default parameters [27].

Sequences were also aligned with a nucleotide identity threshold of 95% against the consensus reference sequence with the alignment tool MOSAIK (<http://bioinformatics.bc.edu/marthlab/Mosaik>), using the recommended parameters for the 454 data to ensure that there was no bias due to the mapping program used. Reads that were not mapped to the reference assembly were then removed from subsequent analysis. This resulted in 1.15% to 10.5% of the raw sequencing reads being removed, and the range of the mean sequencing depth obtained for the PCR-amplified amplicon region being 122-fold to 1,621-fold per residue across the 13 DENV-1 strains.

To identify variants—particularly single-nucleotide polymorphisms (SNPs)—in each DENV-1 population against the reference M37726 genome assembly, LoFreq [28], which is designed to call rare variants from deep sequencing data, was used. In the reference mode, LoFreq reports the variants identified against the reference assembly used, while the consensus mode (-c option) reports variants occurring in the viral population that differs from the consensus genome constructed from the sequence dataset. Therefore, both modes of LoFreq were used in the analysis, with the default parameters applied. Variants were also identified in each virus populations, using VPhaser [29] to compare with the variants identified by LoFreq. To determine the inferred haplotypes and estimate the haplotype frequencies, QuasiRecomb version 1.2 [30] was used to perform a local haplotype reconstruction across the E gene, applying the conservative method, with no recombination assumed and gaps ignored. Since the PCR amplicon spans the prM-E-NS1 gene region, the haplotype inference for the E gene across all 13 samples only considered all sequenced reads that are mapped overlapping the nucleotide positions 935 to 2,419 of the reference M37726 genome assembly.

## Sequence Alignments and Phylogenetic Analysis

Alignment of the Sanger consensus sequences and haplotype sequences derived from 454 reads was performed using the ClustalW program of the Geneious Pro 6.1 software. The aligned nucleic acid sequences were used to construct a bootstrapping phylogenetic tree, using the Neighbor-joining tree building method and Tamura-Nei genetic distance model of the Geneious Pro 6.1.

## Highlighter Plots

Highlighter plots were used to highlight mismatches between nucleotide sequence haplotypes derived from 454 sequence reads. Highlighter plots were performed by using the Gnuplot program in the Los Alamos National Laboratory website-HIV Sequence Database, <http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter.html>.



## Statistical Analysis

Statistical analysis was performed using the GraphPad Prism software package (GraphPad Software Inc.). Unpaired t-test and Analysis of variance (one way ANOVA) were performed to determine the difference between variables. P values lower than 0.05 were considered statistically significant.

## Results

### Phylogenetic analysis

PrM-E-NS1 genes from a total of thirteen DENV-1 strains isolated from 1998 to 2008, including one DENV-1 Infectious Clone (I.C) that was used as a control, were sequenced by Roche/454 pyrosequencing. The E gene consensus sequences of each strain were also derived from Sanger sequencing. The accession numbers for each strain are shown in [Table 1](#). A phylogenetic tree constructed with the consensus DENV-1 E gene nucleotide sequences obtained from Sanger sequencing, showed five distinct branches [24]. DENV-1 complete E gene haplotypes and haplotype frequencies, detected through Roche/454 pyrosequencing, were reconstructed. Reconstructed haplotypes inferred from the viral populations and their frequencies determined through NGS have been used previously to study populations within viruses [30–37].

