2-2021

Effect of Low-Passage Number on Dengue Consensus Genomes and Intra-host Variant Frequencies

Christian K. Fung

Tao Li

Simon Pollett

Maria Theresa Alera

In-Kyu Yoon

See next page for additional authors

Authors

Christian K. Fung, Tao Li, Simon Pollett, Maria Theresa Alera, In-Kyu Yoon, Jun Hang, Louis Macareo, Anon Srikiatkhachorn, Damon Ellison, Alan L. Rothman, Stefan Fernandez, Richard G. Jarman, and Irina Maljkovic Berry

JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Fung *et al.*, *Journal of General Virology* 2021;102:001553
DOI 10.1099/jgv.0.001553

MICROBIOLOGY SOCIETY

OPEN ACCESS

Effect of low-passage number on dengue consensus genomes and intra-host variant frequencies

Christian K. Fung[1], Tao Li[1], Simon Pollett[1], Maria Theresa Alera[2], In-Kyu Yoon[3], Jun Hang[2], Louis Macareo[2], Anon Srikiatkhachorn[2,4], Damon Ellison[1], Alan L. Rothman[4], Stefan Fernandez[2], Richard G. Jarman[1] and Irina Maljkovic Berry[1,*]

Abstract

Intra-host single nucleotide variants (iSNVs) have been increasingly used in genomic epidemiology to increase phylogenetic resolution and reconstruct fine-scale outbreak dynamics. These analyses are preferably done on sequence data from direct clinical samples, but in many cases due to low viral loads, there might not be enough genetic material for deep sequencing and iSNV determination. Isolation of the virus from clinical samples with low-passage number increases viral load, but few studies have investigated how dengue virus (DENV) culture isolation from a clinical sample impacts the consensus sequence and the intra-host virus population frequencies. In this study, we investigate consensus and iSNV frequency differences between DENV sequenced directly from clinical samples and their corresponding low-passage isolates. Twenty five DENV1 and DENV2 positive sera and their corresponding viral isolates (*T. splendens* inoculation and C6/36 passage) were obtained from a prospective cohort study in the Philippines. These were sequenced on MiSeq with minimum nucleotide depth of coverage of 500×, and iSNVs were detected using LoFreq. For both DENV1 and DENV2, we found a maximum of one consensus nucleotide difference between clinical sample and isolate. Interestingly, we found that iSNVs with frequencies ≥5% were often preserved between the samples, and that the number of iSNV positions, and sample diversity, at this frequency cutoff did not differ significantly between the sample pairs (clinical sample and isolate) in either DENV1 or DENV2 data. Our results show that low-passage DENV isolate consensus genomes are largely representative of their direct sample parental viruses, and that low-passage isolates often mirror high frequency within-host variants from direct samples.

INTRODUCTION

Dengue virus (DENV) is the etiologic agent of dengue fever (DF) as well as the more severe forms of illness, dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS), which cause significant health problems each year. Approximately 50–100 million symptomatic DENV infections occur each year, with the greatest burden in tropical and subtropical regions. DENV is an arbovirus, a mosquito-borne RNA virus commonly transmitted by *Aedes* mosquitoes. It belongs to the *Flavivirus* genus of the *Flaviviridae* family and has four serotypes, DENV1, 2, 3 and 4. DENV consists of a genome approximately 11 kb nucleotides (nt) in length. The genome encodes three structural proteins, nucleocapsid (C), precursor membrane (prM) and envelope (E), as well as seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) [1, 2].

Viral genome sequence data, including that of DENV, have increasingly been used in epidemics and outbreaks to provide more precise reconstructions of transmission dynamics and complement conventional non-genomic epidemiologic data.

