

1 High-resolution analysis of intra-host genetic diversity in dengue 2 virus serotype 1 identifies mixed infection

3

4 Khoa T.D. Thai^{1,3,4**}, Matthew R. Henn^{6*}, Michael C. Zody⁶, Vianney Tricou¹, Nguyen Minh
5 Nguyet¹, , Patrick Charlebois⁶, Niall J. Lennon⁶, Lisa Green⁶, Peter J. de Vries^{3,4}, Tran Thinh
6 Hien¹, Jeremy Farrar^{1,2}, H. Rogier van Doorn^{1,2}, Menno D. de Jong^{4,5}, Bruce W. Birren⁶,
7 Edward C. Holmes^{7,8}, Cameron P. Simmons^{1,2}

8

9 ¹ Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh
10 City, Vietnam; ² Center for Tropical Medicine, Nuffield Department of Clinical Medicine,
11 University of Oxford, Center for Clinical Vaccinology and Tropical Medicine, Oxford, UK; ³
12 Division of Infectious Diseases, Tropical Medicine & AIDS, Academic Medical Center,
13 Amsterdam, the Netherlands; ⁴ Center for Infection and Inflammation Research
14 (CINIMA), Academic Medical Center, University of Amsterdam, Amsterdam, the
15 Netherlands; ⁵ Division of Clinical Virology, Academic Medical Center, Amsterdam, the
16 Netherlands; ⁶ Broad Institute of MIT & Harvard, 7 Cambridge Center, Cambridge, MA
17 02142, USA; ⁷ Center for Infectious Disease Dynamics, Pennsylvania State University,
18 University Park; ⁸ Fogarty International Center, National Institutes of Health, Bethesda,
19 Maryland.

20 * These authors contributed equally to this work

21 [†] Corresponding author: Khoa T.D. Thai, Oxford University Clinical Research Unit,
22 Hospital for Tropical Diseases, 190 Ben Ham Tu, District 5, Ho Chi Minh City, Vietnam.
23 Tel: +84 8 9237954; Fax: +84 8 9238904; E-mail: khoat@oucru.org; Division of
24 Infectious Diseases, Tropical Medicine & AIDS, Academic Medical Center, F4-217, PO Box
25 22700, 1100 DE Amsterdam, the Netherlands. Tel.: +31 20 5664380; Fax.: +31 20
26 6972286; Email: t.d.thai@amc.uva.nl

27 Abstract word count: 219
28 Text word count: 4398
29 Tables: 3
30 Figures: 3
31 Supplement tables: 2
32
33 Running title: Intra-host genetic diversity of DENV-1

34 **ABSTRACT**

35 Little is known about the rate at which genetic variation is generated within intra-host
36 populations of dengue virus (DENV) and what implications this diversity has for dengue
37 pathogenesis, disease severity, and host immunity. Previous studies of intra-host DENV
38 variation have used a low frequency of sampling and/or experimental methods that do
39 not fully account for errors generated through amplification and sequencing of viral
40 RNAs. We investigated the extent and pattern of genetic diversity in sequence data in
41 domain III (DIII) of the envelope (E) gene in serial plasma samples ($n=49$) taken from 17
42 patients infected with DENV-1, totaling some 8458 clones. Statistically rigorous
43 approaches were employed to account for artifactual variants resulting from
44 amplification and sequencing, and which we suggest have played a major role in
45 previous studies of intra-host genetic variation. Accordingly, nucleotide sequence
46 diversities of viral populations were very low, with conservative estimates of the
47 average levels of genetic diversity ranging from 0 – 0.0013. Despite such sequence
48 conservation, we observed clear evidence for mixed infection, with the presence of
49 multiple phylogenetically distinct lineages present within the same host, while the
50 presence of stop codon mutations in some samples suggests the action of
51 complementation. In contrast to some previous studies we observed no relationship
52 between the extent and pattern of DENV-1 genetic diversity and disease severity,
53 immune status, or level of viremia.

54 INTRODUCTION

55 DENV is a single-strand positive-sense RNA virus of the family *Flaviviridae* and exists as
 56 four closely related antigenically distinct serotypes, denoted DENV-1 to DENV-4. These
 57 serotypes differ at the consensus level by 25 to 40% at the amino acid (aa) level (15, 30).
 58 Genetic variation within each of the four serotypes is defined as a series of “genotypes”
 59 (or “subtypes”) which can vary from one another by up to ~6-8% and 3% at the
 60 nucleotide and aa levels, respectively (15, 25, 32). For example, at least four major
 61 genotypes of DENV-1 exist each with different geographical distributions (9, 38). The
 62 basis for the genetic diversity in DENV is its error-prone RNA polymerase (10), such that
 63 mutations commonly occur during viral replication, and on which a combination of
 64 genetic drift and negative and/or positive natural selection is able to act. This high
 65 replication error results in DENV existing as a population of closely related variants
 66 within an individually infected host (33, 34) and this intra-host genetic diversity has
 67 been proposed to have implications for pathogenesis of DENV infection, variable disease
 68 outcomes, virus evolution, and host immunity (6).

69 Several previous studies have confirmed that the population of DENV in humans and
 70 within individual *Aedes* mosquitoes contains measurable genetic variation (8, 19, 33,
 71 34). Levels of within-host genetic diversity have been previously shown to vary among
 72 patients. Reported levels of intra-host genetic diversity ranged from 0.21 to 1.67% for
 73 the E gene of DENV-3, with genome-defective DENVs observed in 3.9% - 5.8% of clones
 74 (33, 34). Another similar study showed that the intra-host diversity for the C and NS2B
 75 genes ranged from 0.12 to 1.02% and 0.16 to 1.20%, respectively (34). Lin *et al.* (19)
 76 showed that DENV exhibits substantial sequence diversity in humans and to a lesser
 77 extent in mosquitoes, with the major variant transmitted in both humans and
 78 mosquitoes. Intriguingly, Descloux *et al.* suggested that the level of intra-host genetic
 79 diversity was lower in patients suffering severe dengue disease – dengue hemorrhagic
 80 fever (DHF) and DSS – compared to those experiencing the milder dengue fever (DF),
 81 such that there is a direct link between viral genetic diversity and clinical outcome (8).

82 All previous studies of intra-host DENV genetic diversity have utilized point
 83 measurements in which a limited set of clones ($n = 10-50$) containing short, amplified
 84 segments of the viral genome were sequenced. In most cases the proportion of
 85 mutations due to experimental (PCR/sequencing) error in these studies is uncertain, but

86 likely an important contributor to the levels of diversity observed. As well as a limited
87 sample of population diversity, it is unknown whether the extent of sequence variation
88 changes during the course of infection, and the relationship between intra-host genetic
89 variation and dengue severity is unclear. To address these issues we undertook an
90 expansive study of intra-host DENV variation by sequencing a median of 155 high
91 quality clones from serial plasma samples taken from 17 patients infected with DENV-1
92 and applying a rigorous quality control to exclude artifactual mutations. With these data
93 we explored the relationship between intra-host genetic diversity and clinical outcome,
94 focusing on the sequence encoding the domain III (DIII) of the envelope (E) gene.
95 Importantly, DIII is involved in cell receptor binding and is the major target of virus
96 neutralizing antibodies in humans (3, 21) and hence mutations within this region may
97 have important functional consequences.