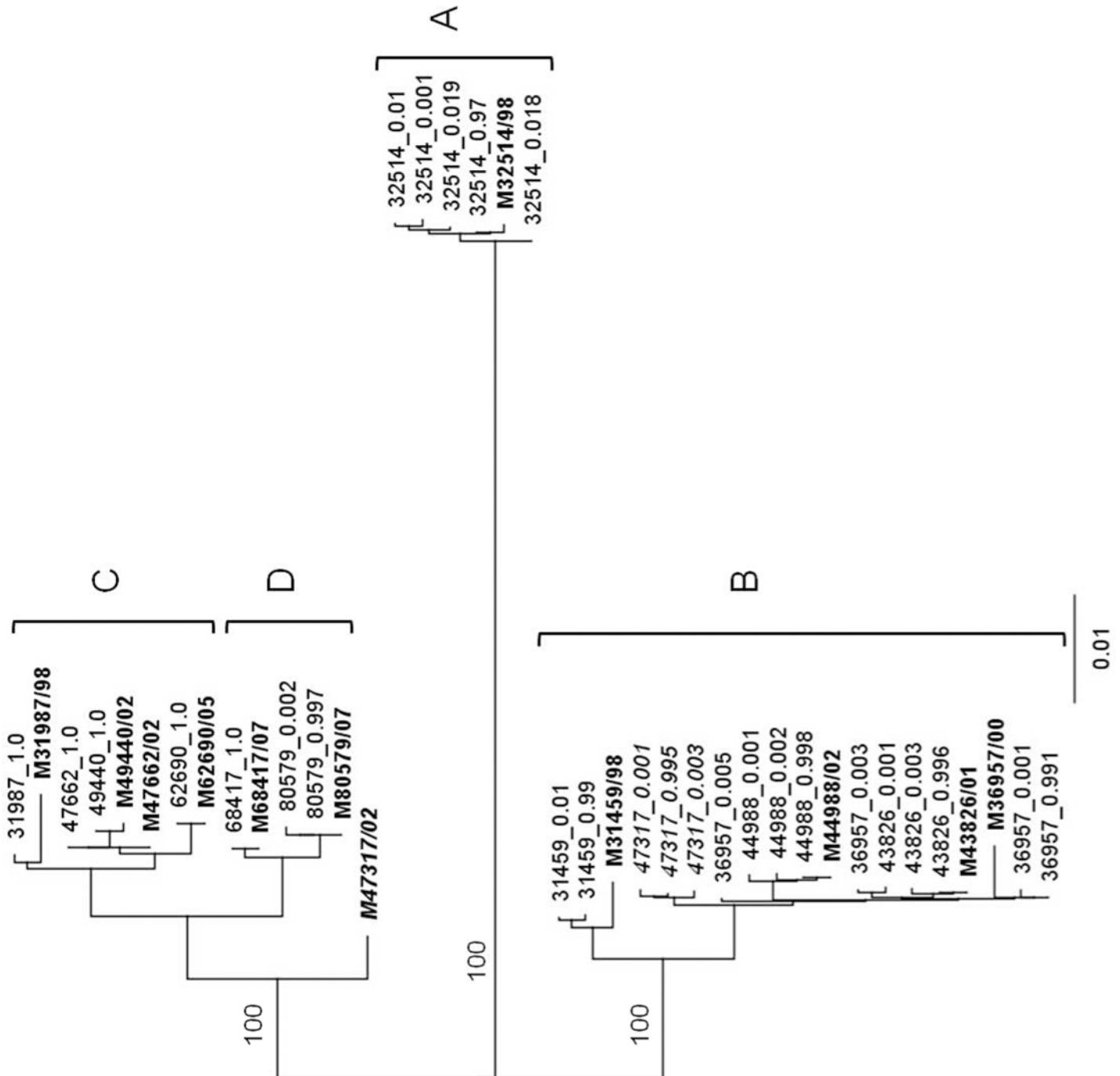
A phylogenetic tree was constructed with the consensus DENV-1 E gene nucleotide sequences obtained from Sanger sequencing and the sequences of each constructed haplotype obtained through Roche/454 pyrosequencing, to determine any disparities ([Fig 1](#)). With the exception of the M47317 strain, each haplotype segregated with the Sanger consensus sequence of the population from which it was drawn ([Fig 1](#)). The intermediate position of the M47317 consensus sequence in this tree suggests that it might have been derived from a recombinant genome ([Fig 2](#)). Alignment of the consensus nucleotide sequences of representative strains from lineage B and C with the consensus sequence for M47317 and the sequences of M47317 haplotypes, suggested a recombination break point at about nucleotide 1087. Phylogenetic trees constructed with sequences from either side of this point supported this conclusion [i.e., nt 1–1087 of the consensus sequence of M47317 segregated with those from lineage C viruses ([Fig 2](#)), while sequences 3' to this point segregated more closely with lineage B viruses and the M47317 haplotypes]. There was no recombination observed within the sequences of M47317 haplotypes with strains from lineage A or D.

### Nature and extent of intra-host genetic diversity in the envelope gene

A combined total of 111,101 raw sequences were acquired from the extinct and skin circulating strains. After quality filtering, an average of 97% of the sequence reads were mapped to the reference (98% were from extinct, and 96% from circulating lineages), using BWA-SW ([S1 Table](#)). To ensure there was no bias due to the mapping program used, we also aligned the sequencing reads using MOSIK. There was no significant difference ( $p > 0.05$ , t test) between the two different programs used ([S1 Table](#)) in the mean coverage and the percentage of mapping.

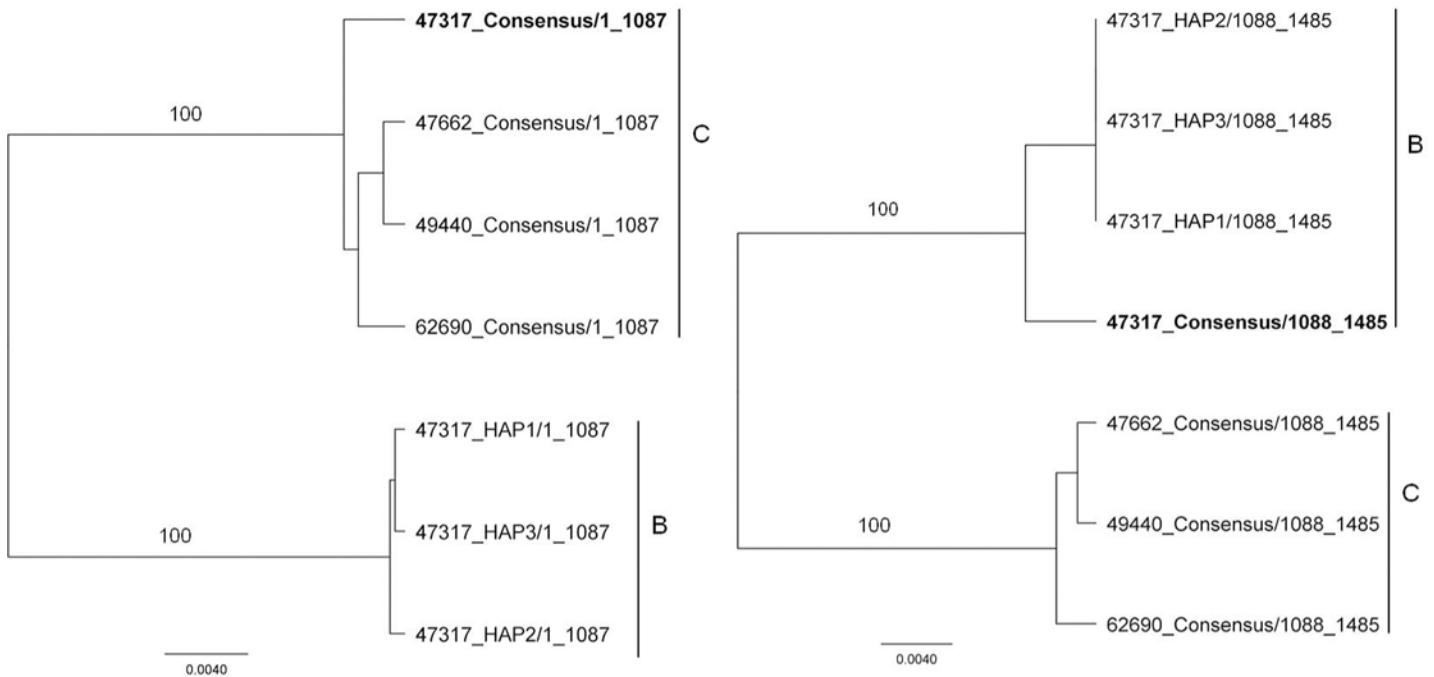
Variants in each DENV-1 strain were identified using LoFreq against reference genome. The number of nucleotide changes and the value of these changes are shown in [Table 2](#). Values for strain M47317 were excluded from analysis because of the disparity between the positions in the phylogenetic tree of the haplotypes and the Sanger consensus sequences. DENV-1 I.C was used as control for this analysis. The total variants and variable sites in DENV-1 populations determined by Roche/454 pyrosequencing were greater than DENV-1 I.C. The nucleotide sequence diversity within viral populations was very low, and estimates of the average diversity ranged from 0.009–0.01. There was a significant difference ( $p < 0.03$ , t-test) in variants