Received 21 August 2020; Accepted 22 December 2020; Published 16 February 2021
Author affiliations: [1]Walter Reed Army Institute of Research, Silver Spring, MD, USA; [2]Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand; [3]Coalition for Epidemic Preparedness Innovations, Washington, DC, USA; [4]University of Rhode Island, Kingston, RI, USA.
***Correspondence:** Irina Maljkovic Berry, irina.maljkovicberry.ctr@mail.mil
Keywords: Dengue; evolution; intra-host variants; passage.
Abbreviations: AA, Access Array; AFRIMS, Armed Forces Research Institute of Medical Sciences; C, nucleocapsid; CHIKV, Chikungunya virus; DENV, Dengue virus; DF, Dengue Fever; DHF, Dengue Hemorrhagic Fever; DNA, Deoxyribonucleic acid; DSS, Dengue shock syndrome; E, Envelope; IFA, immunofluorescence assay; IGV, Integrative Genomics Viewer; iSNV, intra-host single nucleotide variant; NGS, Next generation sequencing; nt, nucleotide; PCR, Polymerase chain reaction; prM, precursor membrane; RNA, Ribonucleic acid; VCF, Variant call format; WRAIR, Walter Reed Army Institute of Research.
GenBank accession numbers: MT832032- MT832080.
Three supplementary tables and two supplementary figures are available with the online version of this article.
001553

This approach of genomic epidemiology for DENV has played a crucial role in analyses of spatial and temporal outbreak dynamics, virus dispersal tracking, genotype-phenotype associations, vaccine effectiveness, vector adaptation and many others [3–5]. Generally, genomic epidemiology studies analyse DENV consensus genomes derived from each host. However, DENV and other RNA viruses exist within a host as a population of intra-host variants, not as a single consensus genome. Advanced next-generation sequencing (NGS) technology can now detect intra-host viral variants present at low frequencies within a sample. Increasingly, the importance of these intra-host single nucleotide variants (iSNVs), and the amount of diversity they create within a host, has been highlighted in recent studies. For instance, Ko *et al*. described the presence and emergence of DENV intra-host variants with differing selection advantages affecting epidemic severity, and Descloux *et al*. suggested a correlation of DENV within-host diversity with disease severity [6, 7]. iSNVs have also been used for more precise tracking of viral transmission and spread within outbreaks and epidemics, providing more granular information on transmission chains and possible hotspots of viral dissemination [8, 9].

NGS of within-host viral populations is preferably done directly on clinical samples, such as blood, serum, nasopharyngeal swabs or saliva, cerebro-spinal fluid and others. However, in many cases, clinical samples may have low viral loads, thus not providing enough genetic material for deep sequencing and complete genome reconstruction with high depth of coverage, requirements that are needed for analyses of low-frequency variants within a sample. This creates bias in some DENV genomic epidemiological investigations as low-viremic individuals are excluded from genomic analysis yet likely play a major role in DENV epidemic dynamics [5, 10].

An alternative approach is to isolate the virus from clinical samples in cell culture and passage it a low number of times (typically less than three passages) in order to amplify the virus thereby allowing an increase in sequencing depth of coverage. Generally, it is thought that isolation and low-passaging does not change the consensus sequence of the virus. For instance, Vasilakis *et al*. investigated how passaging affects the consensus sequence of DENV, and although they did not compare direct clinical sample sequence to the sequence of the passaged isolates, they showed that passaging of isolates many times (n=5 or n=10) results in change of the virus consensus sequence while low-passage number (n=2) does not [11]. They suggested that these changes are most probably not due to a passaging bottleneck, but rather adaptation to cells. Importantly, they also showed that fewer passaging-induced consensus mutations are found in viruses passaged through C6/36 cells (*Aedes albopictus* vector) compared to passaging in human cell lines or vector-human alternating cell lines. Likewise, Chen *et al*. showed an increasing number of mutations in DENV consensus sequences with the increased number of isolate passages (n=20, 30) [12]. However, no studies to date have compared how the sequence of DENV virus from a direct clinical sample may change due to passaging.

In addition to the lack of knowledge on how DENV isolation from a clinical sample impacts the consensus sequence, there is no information on how the intra-host virus population may change with isolation and passaging. This information is scarce for arboviruses in general. Stapleford *et al*. compared Chikungunya virus (CHIKV) iSNVs in isolate samples, with single passage in mammalian or vector cell lines, with iSNVs in viruses from direct clinical samples [13]. They found that the high-frequency iSNVs were maintained over passaging. However, they also noted a decrease in overall diversity, with more diversity maintained when passaging one time through mammalian cells rather than vector cells. No such studies have been performed on DENV.