98 MATERIAL AND METHODS

99 Study population

100 Plasma samples from dengue patients included in the placebo arm of a clinical trial of
 101 chloroquine were used for this study (29). We selected seventeen patients for study
 102 based on the serotype of infection (i.e. DENV-1), serological response (i.e. primary or
 103 secondary), and disease severity (i.e. DF or DHF). Classification of disease severity was
 104 according to 1997 WHO classification criteria (36). For each patient, three sequential
 105 plasma samples, beginning with the enrolment plasma sample, were selected for
 106 analysis. Samples were selected to represent the breadth of viremia levels found in
 107 DENV-1 infected patients. Briefly, we selected three primary DF, seven secondary DF,
 108 and seven secondary DHF patients with a median age of 19, 19, and 20 years,
 109 respectively, with a male/female ratio of 1.3. The median day of illness at admission was
 110 2.2 days (range: 0.6-2.8 days).

111 Whole genome (consensus) sequencing of DENV-1

112 Viral genomes in the enrolment samples were sequenced as part of the Broad Institute's
 113 Genome Resources in Dengue project using a capillary sequencing directed amplification
 114 viral sequencing pipeline as previously reported (24). In short, isolated viral RNAs were
 115 reverse transcribed and then overlapping amplicons that span the complete genome
 116 were amplified using a high fidelity polymerase; resulting products were Sanger
 117 sequenced and resulting sequence coverage was ~8-fold. Resulting sequence reads were
 118 assembled using the Broad Institute's AV454 algorithm (Henn et al., in review).
 119 Consensus assemblies were used for alignment of clone reads as part of the variant
 120 calling process (see below).

121 RNA extraction, real-time PCR, cloning and sequencing

122 Dengue viral RNA was isolated directly from plasma using the QIAamp viral RNA mini
 123 kit (Qiagen, Germany). RNA was reverse transcribed, and DENV-1 viremia levels were
 124 assessed using an internally controlled, serotype specific, real-time reverse-
 125 transcriptase polymerase chain reaction (RT-PCR) assay that has been described
 126 elsewhere (18); results were expressed as cDNA equivalents per ml of serum.

127 The 462 nucleotide region encoding DIII of the E-gene was amplified using the primers:
 128 DIII-E P3; 5'- CAAGAAGGAGCAATGCACAC -3' (corresponding to genome positions 1701
 129 to 1720 of the DENV-1 reference strain (Hawaii, 1944)) and DIII-E P5; 5'-
 130 CCAAAGTCCCATGCGGTGTC - 3' (positions 2182 to 2201). The PCR was performed with
 131 5 µL 5X polymerase buffer (Roche), 1 µL 10 mM DIII-E P3 primer, 1 µL 10 mM DIII-E P5'
 132 primer, 3.5 µL 25mM MgCl₂ (Roche), 1 µL 10 mM dNTPs (Invitrogen), 2.5 U Expand High
 133 Fidelity Plus™ polymerase (Roche) and RNase-free water to a final volume of 25µl. The
 134 PCR conditions were 94°C for 2 min, followed by 45 cycles of 94°C 15 sec, 60°C 30 sec
 135 and 72°C 45 sec, and then 72°C 7 min. The resulting PCR product was visualized on a 1%
 136 agarose gel using ethidium bromide staining and UV light.

137 PCR amplimers were cloned into the T/A cloning vector, pCRII-TOPO, which was
 138 transformed into TOP10 competent cells (Invitrogen). Each transformation culture was
 139 plated out on Luria-Bertani (LB)/ampicillin/isopropylthiogalactoside (IPTG)/X-gal
 140 plates and grown overnight at 37°C. 382 white colonies (suggestive for amplicon
 141 insertion) were selected from each sample and sequenced using dye-terminator
 142 chemistry on ABI 3730xl sequencer (Applied Bioystems) from both ends to generate
 143 paired end reads and quality files.

144 **Variant calling**

145 *Read Alignment and Merging.* Reads from each sample were aligned to the consensus
 146 genome sequence present in the enrolment plasma sample using the BLAST-Like
 147 Alignment Tool (BLAT) version 33 (17). A custom script was used to merge overlapping
 148 forward and reverse reads, simultaneously assign appropriate base quality scores, and
 149 trim the resulting reads to the target amplicon sequence.

150 Overlapping forward and reverse reads were merged into a single contig and assigned
 151 quality scores. To control for poor alignment at the ends of reads, forward and reverse
 152 reads were required to have at least 5 bases aligning into the designed primer (i.e. DIII-E
 153 P3 and DIII-E P5) or were trimmed backwards 5 bases from the end of their alignment.
 154 The quality scores were assigned based on the agreement or disagreement of the bases
 155 between the forward and the reverse reads. The sum of quality scores were assigned for
 156 bases agreeing; bases disagreeing were assigned to the base with the highest quality
 157 score and quality score was assigned as the difference. Gaps were given quality equal to

the lower quality of the adjacent base, or the lowest quality of any contiguous base of the same type (homopolymer adjustment); base(s) were discarded when the gap had higher quality than the inserted base(s) on the opposite strand and bases retained their quality scores if the quality of bases were higher. Indels of the same length in both reads were retained as real. Complex events (e.g. inserts relative to reference opposite deletions, or insertions or deletions of different length) were replaced with a number of Ns equal to the length of the consensus between the two flanking consistent alignments and quality score=0. When the overlapping region (of the forward or reverse) read did not extend to the designed primer, the merged read was extended to include whichever read had the largest number of aligning bases on that side of the overlap, and assigned the raw quality for those bases. In cases where both complement forward or reverse reads did not align, we trimmed the single read to the target amplicon region and retained it for variant calling.

Base variant calling. To reduce false positive base variant calls we employed a Neighborhood Quality Standard (NQS) algorithm (2) to filter bases used for variant calling. Bases not meeting a NQS condition over those regions were excluded, i.e. a base satisfies the NQS condition if the base has PHRED score ≥ 20 , and the neighboring five bases on each side have PHRED scores ≥ 15 . Two variant base data sets were generated for downstream analysis. In the first, highest quality, data set defined as VP base variants were called using the V-Phaser algorithm (Macalalad et al., in review). In short, V-Phaser applies an error probability model defined by a process read error rate, and refined by the inclusion of variant nucleotide phasing information, to define the frequency at which a nucleotide polymorphism needs to be observed to be a true variant given the observed sequence coverage. In general, for the data sets analyzed as part of this study variants were identified as real if they were observed on two or more reads. To explore how erroneous PCR and sequencing may have contributed to the observed levels of genetic diversity we generated a second variant data set, defined as 1HQ that included variants that were seen only once (i.e. singletons). In both the 1HQ and VP data sets only high quality bases that passed NQS were used for base variant calling.

Variant Haplotype Calling. For each aligned read (see *Read Alignment and Merging*) we computed a vector of valid base variant calls (see *Base Variant Calling*). The minimal set of such vectors required to explain all reads was collected using a custom haplotype

190 calling algorithm. For each sample, the algorithm was seeded with a single read and
 191 then reads were assigned a haplotype. If a read matched unambiguously based on the
 192 variant positions to an existing haplotype group (in first iteration match is to seed read)
 193 it was assigned that haplotype, otherwise it was assigned as a new haplotype. This
 194 process was iterated until all reads were grouped into defined haplotypes defined by
 195 variant vectors. We assigned reads that have variant vectors with missing data (e.g. due
 196 to failure to align or presence of a call which is not considered valid) by a similar
 197 process. For reads that the partial vector maps unambiguously to a complete haplotype
 198 the missing information is "corrected" based on the complete vector; those reads that do
 199 not map unambiguously are assigned as "incomplete" haplotypes.

200 **Nucleotide sequence accession numbers.**

201 All nucleotide sequences generated here have been submitted to GenBank and assigned
 202 accession numbers 2262271431-2299350311 (Supplementary table S2).