identified between extinct (A, B) and extant (C, D) lineages, however, no difference ( $p > 0.05$ , t-test) for any other parameters measured in Table 2. While 1.15% to 10.5% of the sequencing reads could not be assembled, there was no significant difference ( $p > 0.16$ , t-test) observed in extinct and circulating lineages (S1 Table).



**Fig 1. Phylogenetic analysis of the E gene haplotype and consensus sequences of DENV-1.** Consensus sequences for each strain have been highlighted. Bootstrap values (100 replications) for key nodes are shown. A distance bar is shown below the tree. Lineage A became extinct in 1998 about the same time lineages B and C appeared. No examples of lineage B have been recovered since 2002, but lineage C still circulated in 2008. Lineage D, first detected in 2006, also still circulated in 2008. For the purposes of analysis, lineages A and B are considered to have become extinct in 1998 and 2002, respectively. Lineages C and D were deemed to be still circulating in 2008.

doi:10.1371/journal.pone.0142473.g001



**Fig 2. Phylogenetic analyses of nt 1–1087 and nt 1088–1485 of the consensus sequences of the E genes of DENV-1 from lineages B and C and of the sequences of the same regions of the E genes of M47317 haplotypes.** Bootstrap values (100 replicates) for key nodes are shown.

doi:10.1371/journal.pone.0142473.g002

**Table 2. Measures of intra-host genetic diversity among DENV-1 lineages.** Genetic diversity was measured through SNPs analysis.

DENV-1 lineage	Name of the lineage	Strain	Total reads analysed	Mean Coverage/nt	Total nt variants <sup>a</sup>	Total number of variable sites		Est. mean genetic diversity (%) <sup>c</sup>	
						n.t	a.a	n.t	a.a
Extinct	A	M32514	5919	454.20	130	10	0	0.023	n.a
		M31459	8330	736.47	51	0	0	n.a	n.a
	M36957	12127	1091.10	57	1	0	0.011	n.a	
	M43826	3294	273.73	57	0	0	n.a	n.a	
	M44988	15211	1334.72	59	0	0	n.a	n.a	
Survived	C	M31987	17729	1621.37	5	0	0	n.a	n.a
		<i>M47317<sub>b</sub></i>	9333	824.04	56	0	0	n.a	n.a
		M47662	6300	561.02	4	1	0	0.009	n.a
	M49440	13439	1217.48	5	0	0	n.a	n.a	
	M62690	3126	287.08	7	1	0	0.002	n.a	
	D	M68417	8635	771.78	19	0	0	n.a	n.a
		M80579	1384	122.77	22	1	0	0.010	n.a
Control	I.C	6274	697.80	15	0	0	n.a	n.a	

n.a; not applicable

<sup>a</sup> The number of bases that are different from the reference genome.

<sup>b</sup> Genetic diversity of recombinant DENV-1 strain showing in italic.

<sup>c</sup> Estimated mean genetic diversity =  $\frac{\text{TotalVariants}}{E(\text{Depth}) \times \text{Length}} = \frac{\sum_{i=1}^n \text{FreqVariant}_i \times \text{Depth}_i}{\sum_{i=1}^n \text{Depth}_i \times \text{Length}}$  where  $n$  is the total number of variants.

doi:10.1371/journal.pone.0142473.t002



**Table 3. Detecting variants, stop codons and deletions using two different sequence analysis platforms.**

Lineage Status	Strain	Year	Detected by LoFreq				Detected by VPhaser			
			Total Variants <sup>a</sup>	Stop Codons	Frameshift	Non-Frameshift	Total Variants <sup>a</sup>	Stop Codons	Frameshift	Non-Frameshift
Extinct	32514	1998	130	0	0	0	148	0	4	0
	31459	1998	51	0	0	0	60	1	7	1
	36957	2000	56	0	0	0	62	1	9	2
	43826	2001	57	0	0	0	62	0	10	0
	44988	2002	59	0	0	0	68	1	2	0
Circulating	31987	1998	5	0	0	0	12	0	10	1
	<i>47317<sup>b</sup></i>	2002	56	0	0	0	58	1	7	1
	47662	2002	3	0	0	0	4	0	0	0
	49440	2002	5	0	0	0	21	0	7	2
	62690	1995	6	0	0	0	12	0	1	0
	68417	2007	19	0	0	0	74	0	13	0
	80579	2008	21	0	0	0	22	0	0	0
	Control	I. C	-	15	0	0	0	15	0	0

<sup>a</sup> The number of bases that are different from the reference genome.