Because of the importance of accurate DENV consensus genomic data from a range of dengue case viral loads, as well as accurate estimations of the intra-host populations and minor variants, we sought to compare DENV consensus sequences from direct clinical samples (sera) to low-passage isolates derived from the same samples. In addition, we compare iSNV (minor variant) frequencies in DENV sequencing both in direct samples and their corresponding isolate genomes. This provides information on the possible consensus changes low-passaging of DENV from direct clinical samples might induce, as well as any changes in the frequency of intra-host population variants. Since low-passaged DENV isolates are frequently and increasingly used in investigations of DENV diversity and spread, this study provides important insight into the effect of passaging on the DENV genome sequence.

METHODS

Specimen collection and viral passaging

Eleven DENV1 culture-positive sera and their corresponding isolates, and 14 DENV2 culture-positive sera and their corresponding isolates, were received from the Armed Forces Research Institute of Medical Sciences (AFRIMS). These specimens were collected between 2012 and 2015 from a prospective cohort study of incident dengue illness set in Cebu, the Philippines (WRAIR no. 1844). The isolates had been passaged as follows: undiluted plasma/serum (0.34 µl/mosquito) was inoculated in 10–20 *Toxorhynchitis splendens* mosquitoes. Amplification of virus in mosquitoes was assessed by an immunofluorescence assay (IFA) assay on head squash preparations. The bodies of IFA-positive mosquitoes were macerated and inoculated onto C6/36 cells. Following one passage, the virus was serotyped, aliquoted and stored at −70 °C or below.

Virus sequencing

200 ul of each clinical specimen and corresponding cultured isolate was used for RNA extraction with QIAamp Viral RNA Mini Kit (Qiagen, MD, USA). Full DENV genomes were amplified via Access Array (AA) system (Fluidigm Corporation, CA, USA) using 48 pairs of in-house DENV serotype-specific primers and the SuperScript III One-Step

Table 1. Number of consensus differences and positions with iSNVs per DENV1 genome from direct samples and their corresponding isolates

Sample name	iSNVs ≥5% direct sample	iSNVs ≥5% isolate	iSNVs ≥1% direct sample	iSNVs ≥1% isolate	Consensus differences*
AB204803_D1_1	0	0	3	5	0
AB204803_D1_2	3	4	8	9	0
AB204803_D1_3	8	3	41†	7	0
AB204803_D1_4	2	1	5	2	0
AB204803_D1_5	0	0	6	4	0
AB204803_D1_6	1	1	5	7	0
AB204803_D1_7	1	1	6	2	0
AB204803_D1_8	3	3	4	6	0
AB204803_D1_9	0	19†	NA	NA	0
AB204803_D1_10	1	0	2	5	1
AB204803_D1_11	1	NA	NA	NA	NA

NA, Not assembled.
*Not including iSNV positions.
†Outlier sample.

RT-PCR system with Platinum Taq High Fidelity polymerase (ThermoFisher Scientific, MA, USA), and 35 cycles, followed by cDNA purification using Agencourt AMPure XP beads (Beckman Coulter, CA, USA). The PCR products were analysed using High Sensitivity DNA D5000 tapes (Agilent Technologies, CA, USA) on the Agilent Tapestation 4200 System (Agilent Technologies) to check cDNA quality and quantity. The NGS libraries were prepared using QIAseq FX DNA Library kit (Qiagen) following the manufacturer's instructions. Then, 50 ng per sample of amplified amplicon was used as input DNA for library preparation. The input DNA was fragmented for 15 min and amplified six cycles. The purified indexed libraries were quantified with Agilent Tapestation using DNA D5000 tapes (Agilent). Equal molar quantities of libraries were pooled based on the Tapestation data. Pooled libraries were denatured and diluted to a final loading concentration of 11.5pM and loaded onto the Miseq sequencing reagent cartridge from Miseq Reagent kit v3, 600 cycles, (illumina, CA, USA) for sequencing.

