

# Chikungunya virus evolution following a large 3'UTR deletion results in host-specific molecular changes in protein-coding regions

Valerie J. Morley,<sup>1,†</sup> María Gabriela Noval,<sup>2</sup> Rubing Chen,<sup>3</sup> Scott C. Weaver,<sup>3</sup> Marco Vignuzzi,<sup>4,‡</sup> Kenneth A. Stapleford,<sup>2</sup> and Paul E. Turner<sup>1,5,\*,&</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 06511-8934, USA, <sup>2</sup>Department of Microbiology, New York University, New York, NY, USA, <sup>3</sup>Institute for Human Infections and Immunity and Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, TX, USA, <sup>4</sup>Viral Populations and Pathogenesis Unit, Institut Pasteur, Paris, France and <sup>5</sup>Program in Microbiology, Yale School of Medicine, New Haven, CT, USA

\*Corresponding author: E-mail: paul.turner@yale.edu

†<http://orcid.org/0000-0001-6805-7562>

‡<http://orcid.org/0000-0002-4400-771X>

§<http://orcid.org/0000-0003-3490-7498>

## Abstract

The 3'untranslated region (UTR) in alphavirus genomes functions in virus replication and plays a role in determining virus host range. However, the molecular evolution of virus UTRs is understudied compared to the evolution of protein-coding regions. Chikungunya virus (CHIKV) has the longest 3'UTR among the alphaviruses (500–700 nt), and 3'UTR length and sequence structure vary substantially among different CHIKV lineages. Previous studies showed that genomic deletions and insertions are key drivers of CHIKV 3'UTR evolution. Inspired by hypothesized deletion events in the evolutionary history of CHIKV, we used experimental evolution to examine CHIKV adaptation in response to a large 3'UTR deletion. We engineered a CHIKV mutant with a 258 nt deletion in the 3'UTR ( $\Delta$ DR1/2). This deletion reduced viral replication on mosquito cells, but did not reduce replication on mammalian cells. To examine how selective pressures from vertebrate and invertebrate hosts shape CHIKV evolution after a deletion in the 3'UTR, we passaged  $\Delta$ DR1/2 virus populations strictly on primate cells, strictly on mosquito cells, or with alternating primate/mosquito cell passages. We found that virus populations passaged on a single host cell line increased in fitness relative to the ancestral deletion mutant on their selected host, and viruses that were alternately passaged improved on both hosts. Surprisingly, whole genome sequencing revealed few changes in the 3'UTR of passaged populations. Rather, virus populations evolved improved fitness through mutations in protein coding regions that were associated with specific hosts.

**Key words:** arbovirus; experimental evolution; emerging pathogen; adaptation; alphavirus.

## 1. Introduction

Arthropod-borne viruses (arboviruses) have highly variable 3'untranslated regions (3'UTRs) that function in viral replication and virus–host interactions (Hyde et al. 2015). The 3'UTR is functionally important, but dramatic variation in 3'UTR sequence among virus lineages suggests that 3'UTR evolution is less constrained than sequence evolution in coding regions. In fact, experimental studies have shown that viruses with large deletions in the 3'UTR are still viable, although deletion mutants are often attenuated (Kuhn, Hong, and Strauss 1990; Kuhn et al. 1992; Men et al. 1996; Mandl et al. 1998; George and Raju 2000; Lo et al. 2003; Chen et al. 2013).

In alphaviruses (family *Togaviridae*, genus *Alphavirus*) and flaviviruses (family *Flaviviridae*, genus *Flavivirus*), 3'UTRs are often characterized by repeated sequence elements (RSEs) (Ou, Trent, and Strauss 1982; Pfeffer, Kinney, and Kaaden 1998; Gritsun and Gould 2006a). Across different alphaviruses, RSEs vary in number (2–5) and length (18–102 nucleotides) (Ou, Trent, and Strauss 1982; Pfeffer, Kinney, and Kaaden 1998; Hyde et al. 2015). This pattern of sequence repetition, along with variation for number of RSEs among related lineages, suggests that duplications and deletions are major mechanisms for 3'UTR sequence evolution in alphaviruses and flaviviruses (Gritsun and Gould 2006b; Chen et al. 2013; Stapleford et al. 2016).

While the functional role of the 3'UTR remains incompletely understood, it has been shown to be important in determining host range. Alphavirus and flavivirus 3'UTRs interact with host proteins in both mammalian and mosquito cells (Pardigon and Strauss 1992; Pardigon, Lenches, and Strauss 1993; Garneau et al. 2008; Sokoloski et al. 2010; Lei et al. 2011). Additionally, the 3'UTR can interact with host-encoded microRNAs, which can alter viral gene expression and replication (Trobaugh et al. 2014). Repeatedly, experimental work has demonstrated that deletions in the 3'UTR tend to reduce replication on mosquito cells, while having little or no impact on replication on mammalian cells (Kuhn, Hong, and Strauss 1990; Weaver et al. 1999; Chen et al. 2013; Trobaugh et al. 2014; Villordo et al. 2015). In one instance, a lineage of eastern equine encephalitis virus that was passaged strictly on mammalian fibroblast cells evolved a 238 nt deletion in 3'UTR, which suggests that this region is dispensable for replication on mammalian cells (Weaver et al. 1999). In flaviviruses the presence of RSEs may be a genetic strategy for obligate cycling between two hosts, in which different repeats can be specialized for function in different host types (Villordo et al. 2015), and 3'UTRs are involved in the production of subgenomic flavivirus RNAs (sfRNAs) (Bavia et al. 2016).

In the current study we focus on chikungunya virus (CHIKV; family *Togaviridae*, genus *Alphavirus*), which has especially diverse 3'UTR sequences across different lineages (Chen et al. 2013; Hyde et al. 2015; Stapleford et al. 2016). CHIKV cycles obligately between primate hosts and *Aedes* spp. mosquito vectors, and it can cause debilitating, chronic febrile illness in humans (Griffin 2013). CHIKV is a nonsegmented, positive-sense, single-stranded RNA virus with two open reading frames. The first open reading frame encodes a polyprotein that is cleaved to form four nonstructural proteins involved in viral replication: nsP1, nsP2, nsP3, and nsP4 (Kuhn 2013). The second open reading frame encodes a polyprotein that is cleaved to form five structural proteins: the capsid protein, the E3 envelope protein, the E2 envelope protein, the membrane-associated protein 6K/TF, and the E1 envelope protein (Kuhn 2013). The CHIKV 3'UTR is the longest in the alphavirus genus, ranging from 500 to

700 nt, with the number of RSEs varying among lineages (Chen et al. 2013; Hyde et al. 2015; Stapleford et al. 2016).

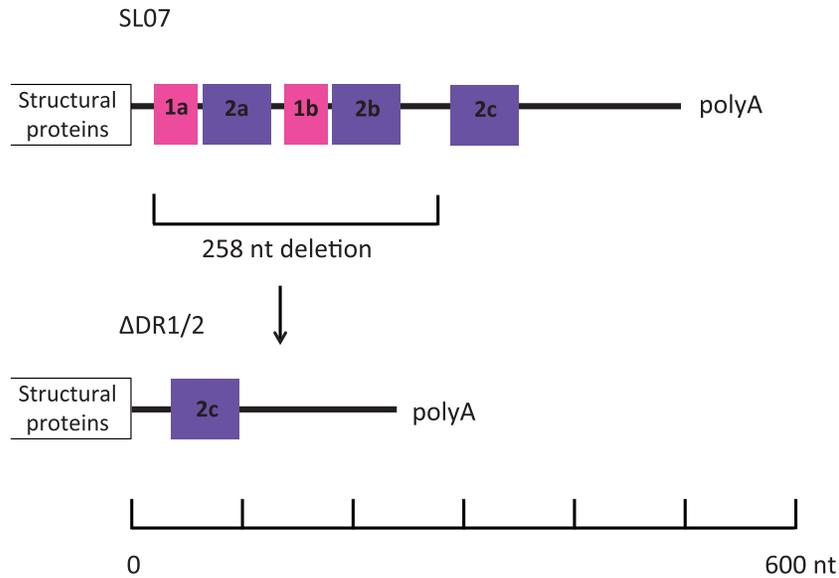
Insertions and deletions are hypothesized to be key mutational mechanisms driving CHIKV 3'UTR evolution and diversification (Chen et al. 2013; Stapleford et al. 2016). For example, compensatory evolution in response to a large 3'UTR deletion was hypothesized to have generated the unique 3'UTR sequence patterns observed in the Asian CHIKV lineage (Chen et al. 2013). The current study was inspired by this and other examples of compensatory evolution following the deletion of 3'UTR RSEs, and we sought to further examine how selective pressures from vertebrate hosts versus invertebrate vectors drive this evolution.