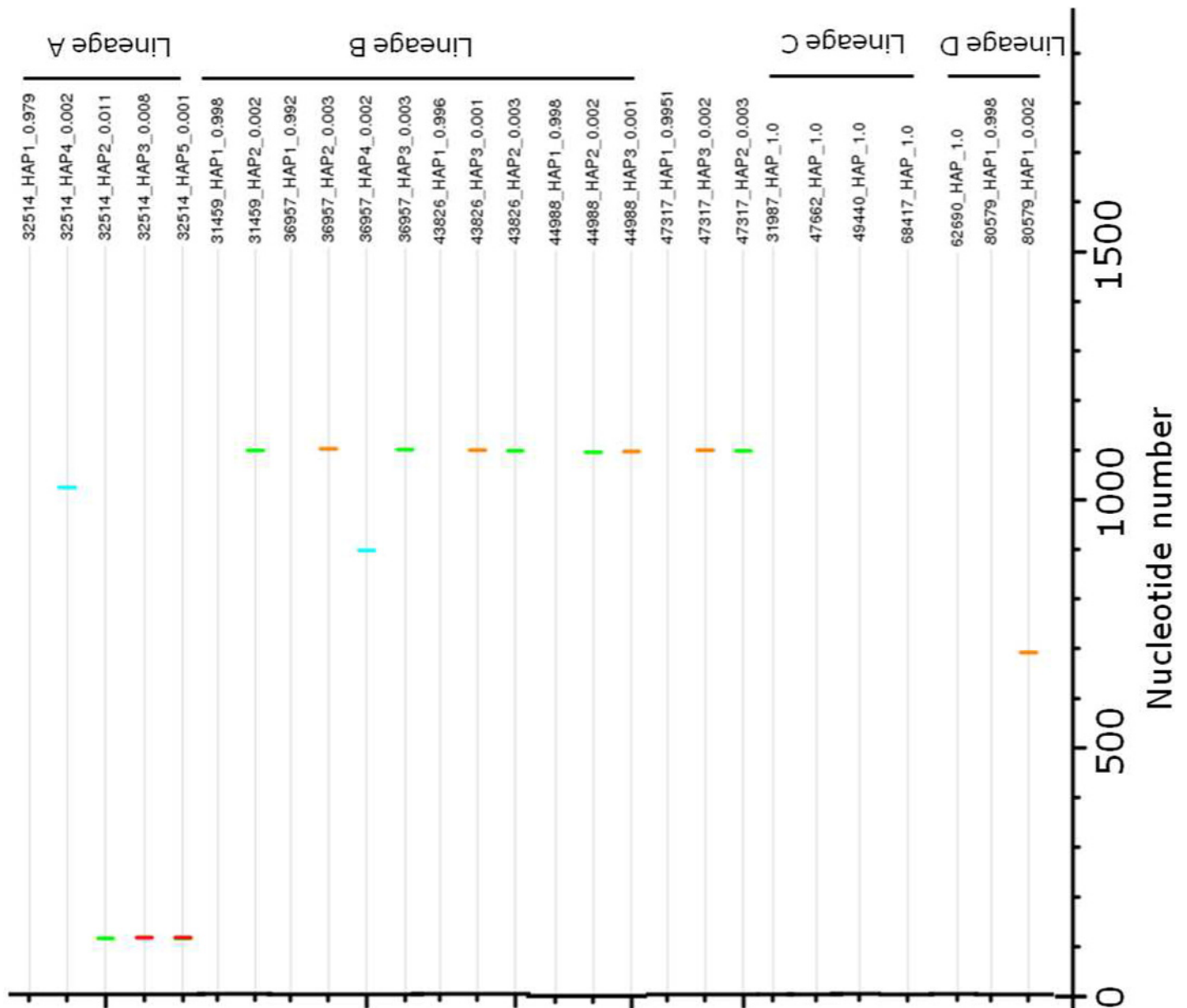
<sup>b</sup> Genetic diversity of recombinant DENV-1 strain showing in italic.

doi:10.1371/journal.pone.0142473.t003

We identified variable sites in the viral community of both lineages; however, there was no trend observed (Table 3). Recent studies showed that the defective viruses that contained intra-genic stop codon mutations, which are most likely to be deleterious, can be transmitted in nature through complementation with functional virus [15]. While we hypothesised that an accumulation of deleterious mutations over time led to the extinction and replacement of DENV lineages, there were no stop codon, frameshift and non-frameshift mutations observed in the E genes of any lineage in this study. Nor were there any insertions or deletions observed in the E genes of any lineage.

To identify any bias in the results due to the analysis platform used, we also identified variants using VPhaser. There was a significant difference ( $p < 0.001$ , one way ANOVA) in the results observed between two platform used (Table 3). Few strains in both lineages presented a stop codon; however, this was a result of a homopolymer associated with an alignment error at position 1046 A->T, K38. Although frameshift and non-frameshift mutations were detected in both lineages using VPhaser, there was no significant difference ( $p > 0.05$ , t-test) identified between the two lineages (Table 3).

The distribution and the nature of the changes (relative to the sequence of the most common haplotype/consensus) within the E gene are shown in Fig 3. Few strains in either lineage displayed polymorphisms in their complete E genes. Overall, only 1 to 5 variable sites were observed. No polymorphic positions between extinct and extant lineages were conserved, and there were no conserved substitutions (e.g., A or T or G or C) that were exclusive to either extinct or extant lineages. The most polymorphic sites were observed between nucleotides 800 to 1100 in an extinct lineage; however, the frequency of these polymorphisms between extinct and extant lineages was not significantly different ( $p > 0.69$ , t test). Studies have shown that the DENV E domain (D) II contains the highest diversity of all viral genome domains [21]. However, in this study, no domain (DI-DIII) bias in conservative and non-conservative changes between extinct and extant lineages was observed.