203 **Evolutionary analysis**

204 *Measurements of genetic variation.* Alignments of full length pseudo-reads (i.e. all valid
 205 variants) from the haplotypes were generated with the MUSCLE software (version 3.7)
 206 (11), using default settings. Because the very low numbers of mutations observed, the
 207 mean pairwise genetic diversity within each sample was calculated from the
 208 uncorrected pairwise distance matrix (p -distance) between taxa and the population
 209 standard error (SE) was estimated with 1000 bootstrap replicates using the MEGA5
 210 program (27). To estimate the mean numbers of synonymous (d_s) and nonsynonymous
 211 (d_N) substitutions per site (ratio d_N/d_s) in each sample we utilized the Jukes-Cantor
 212 substitution model within MEGA5 (27). Mutations detected within each sample were
 213 further characterized as to their frequency and presence in other samples and were
 214 mapped to inferred amino acid (aa) sequences.

215 *Pattern of intra-host evolution.* The evolutionary relationships among the DENV-1
 216 sequences from each sample were inferred through the construction of minimum
 217 spanning networks, utilizing the program TCS 1.21 (5) and following the algorithm of
 218 Templeton *et al.* (28). Inferences from this method depend on the chosen probability of
 219 parsimony and we chose a value of 99% (i.e. a 99% connection limit). This number of
 220 mutational differences associated with the probability just before the (99%) cut-off is

221 the maximum number of mutational connections between pairs of sequences. Networks
222 that are unconnected at the 99% probability of parsimony were linked by decreasing the
223 connection probability. The power of this approach is that it allows the population
224 frequency of each mutation to be assessed, and parsimony-based approach is justified
225 by the small number of total mutations observed.

226 *Global DENV-1 phylogenetic inference.* To determine the frequency of mixed infections in
227 our data sets, the sequences of each individual patient were aligned together with 1390
228 previously published DENV-1 E gene sequences (i.e. 'background data set'), which
229 combines subsets of genotype I ($n=1111$), II ($n=91$) and III ($n=188$). Phylogenetic trees
230 for these data were then estimated using the maximum-likelihood (ML) method
231 available in the RAxML package (version 7.0.4) (26). In all cases we used GTR+ Γ_4 model
232 of nucleotide substitution, as determined by ModelTest v3.7 (23). The reliability of
233 specific groupings on the trees was estimated using bootstrap with 1000
234 pseudoreplicates.

235 RESULTS

236 *Extent and pattern of intra-host genetic variation*

237 The clinical, serological, and demographic features of the 17 DENV-1 infected patients
 238 that participated in this study are shown in Table 1. To determine the intra-host
 239 evolutionary dynamics of DENV-1 in these patients, we studied genetic diversity in 49
 240 serial plasma samples collected during the course of their illness. Overall, we sequenced
 241 8458 clones of the 463 nucleotide region encoding DIII of the E-gene derived from 49
 242 serial plasma samples collected during the course of infection. In the VP data set, 8458
 243 clones were assigned into complete haplotypes with a median of 155 (range: 4 - 362)
 244 clones analyzed at each time point (Table 2); these data excluded singleton mutations
 245 and included only high quality variant positions that were seen frequently enough at a
 246 given sequence coverage to be unlikely to occur as a result of error alone (i.e. typically
 247 observed at least twice). In the 1HQ data set which contains all variants observed
 248 including singletons that may be artefacts resulting from process errors, 8315 and 143
 249 clones were assigned to complete and incomplete haplotypes, respectively. A median of
 250 155 (range: 4 - 361) clones were analyzed at each time point (Table 3).

251 In the VP data set which included only highly confident variant calls, but which may have
 252 excluded some *bona fide* mutations at low frequency, we identified a total of 281 nt
 253 mutations across the 8458 clones of the 463 nt region (Table 2), corresponding to a
 254 mutational frequency of 7.2×10^{-5} (95% confidence interval (CI): $6.4 - 8.1 \times 10^{-5}$)
 255 mutations per nt site. Across all patients and time points, these mutations were
 256 observed at 43 residues (Supplementary Table S1). In all patients, the majority of
 257 sequences (65-100%, mean 97%) recovered were identical to the consensus. A measure
 258 of selection pressure could be calculated in 18 samples, with mean values of pairwise
 259 distance ranging from 0.00005 to 0.00130 (mean, 0.00034, Table 2). There was no
 260 significant difference in the mean pairwise distance between patients with DHF and DF
 261 (0.00030 vs. 0.00041, Table 2). To determine the selection pressure affecting DENV
 262 within each patient, we estimated the mean value of d_N/d_S for each sample. Mean d_N/d_S
 263 values varied between 0.13 and 1.9 with an average value of 0.23. Of the 8458 clones
 264 sequenced, 4 clones contained a total of 6 stop codons (0.05%) (Table 2). In sum, these
 265 stringently filtered data provided a conservative picture of the level of genetic diversity

266 in these samples, but those that are very likely to be real biological variants and suggest
267 that sequence diversities in viral populations may be very low.

268

269 *Intra-host phylogenetic relationships:*

270 To infer the evolutionary history of mutations in each sample we inferred minimum
271 spanning networks (Figure 1 and 2). In five patients (i.e. 49, 121, 154, 323 and 391), the
272 viral population harbored only the consensus sequence. Six patients (i.e. 59, 82, 107,
273 336, 349 and 376) contained haplotypes that are multiple mutational steps (≥ 2) away
274 from the consensus sequence, such that longer branches stem from the consensus
275 sequence. In addition, two patients (82 and 162) harbored multiple phylogenetically
276 distinct viral lineages (i.e. haplotypes) across multiple time points (figure 1A and 1B). A
277 third patient (336) also supported multiple haplotypes when the parsimony probability
278 was reduced to 97% (Figure 1E); in this network, hap 1 ($n = 2$) required seven
279 additional mutational steps, which was suggestive for mixed infections.

280 *Evidence of mixed infections*

281 Notably, one sample (G2542, patient 336) contained two identical clones that differed
282 seven nt (1.5%) from the consensus sequence, and hence far greater than that observed
283 in the majority of other patients (mean = 0.1%). This prompted us to determine
284 whether the high level of genetic diversity in patient 336 was due to mixed infections
285 from different origins within the global diversity of DENV-1. Phylogenetic analyses of
286 the alignment of all haplotypes of each patient with the 'background data set' (a global
287 samples of DENV-1 E DIII sequences from GenBank) provided strong evidence for
288 multiple infections, all involving genotype 1 viruses (Figure 3). Specifically, patient 336
289 harbored a mixed infection with viruses from clade 1 and 5 (clades as described by
290 Raghwani et al. (24)).

291 *Analysis of the 1HQ data set*

292 As a comparison with the high quality but conservative VP data set, and to assess the
293 likely extent of sequencing error, we performed an additional analysis of the 1HQ data
294 set. Among the 8315 clones of the VP data set, 2936 nt mutations were observed,
295 corresponding to a mutation frequency of 7.6×10^{-4} (95% CI: $7.4\text{-}7.9 \times 10^{-4}$) mutations

per nt site (Table 3). A total of 1922 aa mutations were observed. The majority of clones ($n = 1434$, 17.4%) harbored a single aa mutation while 2.8% carried multiple aa mutations. Mean estimates of pairwise genetic diversity varied from 0.00048 to 0.00360 (mean, 0.00164) and the mean values of d_N/d_S ranged from 0 to 1.6 (mean, 0.58). These d_N/d_S values are much higher than those seen between patients, which are normally <0.1 , suggesting that intra-host variation is characterized by transient deleterious mutations or caused by the experimental procedure, which results in an elevation of d_N/d_S values (14). In addition, 36 in-frame stop codons in 32 clones were identified in 17 of the 49 samples studied (Table 3). Genome-defective DENVs observed in 0.38% of clones. All mutations in the VP data set ($n = 43$) were also observed in sequential samples and/or across multiple patients in the 1HQ data set (supplementary Table S1, left). Many mutation positions ($n = 625$ of 845, 74%) were observed in sequential samples and/or across multiple patients in the 1HQ data set, but these mutations lacked statistical rigor to be called a valid variant in the VP data set.