Virus genome assembly and variant calling

DENV reads from each sample were mapped to one of four references: one DENV1 reference (AB204803) and three DENV2 references (KU509277, KM279601 and KU517847), which were determined by first mapping the sequenced samples to a concatenated reference file containing numerous DENV1 and DENV2 genomes with location listed as Philippines and year 2012 to 2015 from GenBank. Using mapping output to support the best-fitting reference for each sample, the sequenced samples were mapped again individually to one of the four references. For the mapping and analysis, we used an in-house pipeline, NGS_Mapper that includes the variant caller LoFreq [14, 15]. Briefly, the default setting of 5/95 was used for the NGS_Mapper calling of intra-host nucleotide variants, allowing for analyses of variant frequencies per nucleotide position down to 5%. In addition, we performed a separate analysis of variants present in the population at a frequency 1% or higher. The minimum nucleotide Phred score threshold was set to 30. This means that any site containing a variant nucleotide of quality score greater than 30 and present at, or greater than, 5 or 1% would be annotated as an ambiguous position. In addition, the minimum depth of coverage to detect a variants was set to 500×. The consensus genomes were quality checked and manually curated using IGV and Geneious version R10, as well as each sample's VCF file for statistical support [16–18]. This post-assembly cleaning process of the final consensus was implemented to ensure removal of any nucleotide variants present due to primer-induced error (variants only occurring in the known primer regions and only found at the ends of the reads), certain types of sequencing error, and strand bias higher than what was observed in confident calls from each genome, respectively. In addition, minor nucleotide variants that were consistently present at either ends of the reads were manually removed from the analyses, as these have previously been shown to be spurious [19]. All genomes have been submitted to GenBank under accession numbers MT832032–MT832080. Shannon entropy was calculated for each sample using $H' = -\sum_{i=1}^{n} fi\,(lnfi)$, where *fi* is the observed frequency of an iSNV, and adjusting for genome length of 10179 nucleotides for DENV1 and 10176 nucleotides for DENV2.

Table 2. DENV1 iSNV CDS positions for ≥5% data, and their variant frequencies for samples mapped to reference AB204803

Sample	CDS nt position	Direct genome nt [AA]	Isolate genome nt [AA]
AB204803_D1_2	1206	G[V]95% A[V]=5%	G[V]=94% A[V]=6%
	1323	T[T]>95%*	T[T]=93% C[T]=7%
	2134	G[V]=86% A[M]=14%	G[V]=88% A[M]=12%
	8829	C[V]=37% T[V]=63%	C[V]=39% T[V]=61%
AB204803_D1_3	66	T[R]=6%; C[R]=94%	T[R]>95%
	844	A[R]=72%; C[R]=28%	A[R]>95%
	1806	G[V]=86%; A[V]=14%	G[V]=93%; A[V]=14%
	1809	A[Q]=88%; G[Q]=12%	A[Q]=92%; G[Q]=18%
	1962	G[K]=81%; T[K]=19%	G[K]>95%
	2061	C[R]=76%; T[R]=24%	C[R]>95%
	3016	C[L]=75% T[L]=25%	C[L]=81% T[L]=19%
	6761	G[R]=39% A[K]=61%	A[K]>95%
AB204803_D1_4	5442	G[E]=6% C[D]=94%	C[D]=>95%
	8310	C[G]=95% A[G]=5%	C[G]=95% A[G]=5%
AB204803_D1_6	8628	C[F]=19% T[F]=81%	C[F]=18% T[F]=82%
AB204803_D1_7	1655	C[T]=60% T[M]=40%	C[T]=60% T[M]=40%
AB204803_D1_8	3512	C[T]=29% T[I]=71%	C[T]=19% T[I]=81%
	6750	G[L]=51% A[L]=49%	G[L]=42% A[L]=58%
	9261	A[R]=89% G[R]=11%	A[R]=88% G[R]=12%
AB204803_D1_10	3306	T[D]=14%; C[D]=86%	C[D]>95%

AA, amino acid; NT, nucleotide.
*Minor variants were present, albeit at a frequency <5%.

RESULTS

Full genomes were obtained for all 11 DENV1 viruses derived from the direct clinical samples and for 10 of their corresponding virus isolates (*T. splendens* inoculation followed by one C6/36 passage). Full genomes were obtained from 14 DENV2 viruses derived from the direct clinical samples and from all of their corresponding virus isolates (*T. splendens* inoculation followed by one C6/36 passage). The overall depth of genome coverage for all obtained sequences was >500× throughout the genome, and positions with intra-host Single Nucleotide Variants (iSNVs) were scanned for. Variants present at a frequency of 5% or higher, and 1% or higher, were called based on criteria described in Methods.