We constructed a mutant with a 258 nt deletion in the 3'UTR from the SL07 CHIKV strain of the Indian Ocean lineage (Fig. 1). This deletion removes two RSEs from the genome ( $\Delta$ DR1/2), leaving one remaining RSE. This deletion is not known to be naturally occurring; it was designed as an extreme example of an RSE deletion, similar to the single RSE deletion described in Chen et al. (2013) hypothesized to have accompanied as a founder effect the CHIKV introduction into Asia many decades ago. We hypothesized that this deletion would impair virus replication on mosquito cells but not on mammalian cells, and that the virus could improve its replication on mosquito cells through compensatory evolution. We used the  $\Delta$ DR1/2 virus to found twenty-four replicate virus populations in each of three treatments (72 populations total): strict passage on African green monkey derived cells (Vero) representing the vertebrate host, strict passage on *Aedes aegypti* cells (Aag2) representing the vector, or alternating passage between the two cell lines. We hypothesized that virus populations evolved in treatments with mosquito cells (Aag2 and alternating) would undergo compensatory mutations to improve replication on mosquito cells. These compensatory mutations might include duplications of the remaining RSE in the 3'UTR to restore a genotype similar to the wild type. Such duplications would likely be rare events; thus the experiment was designed with a large number of replicate virus populations per treatment to increase the likelihood of observing rare mutations. Virus populations in the primate and alternating treatments were allowed to evolve for twenty-five experimental passages. Due to difficulty sustaining virus populations that produced low titers in the mosquito cell passages, viruses were allowed to evolve on Aag2 cells for just seven passages. All endpoint passaged virus populations were assayed for fitness relative to  $\Delta$ DR1/2 on both host types, and all populations were sequenced with whole-genome next-generation sequencing.

## 2. Materials and methods

### 2.1 Generation of recombinant virus $\Delta$ DR1/2

The virus used to found all experimentally evolved lines was constructed by modifying an infectious clone (IC) of CHIKV strain SL07 (Tsetsarkin and Weaver 2011) (Fig. 1). Briefly, two *AvrII* restriction sites were inserted at the 3' and 5' ends of DR1 (RSE 1a + 2a) and DR2 (RSE 1b + 2b) through site-directed mutagenesis. These repeats were excised, the plasmid was religated, and the *AvrII* site was removed through site-directed mutagenesis. The resulting virus has DR1 and DR2 deleted. A marked version of the recombinant virus was used for fitness assays. The virus was marked by introducing a synonymous nucleotide (C to T) via site-directed mutagenesis at nucleotide



**Figure 1.** Map of CHIKV 3'UTR RSEs and cloning strategy. A map of the CHIKV 3'UTR is shown, with RSEs shown as colored boxes. Regions with the same sequence are coded with the same color. A 258 nucleotide region containing two RSEs was deleted from the SL07 clone to create the deletion mutant ΔDR1/2.

1980 of the nsP2 gene (nsP2 amino acid 660) to add a SacII restriction site (Coffey and Vignuzzi 2011).

To generate infectious virus from the CHIKV IC, 10 µg of purified plasmid was linearized with NotI, and the purified linear DNA was transcribed from an SP6 promoter with a mMACHINE mMACHINE kit (Ambion). RNA was transfected into BHK-21 cells using Lipofectamine RNAiMAX reagent (ThermoFisher Scientific). At 48–72 h post-infection, cell culture supernatants containing virus were harvested, aliquoted, and frozen at  $-80^{\circ}\text{C}$ . All work with infectious CHIKV was carried out in a biosafety level 3 laboratory.

## 2.2 Cell culture

**Host cells:** CHIKV was cultured on two host cell lines: Vero and Aag2. Additionally, BHK-21 cells and C6/36 cells were used in growth curve analysis and for plaque assays. Aag2 cells are derived from the mosquito *A. aegypti* (kindly provided by E. Fikrig, Yale University). Vero cells are an African green monkey kidney epithelial line (ATCC #CCL-81). BHK-21 is a golden hamster kidney derived cell line that is highly permissive for virus replication (kindly provided by provided by John J. Holland and Isabel Novella). C6/36 is derived from the mosquito *Aedes albopictus* (kindly provided by E. Fikrig, Yale University). All cell lines were cultured in Dulbecco's modified Eagle medium (ThermoFisher Scientific) supplemented with 10 per cent fetal bovine serum and 1 per cent penicillin and streptomycin. For each transfer cells were seeded in 24 well tissue-culture plates, then allowed to grow for 24 h to achieve confluence prior to virus infection. Mammalian and mosquito cells were incubated at  $37^{\circ}\text{C}$  and  $28^{\circ}\text{C}$ , respectively, with both at 95 per cent relative humidity and 5 per cent  $\text{CO}_2$  atmosphere.

## 2.3 Growth curves

24-Well culture plates were seeded with cells 24 h before infection and allowed to grow to confluence. Each well was infected with  $\sim 10^4$  pfu CHIKV in serum free media (multiplicity of infection (MOI)  $\sim 0.05$ ). Plates were incubated with inoculum for 1 h at  $37^{\circ}\text{C}$  or  $28^{\circ}\text{C}$ , after which the inoculum was removed and

replaced with 1 ml media. Three replicate wells were infected for each planned time point. Cell culture supernatant was harvested at the reported time points and stored at  $-80^{\circ}\text{C}$ . Virus was quantified from thawed samples via plaque assay.

## 2.4 Experimental evolution of virus populations

The ΔDR1/2 virus was used to found twenty-four replicate lineages in each of three treatments (seventy-two lineages total). Virus populations were either passaged strictly on Vero cells, strictly on Aag2 cells, or in alternating passages between Vero and Aag2 cells. To initiate each treatment population, a cell monolayer was infected with the ancestral virus at  $\text{MOI} \sim 0.05$ . The infected monolayer was incubated at  $37^{\circ}\text{C}$  (Vero) or  $28^{\circ}\text{C}$  (Aag2) for 48 h. Then, a subset of each virus population ( $\sim 10^4$  virus particles per transfer) was transferred to fresh cells. A subset of virus lineages (four per treatment) were titered every five passages to ensure that the number of transferred viruses was similar among treatments. A different set of lineages was titered at each checkpoint. This process was repeated for twenty-five passages total in the Vero and alternating treatment. Due to declining titers, passaging in the Aag2 treatment was halted after seven passages. At the end of the experiment aliquots of supernatant containing the virus progeny were drawn from each lineage and frozen at  $-80^{\circ}\text{C}$  for future analysis.

## 2.5 Plaque assays

Viral titers (plaque forming units (PFU) per ml) were estimated using plaque assays, in which serially diluted samples were plated on BHK cells under DMEM medium with 10 per cent fetal bovine serum and solidified with 1 per cent agarose, and incubated for 36 h using the above-described conditions. After incubation, cells were fixed with 10 per cent formaldehyde, media and agarose were removed, and plates were stained with crystal violet to visualize plaques. Each plaque was assumed to have originated from a single infecting virus.

## 2.6 Fitness assays

Relative fitness of virus populations was estimated through competition assays with a marked competitor (protocol modified from Coffey and Vignuzzi (2011)). Two marked competitor viruses were generated: a marked version of the wild type SL07 virus and a marked version of the deletion mutant  $\Delta$ DR1/2. The marked clones were engineered to have a synonymous nucleotide substitution that introduces a SacII restriction site. For experiments where mutations of interest were engineered into both wild-type SL07 and  $\Delta$ DR1/2 genetic backgrounds, engineered viruses were always assayed with a competitor with a matching genetic background (i.e. SL07 was competed with SacII-SL07, and  $\Delta$ DR1/2 was competed with SacII- $\Delta$ DR1/2). For all assays, a 50:50 mixture of the virus of interest and marked competitor was used to inoculate replicate Vero wells and replicate Aag2 wells in 48-well culture plates at MOI = 0.01. For passaged virus populations, assays were replicated in triplicate on each host. For assays with engineered viruses, a total of seven replicates were run on each host over two experimental blocks. Plates were incubated at 37°C (Vero) or 28°C (Aag2) for 48 h, at which point cell culture supernatant was harvested and frozen at -80°C. Viral RNA was extracted from each sample using a PureLink Pro 96 kit (ThermoFisher). The region surrounding the SacII restriction site was amplified via RT-PCR using a Titan One Tube RT-PCR kit (Roche) and primers forward 5'-GTGCGGCTTC TTCAATATG-3' and reverse 5'-CACTGTTCTTAAAGGACTC-3'. The relative abundance of each competitor was measured by the intensity of SacII-digested DNA fragments. About 200 ng purified DNA was digested for 2 h at 37°C. The amount of cut (marked competitor) versus uncut (evolved strain) virus was quantified by running each sample on a 1 per cent gel stained with ethidium bromide. Band intensities were quantified with the BioRad ChemiDoc system. Inoculum samples were run alongside endpoint samples. Fitness ( $\ln W$ ) is represented as the natural log of the output/input ratio of the evolved/marked virus following replication in a given cell line. A fitness value of >0 indicates that the passaged virus is more fit in the assay environment than the marked competitor. In mixed amplicons samples, the cut bands can be detected at ratios down to 1:70, as determined by measuring known ratios of cut and uncut amplicons.