**Fig 3. Distribution of polymorphic sites in the E genes of haplotypes of DENV-1 Lineage A, B, C and D.** Changes from the most prevalent haplotypes/consensus are shown as A: green, T: red, G: orange, C: light blue. The percentage of each haplotype in the population is shown in parenthesis.

doi:10.1371/journal.pone.0142473.g003

## Discussion

RNA virus populations must strike a fine balance between the genetic diversity necessary to survive in changing host environments [9] and the genetic diversity leading to error catastrophe. Descriptions of the nature and magnitude of genetic diversity in DENV populations [15, 17, 38–40] have suffered from two technological constraints: the limited size of the genome sequence that can be cloned into a single plasmid for subsequent sequencing, and the logistics of sequencing statistically large numbers of individually cloned genome fragments.

We investigated intra-host genetic diversity of DENV-1 using NGS technology as it has the potential to address both issues. As this technology has recently become routinely available in

the field, a study to determine whether its use is appropriate for the determination of intra-host genetic diversity is timely and essential. To explore this question, we analysed DENV-1 E gene sequences from each of the populations chosen for this study. The DENV-1.I.C clone E gene was chosen as the unit of analysis by NGS, and was also Sanger sequenced to identify potential sequencing errors and artefacts that might be inherent in the NGS approach.

Our data are largely consistent with those reported in the literature. The genetic diversity in DENV-1 populations determined by 454 pyrosequencing was greater than control I.C, but was less than the previous estimates determined by conventional cloning and sequencing [17, 40, 41]. This observation is consistent with a recent study [20, 21] showing that the nature and magnitude of intra-host genetic diversity in DENV-2 determined by 454 sequencing was significantly lower than previous estimates determined by cloning and sequencing [17, 40, 41].

There was no correlation observed between the nature and extent of the genetic diversity of the DENV-1 E genes and the observed extinction or survival of the lineage from which they were drawn ( $p > 0.5$ ). There were no stop codon, frameshift and non-frameshift mutations observed in the E genes of any lineage. There were no lineage-specific nucleotide substitutions or changes in the DENV E gene that were likely to have contributed to lineage extinction, nor were there substitutions or changes identified that were likely to have provided a fitness advantage to the circulating lineages.

Recently, the DENV E domain (D) II was shown to contain the highest diversity of all viral genome domains [21]. Because E (D) II is highly immunogenic in humans [42], the higher genetic diversity in E DII was possibly due to immune-driven pressure. However, no significant differences in genetic diversity in envelope protein domains that were likely to cause lineage extinction were detected. This is consistent with previous studies [43, 44] that were also unable to detect specific changes in the DENV genomes that might be responsible for lineage extinction. Although no conserved changes were observed in extinct lineages relative to circulating lineages, a few polymorphic sites were identified in all populations of lineage B, particularly at E1083 and E1085. These are synonymous mutations, and do not have any effect on the function of the protein that can result in lineage extinction. However, the nature and magnitude of the genetic diversity of the E genes, determined by 454 sequencing of the DENV-1 in this study (Table 2), was 10- to 100- fold less than previously published reports of DENV diversity [15, 17, 40, 41], and this might have affected the subsequent analysis.

This observation of the nature and extent of intra-host genetic diversity in DENV-1 is consistent with recent reports of the diversity determined by 454 pyrosequencing [20, 21] and by cloning and Sanger sequencing [40], but significantly lower than some previous estimates determined by cloning and Sanger sequencing [15, 17]. Twenty clones from strain M47317 contained 78 polymorphic sites and 1 stop codon, and 20 clones of M49440 contained 89 polymorphic sites and 11 stop codons [15]. The same strains analysed by NGS contained 4 (M47317) and 2 (M49440) polymorphic sites and no stop codons. However, these comparisons have been employing from cDNA generated from different stocks of virus using different primers.

There is a need for comparisons of diversity determined by NGS and by cloning and sequencing of the same cDNA stocks. There is also the need for all NGS sequencing reads to be included in analyses of SNPs rather than only those that can be assembled into complete E gene sequences. However, we acknowledge that an appropriate comparison between pyrosequencing and cloning and Sanger might not be possible due to the amount of the genome that can be cloned, and the difference in platforms and parameters/filters used for sequence analysis. The NGS FLX (454 Life Science, Roche) technology was chosen for this study because it provides longer sequence lengths (average of 300 bases) than other platforms, demonstrates 99.5% accuracy in single-read, and exhibits a very low incidence of substitution errors (<10–6)

[45, 46]. Because of the algorithms used to assemble the short reads generated from 454 sequencing and reconstruct haplotypes of the assembled sequences, the possibility of excluding genuine diversity must be considered. Less than 1.15% to 10.5% of sequence reads could not be assembled with the algorithms employed. Therefore, some diversity could be excluded by the 10% (upper estimate) of sequences that left out of contigs because they are 'error'. However, there was no trend observed in the proportion of sequences that were excluded in extinct and extant lineages (S1 Table).