310

311 DISCUSSION

312 The intra-host population genetic structure of DENV has previously been described as a
 313 population of closely related sequences (1, 7, 14, 19, 33, 34). Our study, which comprises
 314 the largest series of samples and patients as well as stringent filtering of sequence
 315 quality, confirms these observations, but shows that the occurrence of mutations in the
 316 virus population are much lower than previously reported. The mean pairwise genetic
 317 diversity varied between 0.00048 to 0.00360 and 0.00005 to 0.00130 in the 1HQ and VP
 318 data sets, respectively, with no significant difference in the mean pairwise distance
 319 between patients with DHF and DF. The substantially higher sequence variation in our
 320 1HQ data set resembled that described in previous reports (8, 19, 33, 34). However,
 321 given that the 1HQ data set undoubtedly includes a significant number of artifactual
 322 mutations, the high sequence variation in this data set should be regarded as the upper
 323 bound of DENV genetic diversity. As a consequence, it is likely that previous estimates of
 324 intra-host genetic diversity in DENV have been inflated by the erroneous inclusion of
 325 PCR and sequencing errors in the diversity calculations and hence should be treated
 326 with caution.

327 It is important to note that accurate estimations of intra-host sequence variability
 328 depend largely on the accuracy of the experimental procedure, particularly the fidelity of
 329 RT-PCR and sequencing. However, distinguishing *bona fide* from artifactual mutations is
 330 not a trivial exercise. Our rigorous approach to error correction relies on: (i) the
 331 alignment of clonal sequences to a reference sequence for haplotype calling, (ii) the
 332 identification of unambiguous mutations with high quality score of base(s) and (iii)
 333 whether mutations were seen once (1HQ data set) or frequently enough at a given
 334 sequence coverage to be unlikely to be from error (VP data set). Indeed, the 1HQ data set
 335 must harbor a high, but undetermined number of artifactual mutations which were
 336 likely introduced during reverse transcription, PCR amplification or sequencing. The
 337 process error rate (i.e. RT-PCR + cloning + Sanger sequencing) can be expected to be on
 338 the order of $2-8 \times 10^{-6}$ /nt/cycle when a proofreading polymerase is used as reported by
 339 Malet et al. (20) which corresponds to an expectation that ~0.036% of the observed
 340 variants could be errors in our experimental system (% mutations observed in 1HQ data
 341 set 0.024-0.269%). Conversely, the VP data set undoubtedly represents biological
 342 variants, but may underestimate the true intra-host sequence variation, as the variant

343 calling algorithm will call singletons as errors despite that some of these mutations
 344 possibly representing true biological variants. Notably, the probability of a mutation
 345 occurring independently at random across multiple sequential samples is very low,
 346 hence singleton variants observed in multiple samples may have a higher likelihood of
 347 being true biological variants than those observed in a single patient. Indeed, although
 348 RT, PCR and sequencing errors likely contribute to the majority of variants observed in
 349 the 1HQ data set, we were able to identify mutations that occurred in multiple patients
 350 and time points (supplementary Table S1), suggesting that they are biological variants
 351 even though they are at low frequency within individual patients and hence excluded
 352 from the VP data set.

353 Overall, our VP data set indicates that the DIII segment of the E gene in DENV-1 exhibits
 354 limited sequence variation during the course of infection. In addition, it is striking that in
 355 both the VP and HQ1 data sets, we found no clear evidence for adaptive evolution in the
 356 DIII region, either in the form of consistently high d_N/d_S ratios and/or mutations that
 357 exhibited a steady increase in frequency, even though it is thought to be the principle
 358 target for neutralizing antibodies (3, 21). The lack of positive selection in this case is
 359 likely to be a function of the fact that dengue is a self-limiting infection in which innate,
 360 humoral and cellular immune mechanisms removes the virus population before
 361 evidence of positive selection can be detected (4, 13, 22, 31).

362 The relationship between viral genetic variation and disease severity has been well
 363 documented in human immunodeficiency virus 1 (HIV-1) and hepatitis C virus (HCV)
 364 (12, 35). For example, higher HIV-1 sequence diversity has been shown to be associated
 365 with slower disease progression (35). Similarly, disease progression in HCV infection
 366 was associated with measurable genetic evolution, while resolving hepatitis correlated
 367 with evolutionary stasis in the acute phase of HCV infection (12). Because our analysis
 368 considered a relatively large number of sequences per patient and these patients likely
 369 harbored differences in immunological responses, we were able to look for associations
 370 between the intra-host diversity of DENV-1 and disease outcome, immune status or
 371 viremia. Notably, we observed no clear evolutionary patterns in relation to any of these
 372 variables. These results sit in contrast to results reported by Descloux *et al.* (8) who
 373 showed higher intra-host sequence variation in patients with DHF/DSS than those with
 374 DF. The basis for the differences in results between our studies are unknown, but could

375 be related to the methods used to filter sequence quality or to sample size. In addition,
376 Descloux *et al.* assessed a much smaller number of clones (662 clones from 16 sera
377 samples at a single time point), increasing the chance of stochastic effects.

378 Finally, one of the most striking observations from this study was the presence within
379 some patients of phylogenetically distinct lineages or subtypes of genotype 1 DENV-1,
380 indicative of mixed infection. That these mixed infection events were also observed
381 within the high quality VP data sets indicates that they are *bona fide*. This is the first
382 time that intra-serotype mixed infection has been reported in DENV-1, and we likely
383 greatly underestimate its true frequency as we are only able to infer the occurrence of
384 mixed infection when it involves lineages that fall into topologically distinct places on
385 phylogenetic trees (i.e. we cannot identify mixed infection among very closely related
386 viral lineages). Intriguingly, a previous study of DENV-2 evolution also revealed the
387 presence of mixed infection, such that individual patients harbored multiple
388 phylogenetically distinct lineages (1). We therefore conclude that mixed infection is a
389 potentially important contributor to intra-host virus genetic and phenotypic diversity,
390 and provides the raw material for intra-serotype recombination (16, 37). However, we
391 were unable to determine whether these mixed infections represent the simultaneous
392 infection (i.e. co-infection) or superinfection of multiple viral lineages in humans. This is
393 clearly an area that requires additional study.

394 **ACKNOWLEDGEMENT**

395 This study was funded by the Wellcome Trust (CPS) and in part with US Federal funds
 396 from the National Institute of Allergy and Infectious Diseases National Institutes of
 397 Health, Department of Health and Human Services, under Contract Nos.
 398 HHSN272200900006C (BWB) and HHSN272200900018C (BWB). Khoa T.D. Thai is
 399 supported by a 'Mosaic' fellowship from the Netherlands Organization for Scientific
 400 Research (NWO). (Grant: 017.004.074).