## 2.7 Library preparation and sequencing

Genomic viral RNA was isolated from ninety-seven virus samples (ancestral virus stock and twenty-four each from Vero Passage 25, Alternating Passage 24, Alternating Passage 25, and Aag2 Passage 7) using the QIAamp Viral RNA mini kit (Qiagen). cDNA was generated by reverse-transcription with Superscript II (Life Technologies) using random hexamer primers. The entire genome sequence was amplified via PCR using eight primer pairs, generating overlapping PCR fragments 1.5–2.1 kb in length (Supplementary Fig. S2). Some samples with low viral RNA concentrations did not successfully amplify with these reactions, and they were instead amplified in smaller fragments (Supplementary Fig. S3). Amplified genome fragments were purified using the AMPure XP beads (Beckman Coulter), and the eight fragments from each virus sample were combined to create an equimolar mixture. Libraries were prepared from the pooled viral amplicons using the Nextera XT Kit (Illumina Inc.) and the Nextera Index Kit (Illumina Inc.). Samples were sequenced in a single lane via paired-end, 75-bp read Illumina HiSeq 2500 (Illumina Inc.) at the Yale Center for Genome Analysis.

## 2.8 Genome alignment and variant calling

Quality control of reads was conducted using Cutadapt v1.8.3 with a minimum quality score of 20 and a minimum length of 30bp (Martin 2011). PCR primer sites were not removed, and single nucleotide polymorphisms (SNPs) were rarely observed in primer sites. Sequences were mapped to the consensus sequence of the ancestral virus used to found the evolution experiment using BWA v0.7.10 (Li and Durbin 2009). Consensus sequences and minority variant tables were generated using QUASR v7.01 (Watson et al. 2013) a minimum variant frequency cutoff of 0.05. Additional indels were called using VarScan v2.3.9 (Koboldt et al. 2012). Depth-of-coverage plots were generated using deepTools (Ramirez et al. 2016). For the 3'UTR region, coverage for passaged samples was normalized to the coverage of the ancestral stock as a check for insertions and deletions (indels). Coverage for each passaged sample was multiplied by a coefficient to standardize average coverage across samples, and the log2 ratio of the passaged sample coverage to the ancestral coverage was plotted across the region with a bin size of 5. Coverage was visualized using Integrative Genomics Viewer v2.3.61 (Robinson et al. 2011; Thorvaldsdóttir, Robinson, and Mesirov 2013).

## 2.9 Generation of recombinant viruses with Vero or Aag2-associated SNPs

Four additional recombinant viruses were generated: SL07 with 2 Vero-associated mutations (nsP1: R171Q and C: A76A, nt G7794T),  $\Delta$ DR1/2 with 2 Vero-associated mutations (nsP1: R171Q and C: A76A, nt G7794T), SL07 with 3 Aag2-associated mutations (nsP1: V292V, nt G913A; nsP1: M314L; capsid: G7779A), and  $\Delta$ DR1/2 with 3 Aag2-associated mutations (nsP1: V292V, nt G913A; nsP1: M314L; capsid: G7779A). Mutations were introduced by site-directed mutagenesis and confirmed by Sanger sequencing. In addition, each IC was sequenced to confirm the absence of unintended second-site mutations. Virus was generated from IC as described previously, except that two stocks were generated for the SL07 virus with Aag2-associated mutations: one from BHK cells and one from C6/36 cells. The presence of the engineered mutations was confirmed in all virus stocks by Sanger sequencing.

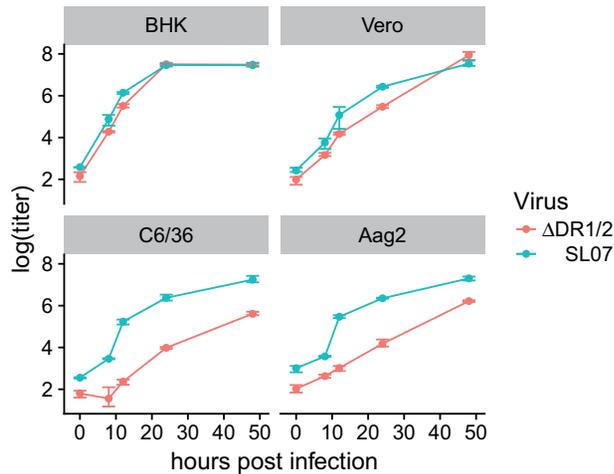
## 2.10 Statistical analyses

Statistical analyses were carried out in R v3.3.1 (R Development Core Team 2014). Figures were created using ggplot2 and cowplot packages in R (Wickham 2009; Wilke 2016). Mixed linear models were created using the nmlme package (Pinheiro et al. 2016) and model contrasts were compared using the package multcomp (Hothorn, Bretz, and Westfall 2008).

## 3. Results

### 3.1 3'UTR deletion affects viral replication on mosquito cells

To assess the effect of the 3'UTR deletion on viral replication, we compared 48-h growth curves for  $\Delta$ DR1/2 and the wild-type SL07 on two types of mammalian cell lines (hamster kidney-derived BHK and primate-derived Vero) and two types of mosquito cell lines (*A. albopictus*-derived C6/36 and *A. aegypti*-derived Aag2) (Fig. 2).  $\Delta$ DR1/2 had impaired replication relative to SL07 on both mosquito cells lines (pairwise t-test with Bonferroni correction,  $P < 0.01$  at all time points). The deletion



**Figure 2.** 3'UTR deletion results in reduced viral replication on mosquito cells. Growth curves were measured in triplicate for SL07 and  $\Delta$ DR1/2 on mammalian (BHK and Vero) and mosquito (C6/36 and Aag2) cell lines over 48 h. Infections were initiated at MOI  $\sim$  0.05. The initial time point (0 hpi) was sampled after the addition of media.  $\Delta$ DR1/2 showed impaired replication relative to SL07 on both mosquito cell lines (pairwise t-test with Bonferroni correction,  $P < 0.01$  at all time points).

had a lesser effect on replication on mammalian cells. On each of the two mammalian cell lines, one time point showed a significant difference in titer between the two viruses (pairwise t-test with Bonferroni correction, Vero 24 h post-infection (hpi)  $P < 0.01$ , BHK 12 hpi  $P < 0.01$ ); at all other time points there was no statistical significance between replication of the two viruses on mammalian cells ( $P > 0.05$ ). Thus, the deletion of two RSEs in the 3'UTR (DR1 and DR2) negatively impacted replication on mosquito cell lines, but minimally affected virus replication on mammalian cell lines. This finding was consistent with previous observations that 3'UTR deletions are detrimental to replication on mosquito cells but not to replication on mammalian cells (Kuhn, Hong, and Strauss 1990; Weaver et al. 1999; Chen et al. 2013; Trobaugh et al. 2014; Villordo et al. 2015).

### 3.2 Passage history influences evolved host range

Because the 3'UTR deletion reduced viral replication on mosquito cells, we expected that experimental passage on a mosquito cell line would result in compensatory evolutionary changes that would restore viral fitness. In contrast, we expected that passage on a mammalian cell line would not impose similar pressure for compensatory evolution. To examine the effect of passage history on viral fitness, passaged viruses were competed against a marked version of the ancestral  $\Delta$ DR1/2 (SacII- $\Delta$ DR1/2) on Vero and Aag2 cells. The introduced SacII site did not affect the replication of the virus (Supplementary Fig. S1). Virus populations from four groups were compared: viruses passaged strictly on Vero cells for twenty-five passages, viruses passaged strictly on Aag2 cells for seven passages, and viruses passaged alternately between Vero and Aag2 and had most recently grown on Vero (passage 24) or Aag2 (passage 25). Thus, virus populations from the alternating treatment appeared twice in this experiment, resulting in a total of ninety-six virus populations being assayed for fitness (twenty-four from each group). Fitness ( $\ln W$ ) was measured as the natural log of the output/input ratio of the passaged/marked virus. A fitness value of  $>0$  indicated that the passaged virus was more fit in the assay environment than the marked ancestor.

For nearly every virus population, passaged viruses had higher fitness than their ancestor on both cell lines ( $\ln W > 0$ ) (Fig. 3A). In many of the competition assays, the passaged virus outcompeted the competitor to such an extent that only an uncut amplicon band could be detected on the gel (Fig. 3B). In these cases, the  $\ln W$  reported is a low bound based on the sensitivity of our assay. Virus populations passaged on Aag2 cells universally increased to the assay saturation point on both hosts, showing improvement relative to their ancestor in their ability to grow on both Vero and Aag2 cells. Thus, passage on Aag2 cells improved fitness on the selected host and led to a correlated increase in fitness on the bypassed Vero cells.