As reported in Romano's study [20], SNPs (A to G) that are detected in the standard cloning and Sanger sequencing also appeared in 454 sequencing reads (at frequencies about 0.4%), but in homopolymer regions of the 454 sequencing reads. Therefore, the SNPs were excluded from the deep sequencing analysis. This was due to the thresholds/filters adopted in the analysis to consider a variant to be true (1% frequency and quality score above 30). However, genetic diversity that excluded when algorithm removes bases with low quality scores [e.g. hypermutated sequence (sequences with more than 5 nt mutations within the first 20 positions of the read in either of the ends); reads with indels that resulted in a frame-shift or variants with higher allele frequencies (5%)] is likely the sequencing errors. Additionally, the algorithms might consider singletons (a class of polymorphism) as errors, and subsequently "correct" the errors, despite the fact that some mutations might represent true biological variants [40].

In this study, variants that identified in each virus populations using the VPhaser analysis platform were compared with the variants identified by the LoFreq platform. There was a significant difference ( $p < 0.001$ ) in the results observed between two platform used (Table 3). We did not identify any stop codon mutation when using the LoFreq platform analysis. While a stop codon mutation appeared in both lineages DENV-1 populations when analysed with VPhaser (Table 3), this mutation appeared in a homopolymer region and was unlikely to represent a true biological variant. While there was no deletion observed when analysed using LoFreq platform, deletions were detected when analysed using VPhaser. This highlighted that there was some discrepancy between sequence diversity estimated by NGS analysis platforms, and indicates the need to evaluate these platforms in order to comprehensively determine intra-host viral genetic diversity.

To our knowledge, this is the first study to investigate the intra-host genetic diversity in DENV-1 E gene by NGS technology. The DENV E gene was chosen based on the considerable body of evidence suggesting that the nature and magnitude of the diversity in DENV E genes is similar to that in most other regions of the open reading frame, and is therefore representative of genomic diversity [21, 39, 43]. Most data relating to diversity in DENV has thus far been obtained from the E gene. This allows for a more comprehensive comparison of our data with the data in the existing literature.

DENV frequently experiences intra- and inter-host population bottlenecks during transmission. The substructure of DENV populations could substantially impact the effects of bottleneck transmission due to the phenotypic variability within the population [24]. We hypothesised that the accumulation of deleterious mutations over time led to the extinction and replacement of DENV lineages. Although this hypothesis was ultimately not supported by our data, the data does provide insights into the DENV population substructures, and extends our understanding of their complexity. Finally, our experimental design revealed a glaring discrepancy between sequence diversity estimated by NGS analyses and the diversity estimated by the classic method of cloning and Sanger sequencing. For NGS to be considered reliable for diversity analyses, this discrepancy must be reconciled.

## Supporting Information

**S1 Table. Mapping 454 pyrosequencing reads using two different sequence analysis programs.**  
(DOCX)

## Acknowledgments

The work was supported by QUT Postgraduate Research Award, Queensland University of Technology, Australia. The authors sincerely thank Dr. Francesca Frentiu for their valuable comments on previous version of this manuscript.

## Author Contributions

Conceived and designed the experiments: MAC WBL JA. Performed the experiments: MAC. Analyzed the data: MAC SB AYC YYT RTHO. Contributed reagents/materials/analysis tools: JA MAC AYC YYT RTHO WBL. Wrote the paper: MAC WBL JA.

## References

1. Domingo E, Flavell RA, Weissmann C. *In vitro* site-directed mutagenesis: generation and properties of an infectious extracistronic mutant of bacteriophage Qbeta. *Gene*. 1976; 1(1):3–25. PMID: [1052322](#)
2. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S. Rapid evolution of RNA genomes. *Science*. 1982; 215(4540):1577–85. doi: [10.1126/science.7041255](#) PMID: [7041255](#)
3. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol*. 1997; 51(1):151–78.
4. Steinhauer DA, Holland JJ. Rapid evolution of RNA viruses. *Annu Rev Microbiol*. 1987; 41:409–33. Epub 1987/01/01. doi: [10.1146/annurev.mi.41.100187.002205](#) PMID: [3318675](#).
5. Domingo E, Holland JJ. Mutation rates and rapid evolution of RNA viruses. Lippincott Williams & Wilkins; 1994. p. 161.
6. Holland JJ, De La Torre JC, Steinhauer DA. RNA virus populations as quasispecies. Springer; 1992. p. 1–20.
7. Pfeiffer JK, Kirkegaard K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog*. 2005; 1(2):e11. PMID: [16220146](#)
8. Coffey LL, Beeharry Y, Bordería AV, Blanc H, Vignuzzi M. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc Natl Acad Sci U S A*. 2011; 108(38):16038–43. doi: [10.1073/pnas.1111650108](#) PMID: [21896755](#)
9. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*. 2006; 439(7074):344–8. PMID: [16327776](#)
10. OhAinle M, Balmaseda A, Macalalad AR, Tellez Y, Zody MC, Saborio S, et al. Dynamics of dengue disease severity determined by the interplay between viral genetics and serotype-specific immunity. *Sci Transl Med*. 2011; 3(114):114ra28. Epub 2011/12/23. doi: [10.1126/scitranslmed.3003084](#) PMID: [22190239](#); PubMed Central PMCID: PMC129678.
11. Duarte EA, Novella IS, Ledesma S, Clarke DK, Moya A, Elena SF, et al. Subclonal components of consensus fitness in an RNA virus clone. *J Virol*. 1994; 68(7):4295–301. PMID: [8207804](#)
12. Eigen M. Error catastrophe and antiviral strategy. *Proc Natl Acad Sci U S A*. 2002; 99(21):13374–6. Epub 2002/10/09. doi: [10.1073/pnas.212514799](#) PMID: [12370416](#); PubMed Central PMCID: PMC129678.
13. Holmes EC, Burch SS. The causes and consequences of genetic variation in dengue virus. *Trends Microbiol*. 2000; 8(2):74–7. PMID: [10664600](#)
14. Holmes EC, Twiddy SS. The origin, emergence and evolutionary genetics of dengue virus. *Infect Genet Evol*. 2003; 3(1):19–28. PMID: [12797969](#)
15. Aaskov J, Buzacott K, Thu HM, Lowry K, Holmes EC. Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science*. 2006; 311(5758):236–8. doi: [10.1126/science.1115030](#) PMID: [16410525](#)