401

REFERENCES

1. **Aaskov, J., K. Buzacott, H. M. Thu, K. Lowry, and E. C. Holmes.** 2006. Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science* **311**:236-238.
2. **Altshuler, D., V. J. Pollara, C. R. Cowles, W. J. Van Etten, J. Baldwin, L. Linton, and E. S. Lander.** 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**:513-516.
3. **Beltramello, M., K. L. Williams, C. P. Simmons, A. Macagno, L. Simonelli, N. T. Quyen, S. Sukupolvi-Petty, E. Navarro-Sanchez, P. R. Young, A. M. de Silva, F. A. Rey, L. Varani, S. S. Whitehead, M. S. Diamond, E. Harris, A. Lanzavecchia, and F. Sallusto.** 2010. The human immune response to Dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity. *Cell Host. Microbe* **8**:271-283.
4. **Chau, T. N., K. L. Anders, I. B. Lien, N. T. Hung, L. T. Hieu, N. M. Tuan, T. T. Thuy, I. T. Phuong, N. T. Tham, M. N. Lanh, J. J. Farrar, S. S. Whitehead, and C. P. Simmons.** 2010. Clinical and virological features of Dengue in Vietnamese infants. *PLoS. Negl. Trop. Dis.* **4**:e657.
5. **Clement, M., D. Posada, and K. A. Crandall.** 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**:1657-1659.
6. **Clyde, K., J. L. Kyle, and E. Harris.** 2006. Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *J. Virol.* **80**:11418-11431.
7. **Craig, S., H. M. Thu, K. Lowry, X. F. Wang, E. C. Holmes, and J. Aaskov.** 2003. Diverse dengue type 2 virus populations contain recombinant and both parental viruses in a single mosquito host. *J. Virol.* **77**:4463-4467.
8. **Descloux, E., V. M. Cao-Lormeau, C. Roche, and L. De, X.** 2009. Dengue 1 diversity and microevolution, French Polynesia 2001-2006: connection with epidemiology and clinics. *PLoS. Negl. Trop. Dis.* **3**:e493.
9. **Diaz, F. J., W. C. Black, J. A. Farfan-Ale, M. A. Lorono-Pino, K. E. Olson, and B. J. Beaty.** 2006. Dengue virus circulation and evolution in Mexico: a phylogenetic perspective. *Arch. Med. Res.* **37**:760-773.

- 434 10. **Domingo, E. and J. J. Holland.** 1997. RNA virus mutations and fitness for survival.
435 *Annu. Rev. Microbiol.* **51**:151-178.
- 436 11. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and
437 high throughput. *Nucleic Acids Res.* **32**:1792-1797.
- 438 12. **Farci, P., A. Shimoda, A. Coiana, G. Diaz, G. Peddis, J. C. Melpolder, A. Strazzer,**
439 **D. Y. Chien, S. J. Munoz, A. Balestrieri, R. H. Purcell, and H. J. Alter.** 2000. The
440 outcome of acute hepatitis C predicted by the evolution of the viral quasispecies.
441 *Science* **288**:339-344.
- 442 13. **Gubler, D. J.** 1998. Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.*
443 **11**:480-496.
- 444 14. **Holmes, E. C.** 2003. Patterns of intra- and interhost nonsynonymous variation
445 reveal strong purifying selection in dengue virus. *J. Virol.* **77**:11296-11298.
- 446 15. **Holmes, E. C. and S. S. Twiddy.** 2003. The origin, emergence and evolutionary
447 genetics of dengue virus. *Infect. Genet. Evol.* **3**:19-28.
- 448 16. **Holmes, E. C., M. Worobey, and A. Rambaut.** 1999. Phylogenetic evidence for
449 recombination in dengue virus. *Mol. Biol. Evol.* **16**:405-409.
- 450 17. **Kent, W. J.** 2002. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**:656-664.
- 451 18. **Laue, T., P. Emmerich, and H. Schmitz.** 1999. Detection of dengue virus RNA in
452 patients after primary or secondary dengue infection by using the TaqMan
453 automated amplification system. *J Clin Microbiol.* **37**:2543-2547.
- 454 19. **Lin, S. R., S. C. Hsieh, Y. Y. Yueh, T. H. Lin, D. Y. Chao, W. J. Chen, C. C. King, and**
455 **W. K. Wang.** 2004. Study of sequence variation of dengue type 3 virus in naturally
456 infected mosquitoes and human hosts: implications for transmission and evolution.
457 *J. Virol.* **78**:12717-12721.
- 458 20. **Malet, I., M. Belnard, H. Agut, and A. Cahour.** 2003. From RNA to quasispecies: a
459 DNA polymerase with proofreading activity is highly recommended for accurate
460 assessment of viral diversity. *J Virol. Methods* **109**:161-170.
- 461 21. **Midgley, C. M., M. Bajwa-Joseph, S. Vasanawathana, W. Limpitikul, B. Wills, A.**
462 **Flanagan, E. Waiyaiya, H. B. Tran, A. E. Cowper, P. Chotiarnwon, J. M. Grimes,**
463 **S. Yoksan, P. Malasit, C. P. Simmons, J. Mongkolsapaya, and G. R. Screaton.**
464 2011. An in-depth analysis of original antigenic sin in dengue virus infection. *J*
465 *Virol.* **85**:410-421.

- 466 22. **Murgue, B., C. Roche, E. Chungue, and X. Deparis.** 2000. Prospective study of the
467 duration and magnitude of viraemia in children hospitalised during the 1996-1997
468 dengue-2 outbreak in French Polynesia. *J. Med. Virol.* **60**:432-438.
- 469 23. **Posada, D. and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA
470 substitution. *Bioinformatics.* **14**:817-818.
- 471 24. **Raghwani, J., A. Rambaut, E. C. Holmes, V. T. Hang, T. T. Hien, J. Farrar, B. Wills,**
472 **N. J. Lennon, B. W. Birren, M. R. Henn, and C. P. Simmons.** 2011. Endemic
473 dengue associated with the co-circulation of multiple viral lineages and localized
474 density-dependent transmission. *PLoS. Pathog.* **7**:e1002064.
- 475 25. **Rico-Hesse, R.** 1990. Molecular evolution and distribution of dengue viruses type
476 1 and 2 in nature. *Virology* **174**:479-493.
- 477 26. **Stamatakis, A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
478 analyses with thousands of taxa and mixed models. *Bioinformatics.* **22**:2688-2690.
- 479 27. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar.** 2011.
480 MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood,
481 Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.*
- 482 28. **Templeton, A. R., K. A. Crandall, and C. F. Sing.** 1992. A cladistic analysis of
483 phenotypic associations with haplotypes inferred from restriction endonuclease
484 mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**:619-633.
- 485 29. **Tricou, V., N. N. Minh, T. P. Van, S. J. Lee, J. Farrar, B. Wills, H. T. Tran, and C. P.**
486 **Simmons.** 2010. A randomized controlled trial of chloroquine for the treatment of
487 dengue in Vietnamese adults. *PLoS. Negl. Trop. Dis.* **4**:e785.
- 488 30. **Vasilakis, N. and S. C. Weaver.** 2008. The history and evolution of human dengue
489 emergence. *Adv. Virus Res.* **72**:1-76.
- 490 31. **Vaughn, D. W., S. Green, S. Kalayanarooj, B. L. Innis, S. Nimmannitya, S.**
491 **Suntayakorn, T. P. Endy, B. Raengsakulrach, A. L. Rothman, F. A. Ennis, and A.**
492 **Nisalak.** 2000. Dengue viremia titer, antibody response pattern, and virus serotype
493 correlate with disease severity. *J. Infect. Dis.* **181**:2-9.
- 494 32. **Vu, T. T., E. C. Holmes, V. Duong, T. Q. Nguyen, T. H. Tran, M. Quail, C. Churcher,**
495 **J. Parkhill, J. Cardoso, J. Farrar, B. Wills, N. J. Lennon, B. W. Birren, P. Buchy, M.**
496 **R. Henn, and C. P. Simmons.** 2010. Emergence of the Asian 1 genotype of dengue
497 virus serotype 2 in viet nam: in vivo fitness advantage and lineage replacement in
498 South-East Asia. *PLoS. Negl. Trop. Dis.* **4**:e757.