Virus populations passaged in the alternating treatment similarly improved on both hosts. Populations passaged in the alternating treatment performed better on Aag2 cells when they had most recently been passaged on Aag2 cells (passage 25 samples) than when they had been most recently passaged on Vero cells (passage 24 samples). Most samples from the alternating passage 25 group reached the assay saturation (cut bands undetectable), while cut bands were visible for most samples from the alternating passage 24 group. Because we could only calculate a low bound for fitness in competitions that hit the assay saturation point, we could not say whether this difference in fitness is statistically significant.

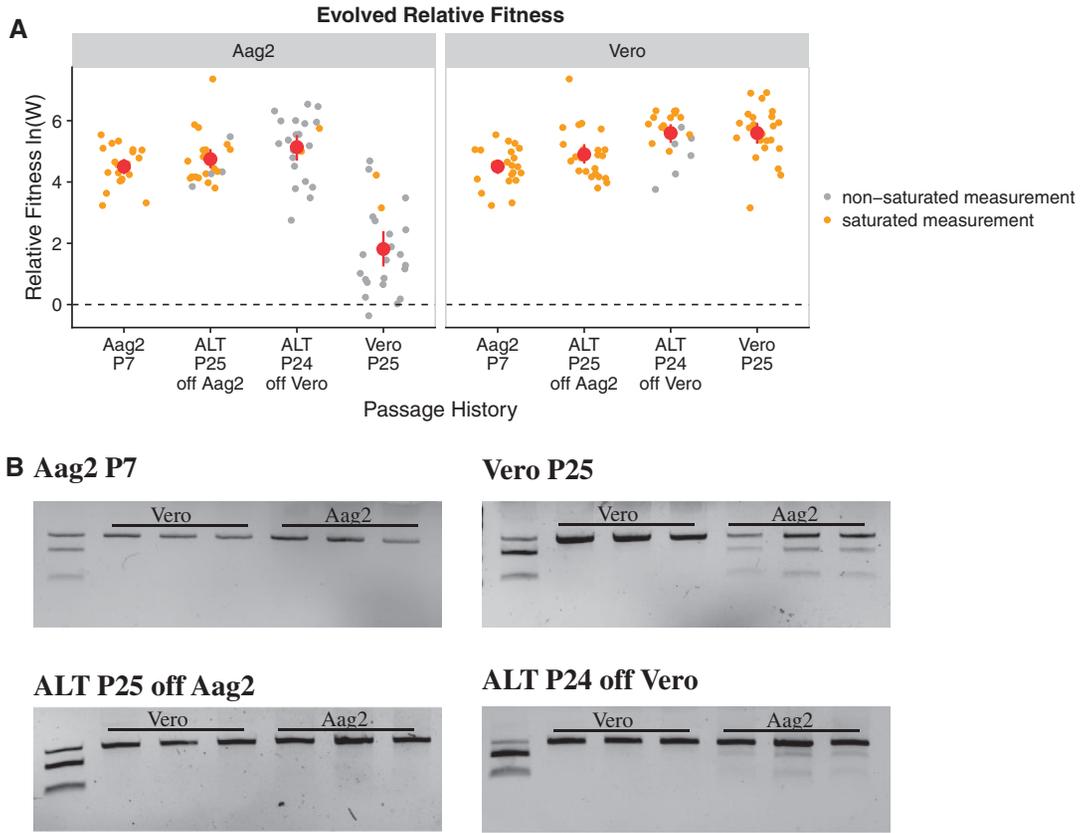
Virus populations passaged strictly on Vero cells showed improved fitness relative to the ancestor on both hosts. However, virus populations passaged on Vero had significantly lower evolved fitness relative to evolved populations from the other treatments (mixed linear model,  $P < 0.0001$ ). Therefore, fitness gains in this treatment were primarily specific to Vero cells.

In summary, passaged virus populations generally showed improved fitness on both cell lines. Passage on Aag2 cells resulted in high correlated fitness gains on both cell lines, while passage on Vero cells resulted in high fitness gains on Vero cells but only modest gains on Aag2 cells. Alternating passage resulted in high fitness gains on both cell lines.

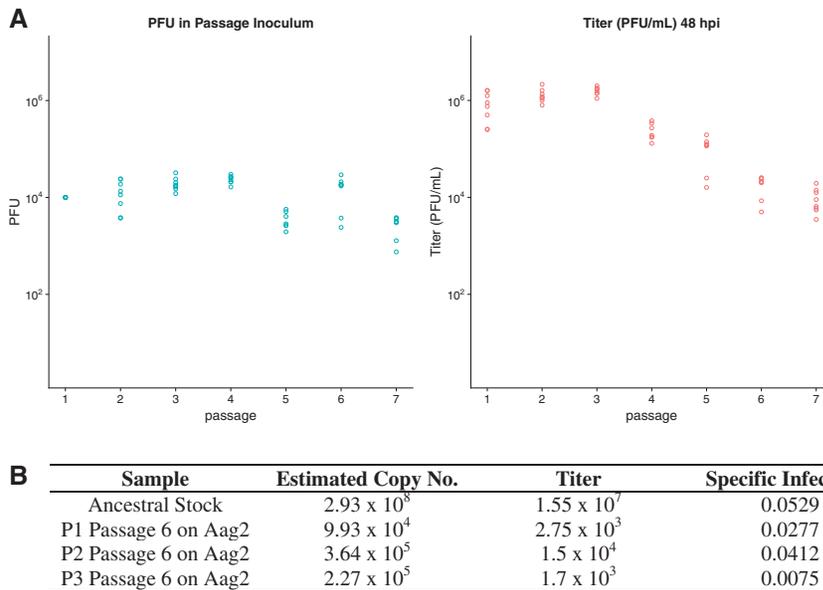
### 3.3 Reduction in detection by plaque assay for Aag2-passaged viruses

Results from fitness assays of Aag2-passaged populations left us with two seemingly contradictory observations. First, we were unable to maintain virus populations when we passaged them on Aag2 cells. This conclusion was based on data from plaque assays conducted on BHK cells, where we observed declining viral titers with Aag2 passage (Fig. 4A). However, our fitness assay showed that virus populations passaged on Aag2 cells had increased fitness relative to their ancestor (Fig. 3). This seemed inconsistent with our observation of declining virus titers.

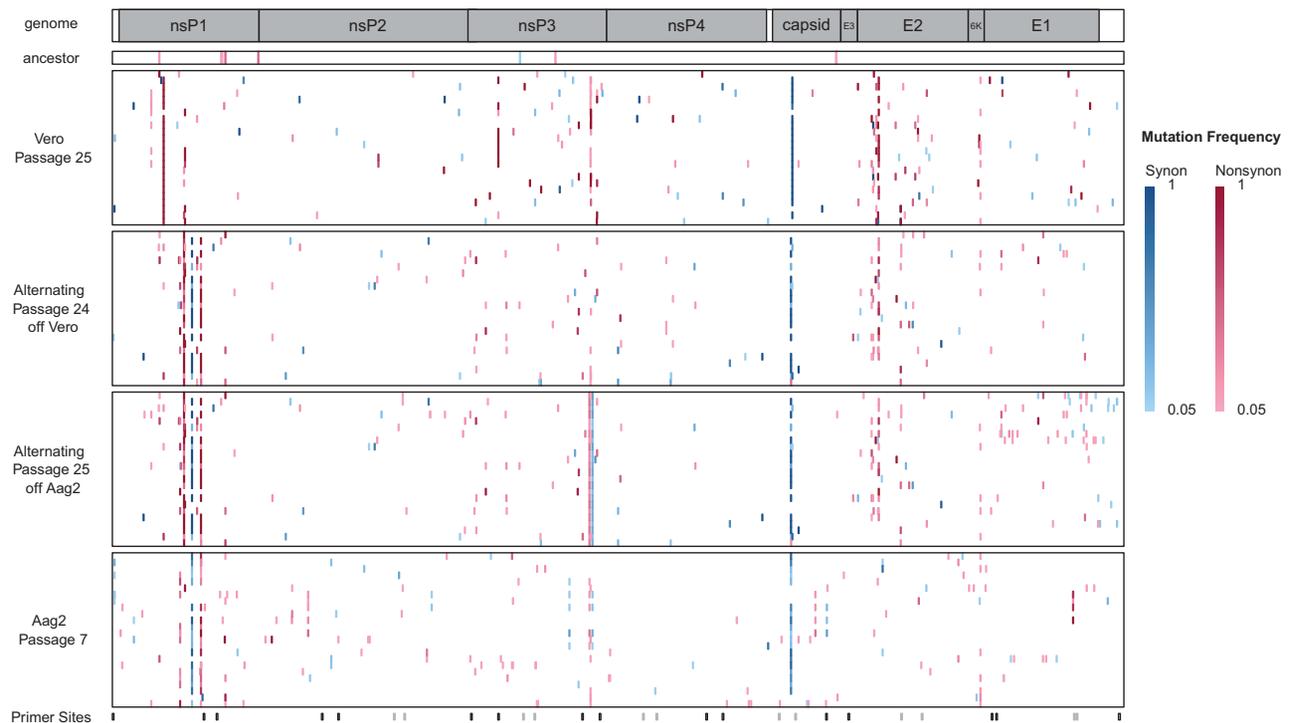
One possibility was that virus populations were not actually declining during passage on Aag2 cells, but that the viruses were losing their ability to form plaques on BHK cells. Thus we would have detected a decline with our plaque assays, but we could have seen increased fitness in our nucleic acid-based fitness assay. To test this possibility, we compared samples of three virus populations that had been frozen after six passages on Aag2 cells with the ancestral virus stock ( $\Delta$ DR1/2) that had been used to initiate these populations. We measured the titer of each virus sample by plaque assay on BHK cells, and we estimated the genome copy number with real-time PCR. We used these data to estimate the specific infectivity for each sample, or the ratio of plaques to virus genome copy number (Fig. 4B). We found that the three Aag2-passaged samples all had lower