DENV1 consensus and iSNVs in direct sample versus corresponding culture

Of 22 DENV1 samples (11 direct-culture pairs), one genome, from the AB204803_D1_11 cultured virus, was not assembled. All except one direct-culture pair had identical consensus genomes (Table 1). One genome, from the AB204803_D1_9 cultured virus, contained an unusual number of iSNV positions already at 5% level cutoff and was deemed an outlier, pointing to possible contamination of this sample (Table 1). Thus, nine DENV1 samples were used for DENV1 iSNV comparisons between viruses sequenced directly from clinical samples and their corresponding cultured isolates. For 5% cutoff analyses, two of the samples did not have any iSNV positions in the genomes derived from either the direct sample or the cultured virus (Table 1). For genomes from direct samples, the average number of iSNV positions was 2.1 (range 0–8) per genome (Table 1). For genomes from their corresponding isolates, the average number of iSNV positions was 1.4 (range 0–4) per genome (Table 1) and it did not differ significantly from the number of iSNV positions in the direct samples (P >0.05, f-test). On the 1% variant frequency cutoff level, one additional direct sample (AB204803_D1_3) was found to be an outlier containing 33 iSNV positions. However, since this sample was included in the 5% analyses, we performed statistical analyses for the 1% data both including and excluding this sample. Excluding this outlier, the average number of iSNV positions was 4.9 (range 2–8) for the direct samples and 5.0 (range 2–9) for their isolates, not differing statistically (P >0.05). Including

Table 3. Number of consensus differences and positions with iSNVs per DENV2 genome from direct samples and their corresponding isolates

Sample name	iSNVs ≥5% direct sample	iSNVs ≥5% isolate	iSNVs ≥1% direct sample	iSNVs ≥1% isolate	Consensus differences*
KU509277_D2_1	3	3	5	3	0
KU509277_D2_2	3	3	6	6	0
KU509277_D2_3	2	2	33	3	0
KU517847_D2_1	5	3	17	7	0
KM279601_D2_1	10	6	81†	9	1
KM279601_D2_2	5	2	42	2	1
KM279601_D2_3	3	1	12	1	1
KM279601_D2_4	5	3	15	9	1
KM279601_D2_5	4	5	26	6	1
KM279601_D2_6	1	3	3	3	0
KM279601_D2_7	1	0	22	1	0
KM279601_D2_8	2	2	2	2	0
KM279601_D2_9	4	4	9	6	0
KM279601_D2_10	1	2	9	6	1

*Not including iSNV positions.
†Outlier sample.

the outlier, the average number of iSNV positions was 8.9 (range 1–41) for the direct samples, making the number of iSNVs significantly higher for the direct samples (P <0.05).

We compared the positions with iSNVs, and variant frequencies within those positions, between the genomes from DENV 1 direct samples and their corresponding cultures, at both 5 and 1% variant frequency cutoffs. For the variant frequency cutoff of 5%, we found that many iSNV positions overlapped between the direct sample and isolate viral populations (Table 2). Out of a total of 19 positions with iSNVs in the clinical DENV1 samples, 12 were also present in their isolates. Isolate samples had a total of 13 iSNV positions of which 12 were also found in the clinical samples. (Table 2). In eight of the positions, a minor variant present at the frequency of ≥5% was found exclusively in either the direct sample or the isolate genome. Interestingly, some of these positions were confirmed to actually contain the corresponding variants, albeit at frequencies <5% (Table S1, available in the online version of this article). The iSNV position overlap between direct sample and its isolate was less frequent for variants occurring at a frequency 1–5% (Table S1).

Shannon entropy was calculated to measure the amount of diversity within each sample. The entropy did not significantly differ between direct samples and isolates at 5% frequency cutoff (P >0.05)(Fig. S1). At 1% frequency cutoff, the Shannon entropy was significantly higher for the direct samples (P=0.01), however, when the AB204803_D1_3 outlier was removed from the calculation, the Shannon entropy was not different between direct samples and their isolates (P >0.05) (Fig. S1).

DENV2 consensus and iSNVs in direct sample versus corresponding culture

Of 28 DENV2 samples (14 direct-culture pairs), all genomes were available for consensus and iSNV comparisons between clinical samples and their corresponding cultured viruses. There were 0–1 consensus genome differences found between the sample pairs (Table 3). For the 5% variant frequency cutoff level, the average number of iSNV positions was 3.5 (range 1–10) for direct samples. For genomes from their corresponding isolates, the average number of iSNV positions was 2.8 (range 0–6) per genome (Table 3). The number of iSNV positions did not differ significantly between direct samples and their isolates (P >0.05, f-test). On the 1% variant frequency cutoff level, one additional direct sample (KM279601_D2_1) was found to be outlier containing 81 iSNV positions. However, since this sample was included in the 5% analyses, we performed statistical analyses for the 1% data both including and excluding this sample. Excluding this outlier, the average number of iSNV positions was 15.5 (range 2–42) for the direct samples and 4.2 (range 0–9) for their isolates, differing statistically (P <0.05). Including the outlier, the average number of iSNV positions was 20.1 (range 2–81) for the direct samples and 4.6 (range 1–9) for the isolates, also significantly different (P <0.05).