- 499 33. **Wang, W. K., S. R. Lin, C. M. Lee, C. C. King, and S. C. Chang.** 2002. Dengue type 3
500 virus in plasma is a population of closely related genomes: quasispecies. *J. Virol.*
501 **76**:4662-4665.
- 502 34. **Wang, W. K., T. L. Sung, C. N. Lee, T. Y. Lin, and C. C. King.** 2002. Sequence
503 diversity of the capsid gene and the nonstructural gene NS2B of dengue-3 virus in
504 vivo. *Virology* **303**:181-191.
- 505 35. **Wolinsky, S. M., B. T. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman, A. J.**
506 **Whetsell, M. R. Furtado, Y. Cao, D. D. Ho, and J. T. Safrit.** 1996. Adaptive
507 evolution of human immunodeficiency virus-type 1 during the natural course of
508 infection. *Science* **272**:537-542.
- 509 36. **World Health Organization.** 1997. Dengue Haemorrhagic Fever. Diagnosis,
510 Treatment and Control. World Health Organization, Geneva, Switzerland.
- 511 37. **Worobey, M., A. Rambaut, and E. C. Holmes.** 1999. Widespread intra-serotype
512 recombination in natural populations of dengue virus. *Proc. Natl. Acad. Sci. U. S. A*
513 **96**:7352-7357.
- 514 38. **Zhang, C., M. P. Mammen, Jr., P. Chinnawirotpisan, C. Klungthong, P.**
515 **Rodpradit, P. Monkongdee, S. Nimmannitya, S. Kalayanarooj, and E. C.**
516 **Holmes.** 2005. Clade replacements in dengue virus serotypes 1 and 3 are
517 associated with changing serotype prevalence. *J. Virol.* **79**:15123-15130.

518 **Table 1.** Summary characteristics of the seventeen patients with DEVN-1 infections

Patient number	Disease outcome	Immune status (Pri/Sec)	Clade (within genotype 1)*	Age	Sex	Consensus assembly GenBank accession number	DENV-1 RNA (copies/mL)		Day of illness	Sequence ID
							Peak viremia	Samples		
49	DF	Pri	1	18	M	FJ410197	7.97x10 ⁸	2.60x10 ⁸	2.7	G2500
								3.97x10 ⁷	3.7	G2501
								8.64x10 ⁵	4.5	G2502
76	DF	Pri	1	21	M	FJ410205	2.33x10 ⁸	2.33x10 ⁸	2.8	G2506
								5.42x10 ⁶	3.8	G2507
								1.10x10 ⁶	4.8	G2508
336	DF	Pri	1	19	F	FJ461319	5.33x10 ⁷	5.33x10 ⁷	2.4	G2542
								1.18x10 ⁷	3.1	G2543
								6.08x10 ⁶	3.4	G2544
82	DF	Sec	N/A	15	M	N/A	1.61x10 ⁹	1.61x10 ⁹	3.0	G2509
								9.35x10 ⁷	4.0	G2510
								1.76x10 ⁷	5.0	G2511
107	DF	Sec	4	18	M	FJ432746	3.76x10 ¹⁰	1.34x10 ¹⁰	1.6	G2512
								1.04x10 ¹⁰	3.3	G2513
								3.88x10 ⁹	3.6	G2514
121	DF	Sec	5	22	M	FJ410212	4.87x10 ⁹	4.87x10 ⁹	1.6	G2515
								2.19x10 ⁸	3.6	G2516
								5.42x10 ⁶	4.9	G2517
171	DF	Sec	4	19	M	FJ410279	6.10x10 ¹⁰	1.49x10 ¹⁰	1.9	G2527
								2.71x10 ¹⁰	3.7	G2528
								2.43x10 ⁸	4.7	G2529
309	DF	Sec	4	18	F	FJ410216	1.03x10 ¹⁰	5.30x10 ⁹	3.0	G2536
								7.06x10 ⁸	4.0	G2537
								7.75x10 ⁷	5.0	G2538
323	DF	Sec	1	31	F	FJ461316	2.31x10 ¹⁰	1.00x10 ¹⁰	1.4	G2539

383	DF	Sec	1	34	F	FJ410247	1.29x10 ¹⁰	5.16x10 ⁸	3.4	G2540
								5.85x10 ⁷	4.4	G2541
								2.22x10 ⁹	1.8	G2551
								2.25x10 ⁹	3.8	G2552
								7.88x10 ⁷	4.8	G2553
59	DHF I	Sec	4	25	M	FJ410199	1.65x10 ¹⁰	1.65x10 ¹⁰	2.3	G2503
								2.65x10 ⁹	3.3	G2504
								1.91x10 ⁷	4.0	G2505
146	DHF II	Sec	2	16	M	FJ410269	3.99x10 ⁹	3.99x10 ⁹	2.4	G2518
								2.64x10 ⁸	3.4	G2519
								1.61x10 ⁸	3.6	G2520
154	DHF II	Sec	1	34	M	FJ410272	2.08x10 ⁹	2.08x10 ⁹	2.3	G2521
								3.04x10 ⁸	3.3	G2522
								5.00x10 ⁶	4.0	G2523
162	DHF II	Sec	1	38	M	FJ410275	1.24x10 ¹⁰	1.24x10 ¹⁰	2.4	G2524
								1.01x10 ⁸	3.4	G2525
								4.94x10 ⁵	4.2	G2526
349	DHF II	Sec	1	20	F	FJ547065	2.36x10 ¹⁰	2.36x10 ¹⁰	2.7	G2545
								1.56x10 ⁹	3.7	G2546
								3.79x10 ⁷	4.7	G2547
376	DHF II	Sec	4	20	F	FJ410244	2.06x10 ⁸	2.06x10 ⁸	2.6	G2548
								6.52x10 ⁷	3.4	G2549
								1.68x10 ⁴	4.4	G2550
391	DHF II	Sec	5	18	F	FJ410220	2.19x10 ⁹	2.19x10 ⁹	2.3	G2554
								1.52x10 ⁸	3.3	G2555
								1.22x10 ⁶	4.0	G2556

* Clades of genotype I as designated by Raghwani, et al. (24).