**Figure 3.** Fitness of passaged virus populations on Aag2 and Vero cells. (A) Fitness of passaged virus populations relative to the *SacII*- $\Delta$ DR1/2 competitor is shown with mean and 95% CI marked in red.  $\ln(W)$  values above zero indicate an increase in fitness relative to the ancestor. Values colored in orange are saturated measurements, and indicate a low bound based on the sensitivity of the fitness assay. (B) Representative examples of fitness assay gels from each treatment. In each lane, the top band is from the passaged virus (uncut by *SacII*), and the bottom two bands are from the *SacII*- $\Delta$ DR1/2 competitor (cut by *SacII*). The first lane in each gel shows the input mixture at the start of the competition, followed by the output at 48 hpi from three competitions on Vero cells and three competitions on Aag2 cells. Lanes where only the uncut band is visible are saturated measurements.



**Figure 4.** Declining virus titers during Aag2 passage. (A) Declining virus titers for eight populations passaged on Aag2 cells. The right panel shows measured titers from plaque assays on BHK cells 48 hpi. The left panel shows estimated inoculum size for each passage. (B) Estimates of specific infectivity for the ancestral virus stock used to initiate experimental passage and three populations (P1, P2, and P3) after six passages on Aag2 cells. Copy number was estimated by quantitative PCR, and titer is from plaque assays on BHK cells.



**Figure 5.** Molecular substitutions in evolved populations. Location of molecular substitutions for all passaged virus populations relative to the ancestral clone. The top bar represents the CHIKV genome, with coding regions for each gene shaded. Each large block represents an experimental treatment, and each block is composed of stacked rows representing the 24 virus lineages in each treatment. Synonymous mutations are marked in blue, and nonsynonymous mutations are marked in red, with darker colors indicating higher frequency mutant alleles. Minority variants detected in the ancestral virus stock are also shown. Primer sites used for PCR amplification are indicated at the bottom. Standard primer sites are marked in black, and additional primer sites used for difficult to amplify samples are marked in gray.

specific infectivity than the ancestral stock; however, this trend was small compared to the substantial drop in viral titer. Taken together, this suggested the modest decrease in specific infectivity on BHK cells was insufficient to fully explain the decay in viral titers.

An alternative explanation for the drop in viral titers could be that virus infectivity was lower on Aag2 cells than on BHK or Vero cells. That is, for a given number of virus particles, fewer viruses actually initiated infection on Aag2 cells than on BHK or Vero cells. Because the calculation of input virus was based on titrations on BHK cells, the number of viruses used to initiate each passage on Aag2 cells could have been effectively much smaller than intended. This inaccuracy could have caused the loss (extinction) of virus populations via dilution effects. However, we note that the bottleneck sizes in our serial-transfer experiments may have been smaller than intended, but they were still sufficiently large for selection to overpower the effects of genetic drift, and for the virus populations to increase in relative fitness before being lost.

### 3.4 Sequencing reveals treatment-specific molecular changes

To examine molecular changes in passaged populations relative to the ancestor, we conducted whole genome next-generation sequencing on the ninety-six viral samples described above (Vero passage 25, alternating passage 24, alternating passage 25, and Aag2 passage 7) and the ancestral virus stock. Mutations were identified at 387 unique loci across the genome (Fig. 5). We observed a high degree of parallel evolution among replicates.

Of the 387 loci where molecular changes were observed, 140 were observed in multiple populations.

Because the ancestor in our evolution experiment had a large deletion in the 3'UTR that affected fitness, we expected to see compensatory genetic changes in the 3'UTR of evolved populations. Surprisingly, we observed very few molecular changes in the 3'UTR of the evolved populations. We observed no fixed (frequency = 1) substitutions or indels in the 3'UTR. We observed polymorphisms in the 3'UTR of eight populations (one from Aag2 passage 7, five from alternating passage 25, and two from Vero passage 25). These polymorphisms all consisted of SNPs, with the exception of one single nucleotide insertion. All mutations recorded in the 3'UTR were at frequencies less than 25 per cent. Insertions and deletions larger than the sequencing read length can be challenging to detect in next generation sequencing data; however, all PCR amplicons used for sequencing were visualized on agarose gels before submission, and no changes in size were observed that would suggest large indels. Additionally, we visually inspected coverage plots from the sequencing data, because increases in read coverage can suggest a duplication. However, coverage plots gave no indication of duplications (Supplementary Figs S2 and S3).

Several molecular substitutions in coding regions of the genome were highly parallel across replicate populations, and some of these mutations were associated with specific hosts. None of these highly parallel variants were detected in the ancestral virus stock, which suggests the mutations arose de novo. In the virus populations passaged strictly on Vero cells, the most common mutations were a nonsynonymous substitution in nsP1 (nsP1: R171Q) and a synonymous substitution in the capsid gene (C: A76A, nt G7794T). The nsP1: R171Q

substitution was observed in twenty-two of twenty-four virus populations passaged strictly on Vero, and it was fixed in all populations where it was observed. The synonymous capsid substitution was observed in twenty of twenty-four virus populations passaged strictly on Vero, and it was fixed in nineteen populations. All virus populations containing the capsid mutation also have the nsP1: R171Q substitution. Interestingly, these mutations were much less commonly observed in virus populations passaged in the alternating treatment. Three populations from the alternating treatment contained both the nsP1: R171Q and capsid mutation as polymorphisms. The frequencies of these two mutations in the populations were consistent with these mutations co-occurring on the same genetic background. These mutations never were observed in strictly Aag2-passaged populations.

Contrastingly, the molecular substitutions most common in the strictly Aag2-passaged cells were also commonly observed in populations passaged in the alternating treatment. Three mutations in the nsP1 coding region and one mutation in the capsid region were especially common. In the nsP1 protein ten strictly Aag2-passaged populations had an amino acid substitution at residue 234 (nsP1: K234T), eighteen populations had a synonymous nucleotide substitution at nt 913 (nsP1: V279V, nt G913A), and eighteen populations had an amino acid substitution at residue 314 (nsP1: M314L). These mutations were observed in 4, 16, and 16 of the alternating populations, respectively. Additionally, twenty strictly Aag2-passaged populations had a synonymous nucleotide substitution in the region encoding the capsid (C: K71K, nt 7779), and sixteen populations from the alternating treatment had this allele. All populations that had the nsP1:V279V mutation also had the nsP1:M314L mutation, and this set overlaps with the populations containing the C: K71K mutation with the exception of three populations.

From these data, we concluded that passaged virus populations evolved increased fitness through mutations in coding regions of the genome, rather than through mutations in the 3'UTR. Mutations that were highly parallel across replicates were identified as candidate mutations underlying fitness increases in passaged virus populations.

### 3.5 Key mutations were beneficial in both $\Delta$ DR1/2 and SL07 backgrounds

Our sequencing identified several candidate mutations in nsP1 and the capsid gene that we hypothesized may have improved fitness in passaged virus populations (Fig. 5). These mutations may have been selected because they compensated for the deletion in the 3'UTR, or they may have been beneficial mutations unrelated to the 3'UTR deletion. To test whether these mutations conferred benefits specifically in the genetic context of the 3'UTR deletion, we engineered sets of candidate mutations into both the wild type (SL07) and deletion ( $\Delta$ DR1/2) genomes and measured fitness. We chose two sets of mutations to test: a set of two mutations commonly observed together after passage on Vero cells (nsP1: R171Q and C: A76A, nt G7794T) and a set of three mutations commonly observed together after passage on Aag2 cells (nsP1: V292V, nt G913A; nsP1: M314L; capsid: G7779A). In total we created four recombinant viruses: SL07 with Vero-associated mutations,  $\Delta$ DR1/2 with Vero-associated mutations, SL07 with Aag2-associated mutations, and  $\Delta$ DR1/2 with Aag2-associated mutations.