We compared the positions with iSNVs, and variant frequencies within those positions, at both 5 and 1% variant frequency cutoffs, for the genomes from DENV2 direct samples and their corresponding cultures (Tables 4, S2 and S3). For 5% cutoff, 24 of 49 direct sample iSNV positions were also found in the isolates, and 24 of 39 isolate iSNV positions

Table 4. DENV2 iSNV CDS positions and their variant frequencies, for ≥5% data, for samples mapped to references KU509277 or KU517847

Sample	CDS nt position	Direct genome	Isolate genome
KU509277_D2_1	2995	T[T]=5%; C[T]=95%	C[T]>95%
	4056	T[A]=72% C[A]=28%	T[A]=72% C[A]=28%
	8085	T[A]=72% C[A]=28%	T[A]=76% C[A]=24%
	8308	A[K]>95%	G[K]=5%; A[K]=95%
KU509277_D2_2	1014	G[K]=72% A[K]=28%	G[K]=86% A[K]=14%
	2338	G[V]=91% A[I]=10%	G[V]=93% A[I]=7%
	4374	A[S]=91% G[S]=9%	A[S]=94% G[S]=6%
KU509277_D2_3	1059	C[R]=>95%*	C[R]=94% A[R]=6%
	1098	G[Q]=95% A[Q]=5%	G[Q]=>95%
	8642	C[T]=84% T[I]=16%	C[T]=82% T[I]=18%
KU517847_D2_1	512	A[N]=5%; G[D]=95%	A[N]=5%; G[D]=95%
	4617	G[R]=6%; A[R]=94%	G[R]=5%; A[R]=95%
	4959	T[I]=15% A[I]=85%	T[I]=12% A[I]=88%
	6128	G[R]=6%; A[K]=94%	A[K]>95%*
	7591	C[L]=6%; T[L]=94%	T[L]>95%

NT, nucleotide; AA, amino acid.
*Minor variants were present, albeit at a frequency <5%.

were also confirmed in the direct samples. As for DENV1, we observed DENV2 positions with a minor variant present at the frequency of ≥5% found exclusively in either the direct sample or the isolate genome, however, some of these positions did have variants below 5% frequency.

At 5% level, there were no significant differences between DENV1 and DENV2 in the number of iSNV positions per genome in either direct or isolate samples (P >0.05). Similarly to the results for DENV1, the iSNV position overlap between direct sample and its isolate was less frequent for DENV2 variants occurring at a frequency 1–5% (Table S3). Shannon entropy was calculated to measure the amount of diversity within each sample. The entropy did not significantly differ between direct samples and isolates (including the KM279601_D2_1 outlier) at 5% frequency cutoff, but was significant at the 1% frequency cutoff (Fig. S2).

DISCUSSION

Genomic epidemiology is increasingly being used in dengue and other viral outbreaks to provide complementary insights into viral transmission and epidemic dynamics [4, 20]. Such information can be used for more precise tracking of viral transmission and locating possible hotspots of viral dissemination. However, consensus genomes provide limited epidemiological inference, particularly when sampled from infected cases closely linked in space and time [5]. To circumvent this problem and increase resolution, iSNVs have recently been used for more precise reconstructions of virus transmission chains [6–9]. For such studies, sequencing of viral genomes from clinical samples is ideal, however, the amount of genetic material may not be sufficient to achieve depth that is required for iSNV identification and analysis. Therefore, viruses from clinical samples may be amplified in cell culture, to achieve sufficient nucleic acid yields for primary analysis, as well as to preserve the virus for additional studies. Often, research reviews highlight sequencing of isolates as a limitation to direct sequencing of clinical samples, despite a relative absence of data showing passaged viruses having a different consensus sequence to its clinical parent. In this study we aimed to investigate whether low-passaging can change DENV genome consensus sequence, or the minor variant admixture, compared to a direct clinical specimen.