DF, dengue fever; DHF I and II, dengue hemorrhagic fever grade 1 and 2; pri, primary dengue; sec, secondary dengue; M, male; F, female

522 **Table 2. Analysis of intra-host variation in DENV sequences in sequential samples from 17 patients in the VP data set.**

Sequence ID	Clone merged sequences (n)	Total number of nt mutations	% of nt mutations	Mean pairwise distance	SE	Global d _N /d _S	Haplotypes (n)	Total number of aa mutations	no. of stop codons	Coverage (mean)
G2500	118	-	-	-	-	-	1	-	-	117.7
G2501	47	-	-	-	-	-	1	-	-	47.0
G2503	323	-	-	-	-	-	1	-	-	322.7
G2504	301	7	0.005	0.00010	0.00006	0.133	3	2	-	300.1
G2505	235	44	0.037	0.00080	0.00022	0.379	11	26	4	234.8
G2506	257	-	-	-	-	-	1	-	-	256.5
G2507	39	3	0.015	0.00031	0.00030	-	2	3	-	39.0
G2508	8	-	-	-	-	-	1	-	-	7.9
G2509	292	63	0.043	0.00073	0.00071	-	2	63	-	291.7
G2510	144	51	0.071	0.00130	0.00078	1.947	10	45	-	143.8
G2511	63	22	0.069	0.00100	0.00097	-	2	22	-	62.9
G2512	329	5	0.003	0.00007	0.00005	-	3	5	-	328.5
G2513	362	-	-	-	-	-	1	-	-	361.7
G2514	309	-	-	-	-	-	1	-	-	308.7
G2515	170	-	-	-	-	-	1	-	-	169.8
G2516	130	-	-	-	-	-	1	-	-	129.9
G2517	155	-	-	-	-	-	1	-	-	154.8
G2518	47	-	-	-	-	-	1	-	-	47.0
G2519	232	4	0.003	0.00007	0.00007	-	2	4	-	231.7
G2520	285	-	-	-	-	-	1	-	-	284.7
G2521	230	-	-	-	-	-	1	-	-	229.1
G2522	4	-	-	-	-	-	1	-	-	3.9
G2523	17	-	-	-	-	-	1	-	-	17.0
G2524	252	33	0.026	0.00049	0.00047	0.000	2	0	-	251.8
G2525	58	3	0.010	0.00022	0.00020	0.000	2	0	-	58.0

	G2526	6	-	-	-	-	-	1	-	-	6.0
	G2527	213	3	0.003	0.00006	0.00006	0.000	2	0	-	212.6
	G2528	207	-	-	-	-	-	1	-	-	206.5
	G2529	15	-	-	-	-	-	1	-	-	15.0
	G2536	330	4	0.002	0.00005	0.00005	0.000	2	0	-	329.6
	G2537	252	-	-	-	-	-	1	-	-	251.8
	G2538	69	-	-	-	-	-	1	-	-	68.9
	G2539	354	-	-	-	-	-	1	-	-	353.8
	G2540	142	-	-	-	-	-	1	-	-	141.8
	G2541	27	-	-	-	-	-	1	-	-	27.0
	G2542	176	14	0.016	0.00034	0.00013	0.000	2	0	-	175.8
	G2543	93	4	0.009	0.00018	0.00017	0.000	2	0	-	92.9
	G2544	9	-	-	-	-	-	1	-	-	9.0
	G2545	348	4	0.002	0.00005	0.00003	0.273	2	2	-	347.7
	G2546	263	5	0.004	0.00008	0.00006	0.000	3	0	-	262.7
	G2547	142	-	-	-	-	-	1	-	-	141.8
	G2548	123	8	0.013	0.00028	0.00015	0.318	4	5	2	122.8
	G2549	13	-	-	-	-	-	1	-	-	13.0
	G2551	325	-	-	-	-	-	1	-	-	324.5
	G2552	320	4	0.002	0.00005	0.00005	0.000	2	0	-	319.8
	G2553	112	-	-	-	-	-	1	-	-	111.9
	G2554	308	-	-	-	-	-	1	-	-	307.8
	G2555	154	-	-	-	-	-	1	-	-	153.9
	G2556	50	-	-	-	-	-	1	-	-	49.9
523	aa: amino acid; d _N /d _S : ratio of nonsynonymous/synonymous substitutions per site; nonsyn.: nonsynonymous; nt: nucleotide; SE:										
524	standard error										

525 **Table 3.** Analysis of intra-host variation in DENV sequences in sequential samples from 17 patients in the 1HQ data set.

Study ID	Clone merged sequences (n)	Total number of nt mutations	% of nt mutations	Mean pairwise distance	SE	Global d _N /d _S	Haplotypes (n)	Total number of aa mutations	no. of stop codons	Coverage (mean)
G2500	116	23	0.042	0.00093	0.00020	0.269	13	12	-	108.4
G2501	47	8	0.035	0.00074	0.00026	0.183	8	3	-	45.5
G2503	317	92	0.057	0.00125	0.00012	0.750	69	64	-	318.5
G2504	285	80	0.060	0.00127	0.00019	0.859	56	62	1	265.8
G2505	233	121	0.104	0.00225	0.00030	0.500	59	76	5	231.8
G2506	254	45	0.038	0.00078	0.00011	0.473	40	28	-	233.1
G2507	39	17	0.088	0.00186	0.00050	0.718	13	12	-	38.5
G2508	8	4	0.111	0.00216	0.00105	-	4	4	-	7.1
G2509	254	156	0.122	0.00277	0.00080	0.609	83	121	3	255.2
G2510	141	97	0.137	0.00276	0.00083	0.923	43	78	-	140.5
G2511	63	47	0.149	0.00271	0.00102	1.682	24	41	1	62.9
G2512	320	94	0.061	0.00127	0.00013	0.899	76	70	-	306.8
G2513	361	135	0.074	0.00163	0.00015	0.668	99	92	3	361.0
G2514	307	69	0.047	0.00103	0.00013	0.506	55	45	1	293.9
G2515	169	37	0.045	0.00100	0.00016	0.398	32	22	-	165.0
G2516	130	33	0.051	0.00109	0.00019	0.758	26	24	2	128.8
G2517	155	71	0.092	0.00198	0.00023	0.518	56	44	-	152.7
G2518	47	12	0.052	0.00119	0.00033	0.679	11	9	-	46.2
G2519	230	116	0.101	0.00216	0.00020	0.592	89	76	-	228.0
G2520	284	90	0.064	0.00138	0.00016	0.653	72	62	-	279.5
G2521	212	82	0.077	0.00155	0.00018	0.442	64	46	2	210.5
G2522	4	3	0.198	0.00324	0.00183	0.613	3	2	-	3.1
G2523	17	8	0.095	0.00203	0.00067	0.307	8	4	-	16.8
G2524	250	99	0.079	0.00162	0.00048	0.247	57	42	-	248.9
G2525	58	25	0.087	0.00185	0.00039	0.435	20	15	1	57.2

G2526	6	5	0.166	0.00360	0.00154	0.461	5	3	-	6.0
G2527	210	80	0.077	0.00174	0.00022	0.344	54	45	1	206.8
G2528	201	68	0.069	0.00150	0.00020	0.823	49	51	-	194.5
G2529	15	5	0.070	0.00144	0.00063	0.465	4	3	-	14.1
G2536	328	142	0.087	0.00189	0.00016	0.482	101	89	2	323.3
G2537	251	93	0.074	0.00167	0.00019	0.535	74	63	3	248.6
G2538	69	38	0.110	0.00238	0.00037	0.325	29	19	1	68.5
G2539	353	87	0.050	0.00109	0.00012	0.667	70	61	-	348.5
G2540	140	50	0.071	0.00151	0.00022	0.697	40	34	-	139.2
G2541	27	5	0.037	0.00080	0.00033	0.460	6	3	-	26.9
G2542	175	50	0.058	0.00123	0.00019	0.234	31	22	-	172.6
G2543	92	38	0.083	0.00178	0.00031	0.426	29	22	-	91.2
G2544	9	1	0.022	0.00048	0.00049	0.000	2	0	-	9.0
G2545	346	128	0.073	0.00158	0.00015	0.556	94	82	4	347.4
G2546	260	109	0.084	0.00174	0.00019	0.782	72	74	-	258.1
G2547	141	63	0.089	0.00196	0.00022	0.871	46	46	1	140.0
G2548	122	67	0.110	0.00240	0.00030	0.395	44	39	4	120.8
G2549	13	4	0.061	0.00133	0.00061	0.923	5	3	-	13.0
G2551	316	100	0.063	0.00131	0.00013	0.320	70	49	-	315.6
G2552	317	90	0.056	0.00123	0.00013	0.606	71	59	-	319.2
G2553	112	26	0.047	0.00100	0.00019	0.581	22	17	-	110.8
G2554	308	57	0.037	0.00081	0.00011	0.725	46	41	1	304.6
G2555	153	42	0.055	0.00116	0.00019	0.530	34	26	-	152.5
G2556	50	24	0.098	0.00207	0.00039	0.745	21	17	-	48.7