We conducted fitness assays with these engineered viruses on Vero cells and on Aag2 cells. Each virus was competed against a Sac-II marked competitor with a like genetic

background (SL07 with SacII-SL07, and  $\Delta$ DR1/2 with SacII- $\Delta$ DR1/2). As expected, we found that Vero-associated mutations were beneficial on Vero cells, and Aag2-associated mutations were beneficial on Aag2 cells. However, the benefits conferred by these mutations were not specific to the  $\Delta$ DR1/2 background (Fig. 6). Both sets of mutations were beneficial in both genetic backgrounds. On average, the gain in fitness was greater on the wild type background in both cases. These results suggest that the candidate mutations we identified evolved in response to experimental passage on different host cells, but were not a response to the 3'UTR deletion.

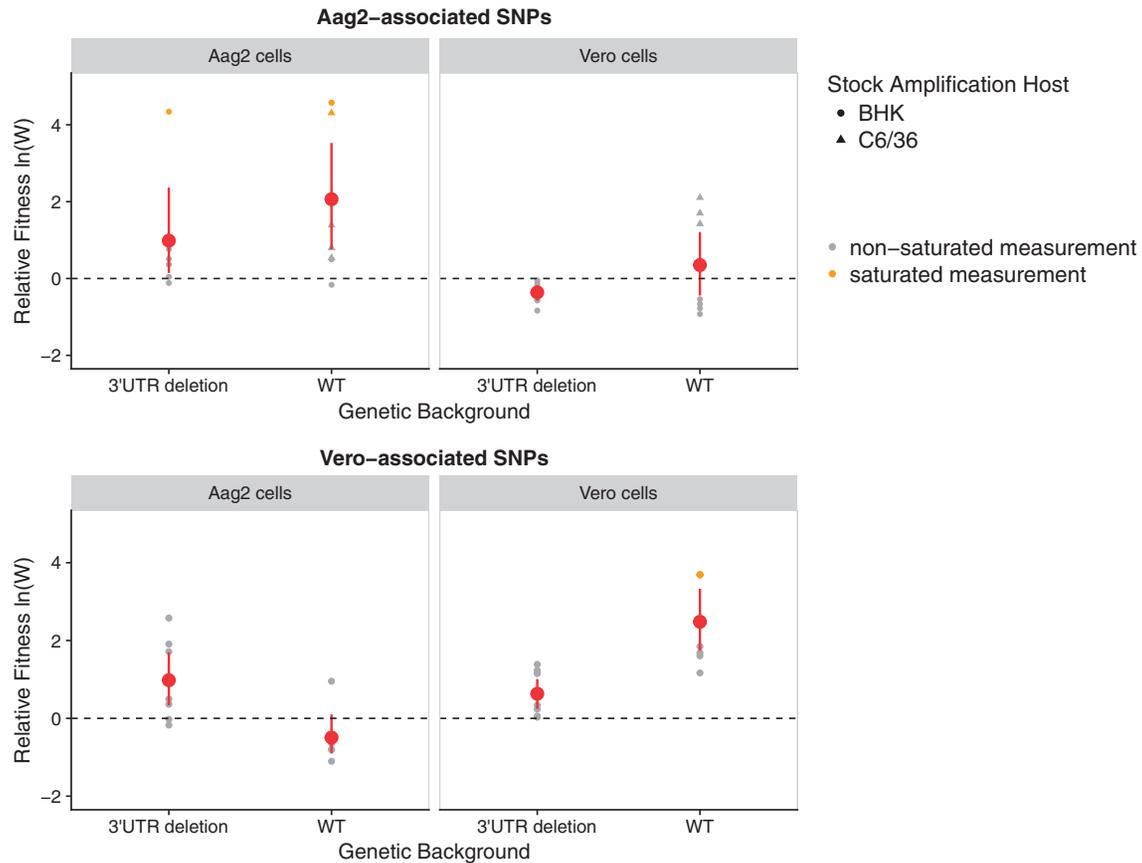
Additionally, to test whether the host cell line used to generate the virus affected viral fitness, we generated stocks of the SL07 virus with Aag2-associated mutations on both mammalian (BHK) and mosquito (C6/36) cell lines. In fitness assays, we observed higher fitness on Vero cells for virus derived from mosquito cells (Fig. 6).

These experiments confirmed that candidate mutations we identified in the nsP1 and capsid genes likely contributed to improved fitness in passaged virus populations. These mutations appear to be beneficial in the experimental environment but are not specific to the 3'UTR deletion.

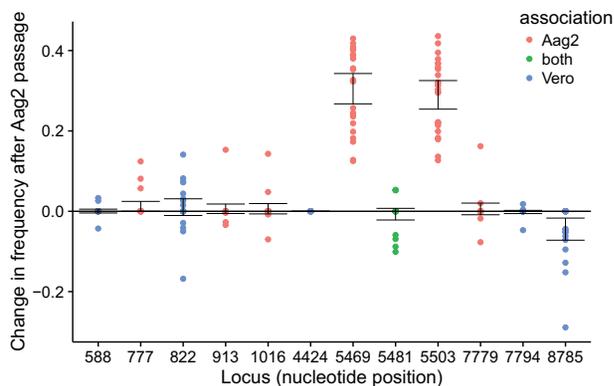
### 3.6 Enrichment of host-specific genotypes in alternating passages

When populations experience fluctuating environments, such as in our alternating treatment, the selection coefficients of alleles may differ between environments. This can cause the enrichment of environment-specific genotypes after passage in a given environment, and a subsequent decline in frequency for that genotype when the environment changes. Whether a given genotype fluctuates with environmental conditions depends on the pleiotropic effects of an allele in the different environments.

Our initial genetic analysis of populations passaged in the alternating treatment suggested that differing selective pressures in the Vero and Aag2 environments might be causing the enrichment of different genotypes after passage on different hosts. To further investigate this possibility, we focused on a subset of the most highly convergent mutations in the evolved populations. We considered mutations that were observed at least ten times in any one treatment, resulting in a set of twelve mutations (Fig. 7). Eleven of these twelve mutations were highly host specific and occurred in only one of the single-host treatments (i.e. a substitution at nt 7794 was observed in 20 populations in the Vero treatment, but never in the Aag2 treatment). With the exception of a mutation at nt 4424, all of the mutations in this set also were observed in the alternating treatment. For these mutations, we calculated changes in frequency in each population between passage 24 (harvested after replication on Vero) and passage 25 (harvested after replication on Aag2). For 8/11 of the alleles in question, changes in frequency were not significantly different from zero (t-test with Bonferroni correction,  $P > 0.05$ ). However, we detected two alleles that were significantly enriched after passage on Aag2 cells: nt 5469, which causes the amino acid substitution nsP3: P465Q (t-test with Bonferroni correction,  $P = 2.12e-12$ ) and the synonymous change nt C5503A (t-test with Bonferroni correction,  $P = 7.96e-13$ ). Both of these alleles were solely Aag2-associated in the single-host treatments. Additionally, one allele showed significant decrease in frequency after passage on Aag2 cells: nt 8785 that causes the amino acid substitution E2: G82R (t-test with Bonferroni correction,  $P = 8.34e-02$ ). In single-host treatments



**Figure 6.** Fitness effects of key mutations in deletion and wild type genetic backgrounds. Mutations associated with passage on Vero cell or Aag2 cells were engineered into two genetic backgrounds: CHIKV wild type (SL07) and the 3'UTR deletion virus ( $\Delta$ DR1/2). Fitness was measured through competition assays, in which each engineered virus was competed against a *SacII*-marked competitor with the same genetic background (SL07 with *SacII*-SL07,  $\Delta$ DR1/2 with *SacII*- $\Delta$ DR1/2). Fitness was measured on two cell types: Aag2 and Vero. For the SL07 virus with Aag2-associated mutations, virus stocks were generated on both BHK and C6/36 cells.



**Figure 7.** Change in allele frequencies in alternating treatment after passage on Aag2. For a subset of highly convergent alleles, here we show the change in frequency in the alternating treatment between passage 24 (harvested after replication on Vero) and passage 25 (harvested after replication on Aag2). Loci are colored by where the mutant allele was observed in the single-host treatments: only in the Vero treatment, only in the Aag2 treatment, or in both treatments.

this allele was only observed in Vero populations. Together these results suggest that many alleles we identified as being favored on a single host were at least tolerated on the alternate host. However, we have identified three alleles that seemed to experience fluctuating selection, perhaps due to tradeoffs between the two host types.

## 4. Discussion

The deletion mutant we constructed ( $\Delta$ DR1/2) had impaired replication on mosquito cells relative to the wild-type virus; therefore, we expected to see compensatory evolution and perhaps duplications in the 3'UTR after serial passage on mosquito cells. As we predicted,  $\Delta$ DR1/2 virus populations passaged on Aag2 cells or with alternating mosquito/mammalian passages improved their ability to replicate on mosquito cells. Surprisingly however, sequencing revealed very few molecular changes in the 3'UTRs of passaged populations. Instead, virus populations improved their fitness and their ability to replicate on mosquito cells through mutations in coding regions of the genome while retaining the 3'UTR deletion.