16. Li D, Lott WB, Lowry K, Jones A, Thu HM, Aaskov J. Defective interfering viral particles in acute dengue infections. *PLoS One*. 2011; 6(4):e19447. Epub 2011/05/12. doi: [10.1371/journal.pone.0019447](https://doi.org/10.1371/journal.pone.0019447) PMID: [21559384](https://pubmed.ncbi.nlm.nih.gov/21559384/); PubMed Central PMCID: PMC3084866.
17. Wang WK, Lin SR, Lee CM, King CC, Chang SC. Dengue type 3 virus in plasma is a population of closely related genomes: quasispecies. *J Virol*. 2002; 76(9):4662–5. Epub 2002/04/05. PMID: [11932434](https://pubmed.ncbi.nlm.nih.gov/11932434/); PubMed Central PMCID: PMC155095.
18. Wang W-K, Sung T-L, Lee C-N, Lin T-Y, King C-C. Sequence diversity of the capsid gene and the non-structural gene NS2B of dengue-3 virus in vivo. *Virology*. 2002; 303(1):181–91. PMID: [12482670](https://pubmed.ncbi.nlm.nih.gov/12482670/)
19. Astrovskaya I, Tork B, Mangul S, Westbrook K, Mändou I, Balfe P, et al. Inferring viral quasispecies spectra from 454 pyrosequencing reads. *BMC Bioinformatics*. 2011; 12(Suppl 6):S1. doi: [10.1186/1471-2105-12-S6-S1](https://doi.org/10.1186/1471-2105-12-S6-S1) PMID: [21989211](https://pubmed.ncbi.nlm.nih.gov/21989211/)
20. Romano CM, Lauck M, Salvador FS, Lima CR, Villas-Boas LS, Araújo ESA, et al. Inter- and intra-host viral diversity in a large seasonal DENV2 outbreak. *PLoS One*. 2013; 8(8):e70318. doi: [10.1371/journal.pone.0070318](https://doi.org/10.1371/journal.pone.0070318) PMID: [23936406](https://pubmed.ncbi.nlm.nih.gov/23936406/)
21. Parameswaran P, Charlebois P, Tellez Y, Nunez A, Ryan EM, Malboeuf CM, et al. Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and inter-host diversity. *J Virol*. 2012; 86(16):8546–58. Epub 2012/06/01. doi: [10.1128/jvi.00736-12](https://doi.org/10.1128/jvi.00736-12) PMID: [22647702](https://pubmed.ncbi.nlm.nih.gov/22647702/); PubMed Central PMCID: PMC3421746.
22. Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol*. 1990; 44:649–88. Epub 1990/01/01. doi: [10.1146/annurev.mi.44.100190.003245](https://doi.org/10.1146/annurev.mi.44.100190.003245) PMID: [2174669](https://pubmed.ncbi.nlm.nih.gov/2174669/).
23. Rice CM, Strauss E. G and Strauss J. H. *Flavivirus genome organization, expression, and replication. Togaviridae and Flaviviridae*. 1986:pp. 279–326 Edited by Schlesinger S. and Schlesinger M.J Plenum, New York.
24. Choudhury MA, Lott WB, Aaskov J. Distribution of fitness in populations of dengue viruses. *PLoS One*. 2014; 9(9):e107264. doi: [10.1371/journal.pone.0107264](https://doi.org/10.1371/journal.pone.0107264) PMID: [25222471](https://pubmed.ncbi.nlm.nih.gov/25222471/)
25. Thu HM, Lowry K, Myint TT, Shwe TN, Han AM, Khin KK, et al. Myanmar dengue outbreak associated with displacement of serotypes 2, 3, and 4 by dengue 1. *Emerg Infect Dis*. 2004; 10(4):593–97. PMID: [15200847](https://pubmed.ncbi.nlm.nih.gov/15200847/)
26. Andrews S. FastQC: A quality control tool for high throughput sequence data. Available: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2014.
27. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010; 26(5):589–95. doi: [10.1093/bioinformatics/btp698](https://doi.org/10.1093/bioinformatics/btp698) PMID: [20080505](https://pubmed.ncbi.nlm.nih.gov/20080505/)
28. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. 2012; 40(22):11189–201. doi: [10.1093/nar/gks918](https://doi.org/10.1093/nar/gks918) PMID: [23066108](https://pubmed.ncbi.nlm.nih.gov/23066108/)
29. Macalalad AR, Zody MC, Charlebois P, Lennon NJ, Newman RM, Malboeuf CM, et al. Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data. *PLoS Comput Biol*. 2012; 8(3):e1002417. doi: [10.1371/journal.pcbi.1002417](https://doi.org/10.1371/journal.pcbi.1002417) PMID: [22438797](https://pubmed.ncbi.nlm.nih.gov/22438797/)
30. Zagordi O, Töpfer A, Prabhakaran S, Roth V, Halperin E, Beerenwinkel N, editors. *Probabilistic inference of viral quasispecies subject to recombination. Research in Computational Molecular Biology*; 2012: Springer.
31. Zagordi O, Geyrhofer L, Roth V, Beerenwinkel N. Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. *J Comput Biol*. 2010; 17(3):417–28. Epub 2010/04/10. doi: [10.1089/cmb.2009.0164](https://doi.org/10.1089/cmb.2009.0164) PMID: [20377454](https://pubmed.ncbi.nlm.nih.gov/20377454/).
32. Zagordi O, Klein R, Daumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res*. 2010; 38(21):7400–9. Epub 2010/07/31. doi: [10.1093/nar/gkq655](https://doi.org/10.1093/nar/gkq655) PMID: [20671025](https://pubmed.ncbi.nlm.nih.gov/20671025/); PubMed Central PMCID: PMC2995073.
33. Bull RA, Eden J-S, Luciani F, McElroy K, Rawlinson WD, White PA. Contribution of Intra- and Interhost Dynamics to Norovirus Evolution. *J Virol*. 2012; 86(6):3219–29. doi: [10.1128/jvi.06712-11](https://doi.org/10.1128/jvi.06712-11) PMID: [22205753](https://pubmed.ncbi.nlm.nih.gov/22205753/)
34. Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog*. 2011; 7(9):e1002243. Epub 2011/09/14. doi: [10.1371/journal.ppat.1002243](https://doi.org/10.1371/journal.ppat.1002243) PMID: [21912520](https://pubmed.ncbi.nlm.nih.gov/21912520/); PubMed Central PMCID: PMC3164670.
35. Eriksson N, Pachter L, Mitsuya Y, Rhee S, Wang C, Gharizadeh B, et al. Viral population estimation using pyrosequencing. *PLoS Comput Biol*. 2008; 4(4):e1000074. doi: [10.1371/journal.pcbi.1000074](https://doi.org/10.1371/journal.pcbi.1000074) PMID: [18437230](https://pubmed.ncbi.nlm.nih.gov/18437230/)