We compared virus from direct clinical samples and their corresponding cultured viruses for both DENV1 and DENV2, and found that there were no significant differences in the number of iSNV positions, or in the Shannon entropy, between the sample pairs (direct and isolate) when looking at variants present at 5% or higher. This diversity preservation was also true when variant frequency cutoff was 1% upon removal of outlier sample for DENV1, but not for DENV2. This suggests there was no increase or decrease in ≥5% iSNV diversity following amplification of DENV through *T. splendens* inoculation and one C6/36 passage, however, the isolate diversity did decrease for DENV2 when including variants at frequencies 1–5%. These results are mainly in concordance with a previous study showing that the amount of dengue

intra-host diversity generally does not change when the virus is transmitted from human to mosquito [21]. Also in concordance with previous human–mosquito transmission studies, we found that DENV iSNVs occurring at higher frequencies (≥5%) were often preserved after mosquito and C6/36 culture passage [13, 21], although instances with iSNVs that were exclusive to either direct samples or cultured samples were observed. This was more pronounced when looking at variants present at lower frequencies (1–5%), suggesting that like in human–mosquito transmission, virus isolation may result in removal of low-frequency variants, followed by eventual restoration of diversity upon virus replication in culture, albeit with a different low-frequency iSNV repertoire. The lower diversity at 1% cutoff in DENV2 isolate samples might mirror this purge of low-frequency minor variants from direct samples, where the iSNV repertoire had not been restored yet. The discrepant results between DENV1 and DENV2 diversity at 1% variant frequency cutoff may be due to differences between the two serotypes, or may be that the threshold for low-frequency iSNV restoration is somewhere between 1–2 passages. This becomes important to consider in studies utilizing within-host variation for more precise transmission inferences. However, we show that higher frequency iSNVs (≥5%) were often preserved, and if these are found in low-passage isolates from epidemiologically linked cases, they may represent true instances of minor variant transmission.

We found none to very few (up to one per genome) consensus sequence changes between direct sample and isolate viruses. Consensus conservation has also been observed in human–mosquito transmission studies, although it has been suggested that genome changes might occur and also differ depending on the mosquito species [21, 22]. A single nucleotide difference would not have a great impact on phylogenetic reconstructions of DENV full genomes for molecular epidemiology studies, and our study shows that low virus passage may be used for phylogenetic inferences when not enough viral material is available in clinical samples. However, a single (nonsynonymous) nucleotide difference might have epidemiological importance in other contexts, such as in arbovirus adaptation to the vector [23, 24]. In this study, the viruses were not passaged more than once, and a low-passage approach is commonly used for virus amplification for genomic surveillance studies [3, 11]. Further studies can be performed to investigate the impact of increasing number of passages on DENV consensus sequence and maintenance of high- and low-frequency variants. In addition, further analyses on other DENV serotypes (DENV3 and DENV4) are needed, as well as confirmation of our results using other *in vitro* passage protocols and approaches that more thoroughly account for PCR, sequencing and basecalling algorithm errors.

In conclusion, we show that limited culture passaging had minimal effect on the consensus sequence of DENV from direct clinical samples, and that iSNV frequencies at ≥5% were often maintained during passage. These data provide more confidence in using virus isolates as an alternative approach to deep sequence DENV intra-host viral populations, and mitigate epidemiological sampling limitations previously constrained by dengue case viral load. Our results serve as an important technical reference for DENV genomic epidemiology, virus dispersal tracking and vaccine effectiveness studies.

Funding information
This work was funded by the Armed Forces Health Surveillance Branch (AFHSB) and its Global Emerging Infections Surveillance (GEIS) Section, FY2019 ProMIS ID P0116_19_WR and the National Institutes of Health grant P01AI034533.

Acknowledgement
Disclaimer: This material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the authors, and are not to be construed as official, or as reflecting true views of the Department of the Army, the Department of Defense or the National Institutes of Health.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
The investigators have adhered to the policies for protection of human subjects as prescribed in AR 70–25. The work has been reviewed by the Human Subject Protection Branch, WRAIR, under protocol #1844.