Sequencing of viral populations revealed a large number of mutations relative to the ancestral  $\Delta$ DR1/2 clone. Across all populations, we identified mutations at 387 unique loci. Notably, a previous study that experimentally passaged CHIKV in similar treatments (but different mosquito and mammalian cell lines) for seven passages observed no changes in the consensus sequence of the passaged viruses (Coffey and Vignuzzi 2011). Mutations in the nsP1 gene and the capsid protein were the most highly convergent molecular changes, and passage on mosquito versus mammalian cells selected different mutations in these genes. Reverse genetic experiments confirmed that these mutations improved fitness on their selected hosts. However, these mutations were beneficial in both the  $\Delta$ DR1/2 and the wild type SL07 3'UTR genetic backgrounds, which

suggested that these beneficial mutations were not specific to the 3'UTR deletion.

The function of the beneficial mutations observed in the nsP1 and capsid genes is unknown, but two of the nsP1 amino acid substitutions we identified have been described in naturally occurring CHIKV isolates. The nsP1 protein is the main enzyme involved in mRNA capping. It is a methyltransferase and a guanylyltransferase that also anchors the viral replicase complex to lipid membranes (Kuhn 2013). The nsp1: R171Q substitution has been reported in CHIKV isolates from Sri Lanka in 2006 (Lim et al. 2009), Comoros in 2005 (Kariuki Njenga et al. 2008; Wasonga et al. 2015), India in 2008–2013 (Abraham et al. 2016), and the Caribbean in 2014 (Sahadeo et al. 2017). The isolates from Sri Lanka, Comoros, and India belonged to the ESCA (East/Central/South African) genotype, as did our wild type virus SL07. The Caribbean isolates belonged to the Asian genotype. These studies reported co-circulating CHIKV variants nsP1: 171R and nsP1: 171Q, sometimes within the same human patient (Kariuki Njenga et al. 2008; Lim et al. 2009; Wasonga et al. 2015; Abraham et al. 2016). The function of the nsP1: R171Q mutation has not been described, although nsP1: 171Q variants were shown to have increased cytopathogenicity on primate cells (Abraham et al. 2016). None of the nsP1: 171Q isolates reported contained the synonymous capsid mutation we observed in our experimental populations or unusual 3'UTR structure. Additionally, the nsP1: M314L variant we observed after passage on mosquito cells has been reported in natural isolates from Mauritius (Kowalzik et al. 2008) and India in 2006 (Santhosh et al. 2008; Kumar et al. 2014). Again, the function remains unknown, and in natural isolates it did not co-occur with the capsid mutation we observed in our samples.

The mutations in the nsP1 and capsid genes explain part of the observed increase in fitness, but understanding the full picture likely involves the many other mutations we observed. We observed molecular changes at hundreds of loci across the genome, and each passaged population carried multiple mutations relative to the  $\Delta$ DR1/2 ancestor. Many of these mutations occurred in parallel across replicate populations, suggesting that they may be adaptive. Polymorphic loci were also common in passaged virus populations. Two important principles of RNA virus evolution suggest that the sum of these mutations may be important: epistasis and quasispecies.

Epistasis occurs when the fitness effect of a mutation depends on the genetic background in which it occurs. This can mean that the combined fitness effect of co-occurring mutations is greater than the sum of their parts, or that certain mutations are only beneficial in the context of certain genetic backgrounds. Epistatic interactions have been shown to be important for CHIKV evolution in nature (Tsetsarkin et al. 2011; Tsetsarkin et al. 2014) and for the evolution of related alphaviruses during experimental passage (Morley, Mendiola, and Turner 2015). The observed fitness of our virus populations likely depends on the combined fitness effects of multiple mutations that accumulated during experimental passage. The combined fitness effects of these mutations may be additive but are likely complicated by epistasis.

Additionally, the observed fitness of an RNA virus population depends on the genetic diversity within the population. The error-prone replication of RNA viruses results in a high mutation rate, which means that genetic diversity is generated quickly as viruses replicate. These results in virus populations that are not composed of a single genotype, but rather a mutant swarm of related genotypes, sometimes called a viral quasispecies. Experimental work has shown that the observed fitness of

a viral population depends on the sum of the quasispecies rather than on the isolated fitness of a single dominant genotype (Bordería et al. 2015). Thus, the minority variants we described likely play an important role in determining the observed fitness of the viral population.

The fitness increases we observed in our passaged virus populations likely stem from the interaction of multiple mutations within genomes and multiple genotypes within virus populations. The small number of mutations we investigated through reverse genetics did not show specificity to the 3'UTR deletion, but there are many other mutations we detected that have potential to be important. Outside of the mutations investigated here in detail, there were also highly convergent mutations in the E2 gene, nsP3, and others in nsP1.

Our experiment also contributes to a substantial body of research in which arboviruses are experimentally passaged on alternating hosts or strictly on one host type (Weaver et al. 1999; Turner and Elena 2000; Cooper and Scott 2001; Greene et al. 2005; Coffey et al. 2008; Vasilakis et al. 2009; Coffey and Vignuzzi 2011; Deardorff et al. 2011). The aim of these experiments is to study fitness trade-offs between hosts. Three main generalizations can be taken away from this body of work: (1) Selection on one host type leads to increased fitness on that host type. (2) This often, but not always, is correlated with a reduction in fitness on the bypassed host. (3) Populations evolved in an alternating host regime gain fitness on both host types, often to the same or greater extent as populations serially passaged on a single host type. The results from the current experiment were generally in line with these expectations. When virus populations were passaged strictly on one host type, they evolved increased fitness on the selected host type. Virus populations from the alternating treatment evolved improved fitness on both host types. Interestingly, populations passaged strictly on Aag2 cells showed a correlated improvement on Vero cells, but this was not true to the same extent in the opposite direction.

In conclusion, we found that CHIKV populations recovered fitness after a large 3'UTR deletion through molecular changes in coding regions of the genome. The mutations that were selected were often highly parallel across replicates, and were specific to the host cell line used for experimental passage. The most convergent mutations were beneficial mutations that improved fitness in the experimental environment, but were not specific to the 3'UTR deletion.

## Data availability

Raw sequencing data are available through the SRA project accession number SRP134110.

## Supplementary data

Supplementary data are available at *Virus Evolution* online.

**Conflict of interest:** None declared.

## Acknowledgements

This work was supported by grant no. DEB-1051093 from the US National Science Foundation (NSF) and grant no. R01-AI091646–01 from the US National Institutes of Health (NIH). V.J.M. was supported by the U.S. National Science Foundation (NSF) Graduate Research Fellowship grant #DGE-1122492, and P.E.T by the NSF BEACON Center for the Study of Evolution in Action.