36. Zagordi O, Bhattacharya A, Eriksson N, Beerenwinkel N. ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. *BMC Bioinformatics*. 2011; 12(1):119. doi: [10.1186/1471-2105-12-119](https://doi.org/10.1186/1471-2105-12-119)
37. Poh W-T, Xia E, Chin-inmanu K, Wong L-P, Cheng AY, Malasit P, et al. Viral quasispecies inference from 454 pyrosequencing. *BMC Bioinformatics*. 2013; 14(1):355.
38. Worobey M, Rambaut A, Holmes EC. Widespread intra-serotype recombination in natural populations of dengue virus. *Proc Natl Acad Sci U S A*. 1999; 96(13):7352–57. PMID: [10377418](https://pubmed.ncbi.nlm.nih.gov/10377418/)
39. Zhang C, Mammen MP Jr., Chinnawirotpisan P, Klungthong C, Rodpradit P, Nisalak A, et al. Structure and age of genetic diversity of dengue virus type 2 in Thailand. *J Gen Virol*. 2006; 87(4):873–83. doi: [10.1099/vir.0.81486-0](https://doi.org/10.1099/vir.0.81486-0)
40. Thai KTD, Henn MR, Zody MC, Tricou V, Nguyet NM, Charlebois P, et al. High-resolution analysis of intrahost genetic diversity in dengue virus serotype 1 infection identifies mixed infections. *J Virol*. 2012; 86(2):835–43. doi: [10.1128/jvi.05985-11](https://doi.org/10.1128/jvi.05985-11) PMID: [22090119](https://pubmed.ncbi.nlm.nih.gov/22090119/)
41. Chao DY, King CC, Wang WK, Chen WJ, Wu HL, Chang GJ. Strategically examining the full-genome of dengue virus type 3 in clinical isolates reveals its mutation spectra. *Virol J*. 2005; 2:72. Epub 2005/08/27. doi: [10.1186/1743-422x-2-72](https://doi.org/10.1186/1743-422x-2-72) PMID: [16120221](https://pubmed.ncbi.nlm.nih.gov/16120221/); PubMed Central PMCID: PMC1208963.
42. Lai C-Y, Tsai W-Y, Lin S-R, Kao C-L, Hu H-P, King C-C, et al. Antibodies to Envelope Glycoprotein of Dengue Virus during the Natural Course of Infection Are Predominantly Cross-Reactive and Recognize Epitopes Containing Highly Conserved Residues at the Fusion Loop of Domain II. *J Virol*. 2008; 82(13):6631–43. doi: [10.1128/jvi.00316-08](https://doi.org/10.1128/jvi.00316-08) PMID: [18448542](https://pubmed.ncbi.nlm.nih.gov/18448542/)
43. Thu HM, Kym L, Jiang L, Hlaing T, Holmes EC, Aaskov J. Lineage extinction and replacement in dengue type 1 virus populations are due to stochastic events rather than to natural selection. *Virology*. 2005; 336(2):163–72. PMID: [15892958](https://pubmed.ncbi.nlm.nih.gov/15892958/)
44. Wittke V, Robb TE, Thu HM, Nisalak A, Nimmannitya S, Kalayanrooj S, et al. Extinction and rapid emergence of strains of dengue 3 virus during an Interepidemic period. *Virology*. 2002; 301(1):148–56. PMID: [12359455](https://pubmed.ncbi.nlm.nih.gov/12359455/)
45. Droege M, Hill B. The Genome Sequencer FLX System—longer reads, more applications, straight forward bioinformatics and more complete data sets. *J Biotechnol*. 2008; 136(1–2):3–10. Epub 2008/07/12. doi: [10.1016/j.jbiotec.2008.03.021](https://doi.org/10.1016/j.jbiotec.2008.03.021) PMID: [18616967](https://pubmed.ncbi.nlm.nih.gov/18616967/).
46. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012; 2012:251364. Epub 2012/07/26. doi: [10.1155/2012/251364](https://doi.org/10.1155/2012/251364) PMID: [22829749](https://pubmed.ncbi.nlm.nih.gov/22829749/); PubMed Central PMCID: PMC3398667.