526 aa: amino acid; d_N/d_S: ratio of nonsynonymous/synonymous substitutions per site; nonsyn.: nonsynonymous; nt: nucleotide; SE:

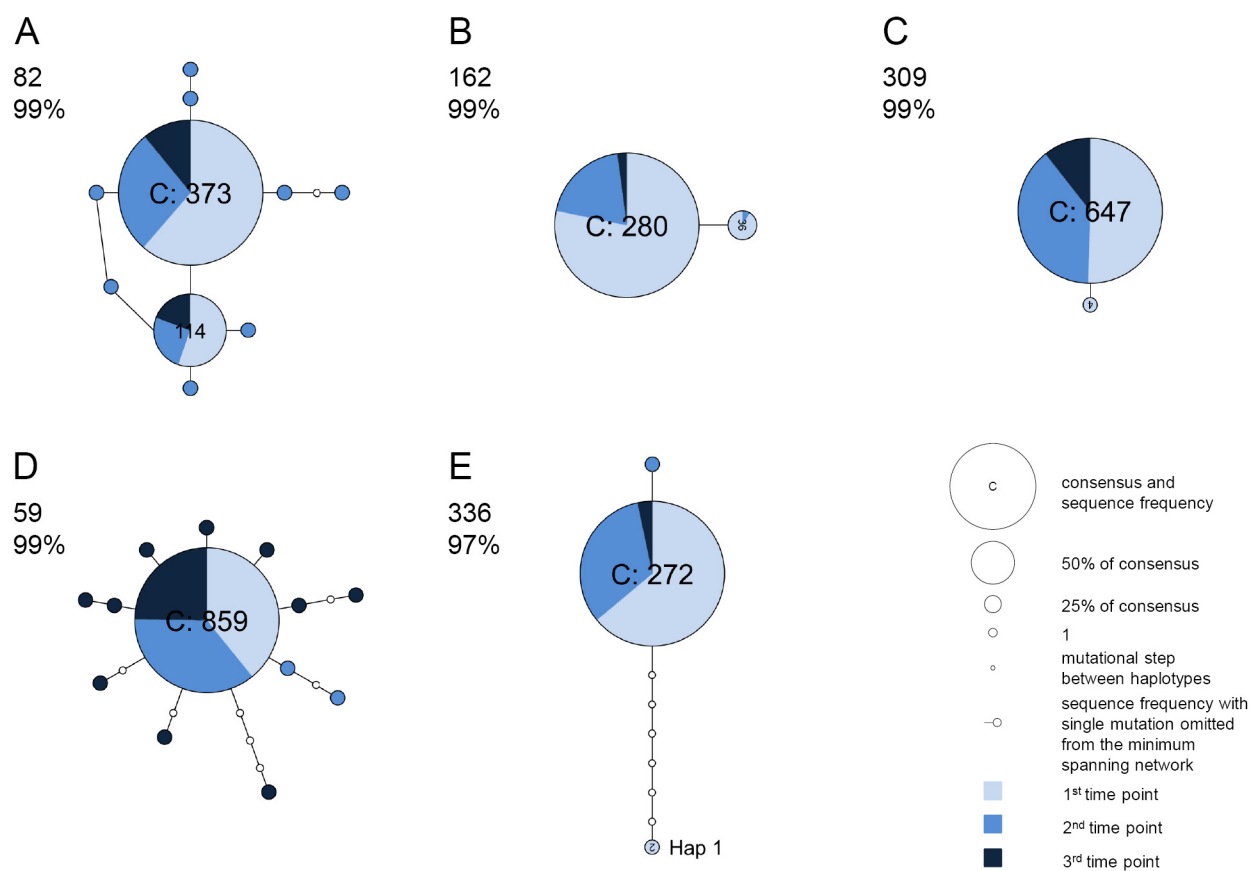
527 standard error

528 **FIGURE LEGENDS**

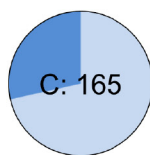
529 **Figure 1.** Minimum spanning networks of intra-host DENV-1 sequence data (VP data
530 set). Each network was inferred by compiling sequences from multiple days. Numbers in
531 the upper left corner correspond to the patient number; percentages indicate the
532 probability of parsimony used to construct the network. Haplotypes with the high
533 ancestral probability are displayed as circles. Circles sizes are proportional to the
534 number of sequences that exhibit each variant, and the pie chart in each circle indicate
535 the percentage of each variant at different time points. Connecting lines indicate a single
536 mutation shared among haplotypes. (A and B) Minimum spanning network in which
537 multiple viral lineages were observed across time-points (patient 82 and 162). (C)
538 Minimum spanning network in which one mutation was shared between haplotypes
539 (patient 309). (D) Minimum spanning network with star-like typology (patient 59). (E)
540 Minimum spanning network with reduced parsimony probability (patient 336).

541 **Figure 2.** Minimum spanning networks of intra-host DENV sequence data (VP data set)
542 in which one (B, E, G, J and K) and/or two (C, I and J) mutations were shared between
543 haplotypes. All sequences were identical to the consensus in A, D, F, H and L. Refer to
544 figure 1 for more information.

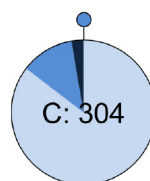
545 **Figure 3.** Maximum likelihood (ML) phylogenetic tree for all ($n = 89$) consensus
546 sequences derived from clones in the VP data set in relation to 1390 equivalent
547 background DENV-1 sequences collected from GenBank. Red colored lines represent
548 clones from sample G2542 and red arrow bars signify mixed infection. Clades are
549 indicated as numbers. Horizontal branches are drawn to a scale of nucleotide
550 substitutions per site, and the tree is midpoint rooted, nodes are ordered increasingly
551 and presented as a polar tree.



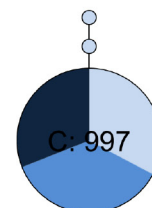
A
49
99%



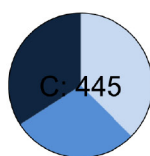
B
76
99%



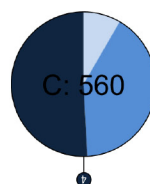
C
107
99%



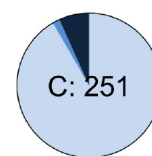
D
121
99%



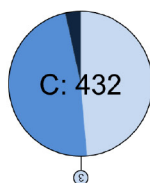
E
146
99%



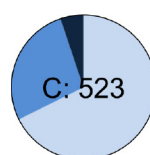
F
154
99%



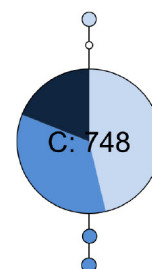
G
171
99%



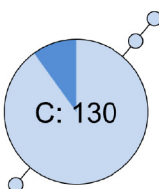
H
323
99%



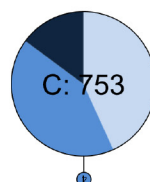
I
349
99%



J
376
99%



K
383
99%



L
391
99%