## References

- Abraham, R. et al. (2016) 'Correlation of Phylogenetic Clade Diversification and *in Vitro* Infectivity Differences among Cosmopolitan Genotype Strains of Chikungunya Virus', *Infection, Genetics, and Evolution*, 37: 174–84.
- Bavia, L. et al. (2016) 'A Glance at Subgenomic Flavivirus RNAs and microRNAs in Flavivirus Infections', *Virology Journal*, 13: 84.
- Bordería, A. V. et al. (2015) 'Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype', *PLoS Pathogens*, 11: e1004838.
- Chen, R. et al. (2013) 'Chikungunya Virus 3' Untranslated Region: Adaptation to Mosquitoes and a Population Bottleneck as Major Evolutionary Forces', *PLoS Pathogens*, 9: e1003591.
- Coffey, L. L., and Vignuzzi, M. (2011) 'Host Alternation of Chikungunya Virus Increases Fitness While Restricting Population Diversity and Adaptability to Novel Selective Pressures', *Journal of Virology*, 85: 1025–35.
- et al. (2008) 'Arbovirus Evolution *in Vivo* Is Constrained by Host Alternation', *Proceedings of the National Academy of Sciences*, 105: 6970–5.
- Cooper, L. A., and Scott, T. W. (2001) 'Differential Evolution of Eastern Equine Encephalitis Virus Populations in Response to Host Cell Type', *Genetics*, 175: 1403–12.
- Deardorff, E. R. et al. (2011) 'West Nile Virus Experimental Evolution *in Vivo* and the Trade-off Hypothesis', *PLoS Pathogens*, 7: e1002335.
- Garneau, N. L. et al. (2008) 'The 3' Untranslated Region of Sindbis Virus Represses Deadenylation of Viral Transcripts in Mosquito and Mammalian Cells', *Journal of Virology*, 82: 880–92.
- George, J., and Raju, R. (2000) 'Alphavirus RNA Genome Repair and Evolution: Molecular Characterization of Infectious Sindbis Virus Isolates Lacking a Known Conserved Motif at the 3' End of the Genome', *Journal of Virology*, 74: 9776–85.
- Greene, I. P. et al. (2005) 'Effect of Alternating Passage on Adaptation of Sindbis Virus to Vertebrate and Invertebrate Cells', *Journal of Virology*, 79: 14253–60.
- Griffin, D. (2013), 'Alphaviruses', in D. M. Knipe and P. M. Howley (eds.), *Fields Virology*, 6th edn, Vol. 1. Philadelphia, PA: Lippincott Williams & Wilkins.
- Gritsun, T. S., and Gould, E. A. (2006a) 'Direct Repeats in the 3' Untranslated Regions of Mosquito-Borne Flaviviruses: Possible Implications for Virus Transmission', *Journal of General Virology*, 87: 3297–305.
- , and — (2006b) 'The 3' Untranslated Regions of Kamiti River Virus and Cell Fusing Agent Virus Originated by Self-Duplication', *Journal of General Virology*, 87: 2615–9.
- Hothorn, T., Bretz, F., and Westfall, P. (2008) 'Simultaneous Inference in General Parametric Models', *Biometrical Journal*, 50: 346–63.
- Hyde, J. L. et al. (2015) 'The 5' and 3' ends of alphavirus RNAs – Non-Coding Is Not Non-Functional', *Virus Research*, 206: 99–107.
- Kariuki Njenga, M. et al. (2008) 'Tracking Epidemic Chikungunya Virus into the Indian Ocean from East Africa', *Journal of General Virology*, 89: 2754–60.
- Koboldt, D. et al. (2012) 'VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing', *Genome Research*, 22: 568–76.
- Kowalik, S. et al. (2008) 'Characterisation of a Chikungunya Virus from a German Patient Returning from Mauritius and Development of a Serological Test', *Medical Microbiology and Immunology*, 197: 381–6.
- Kuhn, R. J., Hong, Z., and Strauss, J. H. (1990) 'Mutagenesis of the 3' Nontranslated Region of Sindbis Virus RNA', *Journal of Virology*, 64: 1465–76.
- et al. (1992) 'Attenuation of Sindbis Virus Neurovirulence by Using Defined Mutations in Nontranslated Regions of the Genome RNA', *Journal of Virology*, 66: 7121–7.
- (2013), 'Togaviridae', in D. M. Knipe and P. M. Howley (eds.), *Fields Virology*, 6th edn, Vol. 1, pp. 629–50. Philadelphia, PA: Lippincott Williams & Wilkins.
- Kumar, A. et al. (2014) 'A Novel 2006 Indian Outbreak Strain of Chikungunya Virus Exhibits Different Pattern of Infection as Compared to Prototype Strain', *PLOS One*, 9: e85714.
- Lei, Y. et al. (2011) 'Functional Interaction between Cellular p100 and the Dengue Virus 3' UTR', *Journal of General Virology*, 92: 796–806.
- Li, H., and Durbin, R. (2009) 'Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform', *Bioinformatics*, 25: 1754–60.
- Lim, C. et al. (2009) 'Chikungunya Virus Isolated from a Returnee to Japan from Sri Lanka: Isolation of Two Sub-Strains with Different Characteristics', *The American Journal of Tropical Medicine and Hygiene*, 81: 865–8.
- Lo, M. K. et al. (2003) 'Functional Analysis of Mosquito-Borne Flavivirus Conserved Sequence Elements within 3' Untranslated Region of West Nile Virus by Use of a Reporting Replicon That Differentiates between Viral Translation and RNA Replication', *Journal of Virology*, 77: 10004–14.
- Mandl, C. W. et al. (1998) 'Spontaneous and Engineered Deletions in the 3' Noncoding Region of Tick-Borne Encephalitis Virus: Construction of Highly Attenuated Mutants of a Flavivirus', *Journal of Virology*, 72: 2132–40.
- Martin, M. (2011) 'Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads', *EMBnet Journal*, 17: 10–2.
- Men, R. et al. (1996) 'Dengue Type 4 Virus Mutants Containing Deletions in the 3' Noncoding Region of the RNA Genome: Analysis of Growth Restriction in Cell Culture and Altered Viremia Pattern and Immunogenicity in Rhesus Monkeys', *Journal of Virology*, 70: 3930–7.
- Morley, V. J., Mendiola, S. Y., and Turner, P. E. (2015) 'Rate of Novel Host Invasion Affects Adaptability of Evolving RNA Virus Lineages', *Proceedings of the Royal Society B: Biological Sciences*, 282: 20150801.
- Ou, J. H., Trent, D. W., and Strauss, J. H. (1982) 'The 3'-Non-Coding Regions of Alphavirus RNAs Contain Repeating Sequences', *Journal of Molecular Biology*, 156: 719–30.
- Pardigon, N., and Strauss, J. H. (1992) 'Cellular Proteins Bind to the 3' End of Sindbis Virus minus-Strand RNA', *Journal of Virology*, 66: 1007–15.
- , Lenches, E., and Strauss, J. H. (1993) 'Multiple Binding Sites for Cellular Proteins in the 3' End of Sindbis Alphavirus Minus-Sense RNA', *Journal of Virology*, 67: 5003–11.
- Pfeffer, M., Kinney, R. M., and Kaaden, O. R. (1998) 'The alphavirus 3'-Nontranslated Region: Size Heterogeneity and Arrangement of Repeated Sequence Elements', *Virology*, 240: 100–8.
- Pinheiro, J. et al. (2016), 'Nlme: Linear and Nonlinear Mixed Effects Models'. <https://CRAN.R-project.org/package=nlme>.
- R Development Core Team. (2014) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramirez, F. et al. (2016) 'deepTools2: A Next Generation Web Server for Deep-Sequencing Data Analysis', *Nucleic Acids Research*, 44: W160–5.

- Robinson, J. T. et al. (2011) 'Integrative Genomics Viewer', *Nature Biotechnology*, 29: 24–6.
- Sahadeo, N. S. D. et al. (2017) 'Understanding the Evolution and Spread of Chikungunya Virus in the Americas Using Complete Genome Sequences', *Virus Evolution*, 3: vex010–vex10.
- Santhosh, S. R. et al. (2008) 'Comparative Full Genome Analysis Revealed E1: A226V Shift in 2007 Indian Chikungunya Virus Isolates', *Virus Research*, 135: 36–41.
- Sokoloski, K. J. et al. (2010) 'Sindbis Virus Usurps the Cellular HuR Protein to Stabilize Its Transcripts and Promote Productive Infections in Mammalian and Mosquito Cells', *Cell Host & Microbe*, 8: 196–207.
- Stapleford, K. A. et al. (2016) 'Whole-Genome Sequencing Analysis from the Chikungunya Virus Caribbean Outbreak Reveals Novel Evolutionary Genomic Elements', *PLoS Neglected Tropical Diseases*, 10: e0004402.
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013) 'Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration', *Briefings in Bioinformatics*, 14: 178–92.
- Trobaugh, D. W. et al. (2014) 'RNA Viruses Can Hijack Vertebrate microRNAs to Suppress Innate Immunity', *Nature*, 506: 245–8.
- Tsetsarkin, K. A., and Weaver, S. C. (2011) 'Sequential Adaptive Mutations Enhance Efficient Vector Switching by Chikungunya Virus and Its Epidemic Emergence', *PLoS Pathogens*, 7: e1002412.
- et al. (2011) 'Chikungunya Virus Emergence Is Constrained in Asia by Lineage-Specific Adaptive Landscapes', *Proceedings of the National Academy of Sciences*, 108: 7872–7.
- et al. (2014) 'Multi-Peaked Adaptive Landscape for Chikungunya Virus Evolution Predicts Continued Fitness Optimization in *Aedes albopictus* Mosquitoes', *Nature Communications*, 5: 4084.
- Turner, P. E., and Elena, S. F. (2000) 'Cost of Host Radiation in an RNA Virus', *Genetics*, 156: 1465–70.
- Vasilakis, N. et al. (2009) 'Mosquitoes Put the Brake on Arbovirus Evolution: Experimental Evolution Reveals Slower Mutation Accumulation in Mosquito than Vertebrate Cells', *PLoS Pathogens*, 5: e1000467.
- Villordo, S. M. et al. (2015) 'Dengue Virus RNA Structure Specialization Facilitates Host Adaptation', *PLoS Pathogens*, 11: e1004604.
- Wasonga, C. et al. (2015) 'Genetic Divergence of Chikungunya Virus Plaque Variants from the Comoros Island (2005)', *Virus Genes*, 51: 323–8.
- Watson, S. J. et al. (2013) 'Viral Population Analysis and Minority-Variant Detection Using Short Read Next-Generation Sequencing', *Philosophical Transactions of the Royal Society B*, 368: 20120205.
- Weaver, S. C. et al. (1999) 'Genetic and Fitness Changes Accompanying Adaptation of an Arbovirus to Vertebrate and Invertebrate Cells', *Journal of Virology*, 73: 4316–26.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Wilke, C. O. (2016) *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*, 0.7.0 edn. <https://CRAN.R-project.org/package=cowplot>.