References

1. Chen R, Vasilakis N. Dengue-quo tu et quo vadis? *Viruses* 2011;3:1562–1608.
2. Halstead S. Recent advances in understanding dengue. *F1000Res* 2019;8:1279.
3. Salje H, Lessler J, Maljkovic Berry I, Melendrez MC, Endy T *et al*. Dengue diversity across spatial and temporal scales: local structure and the effect of host population size. *Science* 2017;355:1302–1306.
4. Pollett S, Fauver JR, Maljkovic Berry I, Melendrez M, Morrison A *et al*. Genomic epidemiology as a public health tool to combat mosquito-borne virus outbreaks. *J Infect Dis* 2020;221:S308–S318.
5. Pollett S, Melendrez MC, Maljkovic Berry I, Duchêne S, Salje H *et al*. Understanding dengue virus evolution to support epidemic surveillance and counter-measure development. *Infect Genet Evol* 2018;62:279–295.
6. Ko H-Y, Li Y-T, Chao D-Y, Chang Y-C, Li Z-RT *et al*. Inter- and intra-host sequence diversity reveal the emergence of viral variants during an overwintering epidemic caused by dengue virus serotype 2 in southern Taiwan. *PLoS Negl Trop Dis* 2018;12:e0006827.
7. Descloux E, Cao-Lormeau VM, Roche C, De Lamballerie X. Dengue 1 diversity and microevolution, French Polynesia 2001-2006: connection with epidemiology and clinics. *PLoS Negl Trop Dis* 2009;3:e493.
8. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A *et al*. Tracking virus outbreaks in the twenty-first century. *Nat Microbiol* 2019;4:10–19.
9. Ladner JT, Grubaugh ND, Pybus OG, Andersen KG. Precision epidemiology for infectious disease control. *Nat Med* 2019;25:206–211.
10. Ten Bosch QA, Clapham HE, Lambrechts L, Duong V, Buchy P *et al*. Contributions from the silent majority dominate dengue virus transmission. *PLoS Pathog* 2018;14:e1006965.
11. Vasilakis N, Deardorff ER, Kenney JL, Rossi SL, Hanley KA *et al*. Mosquitoes put the brake on arbovirus evolution: experimental evolution reveals slower mutation accumulation in mosquito than vertebrate cells. *PLoS Pathog* 2009;5:e1000467.
12. Chen WJ, Wu H-R, Chiou S-S. E/NS1 modifications of dengue 2 virus after serial passages in mammalian and/or mosquito cells. *Intervirology* 2003;46:289–295.
13. Stapleford KA, Moratorio G, Henningsson R, Chen R, Matheus S *et al*. Whole-Genome sequencing analysis from the Chikungunya

virus Caribbean outbreak reveals novel evolutionary genomic elements. *PLoS Negl Trop Dis* 2016;10:e0004402.

14. GitHub - VDBWRAIR/ngs_mapper. Available from: https://github.com/VDBWRAIR/ngs_mapper.
15. Wilm A, Aw PPK, Bertrand D, Yeo GH, Ong SH *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;40:11189–11201.
16. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M *et al*. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647–1649.
17. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES *et al*. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–26.
18. Thorvaldsdottir H, Robinson JT, Mesirov JP, Viewer IG. IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–192.
19. McCrone JT, Lauring AS. Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling. *J Virol* 2016;90:6884–6895.
20. Maljkovic Berry I, Melendrez MC, Bishop-Lilly KA, Rutvisuttinunt W, Pollett S *et al*. Next generation sequencing and bioinformatics methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *J Infect Dis* 2020;221:S292–S307.
21. Sim S, Aw PPK, Wilm A, Teoh G, Hue KD *et al*. Tracking dengue virus Intra-host genetic diversity during Human-to-Mosquito transmission. *PLoS Negl Trop Dis* 2015;9:e0004052.
22. Sessions OM, Wilm A, Kamaraj US, Choy MM, Chow A *et al*. Analysis of dengue virus genetic diversity during human and mosquito infection reveals genetic constraints. *PLoS Negl Trop Dis* 2015;9:e0004044.
23. Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A single mutation in Chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog* 2007;3:e201.
24. Maljkovic Berry I, Eyase F, Pollett S, Konongoi SL, Joyce MG *et al*. Global outbreaks and origins of a chikungunya virus variant carrying mutations which may increase fitness for *Aedes aegypti*: revelations from the 2016 mandera, kenya outbreak. *Am J Trop Med Hyg* 2019;100:1249–1257.