Virus Evolution, 2018, 4(1): vey001

doi: 10.1093/ve/vey001
Research article

Mosquito bottlenecks alter viral mutant swarm in a tissue and time-dependent manner with contraction and expansion of variant positions and diversity

Edward I. Patterson,[1,†,§] Kamil Khanipov,[2,†] Mark M. Rojas,[2] Tiffany F. Kautz,[1] Dedeke Rockx-Brouwer,[1,‡] Georgiy Golovko,[2] Levent Albayrak,[2] Yuriy Fofanov,[2] and Naomi L. Forrester[1,*]

[1]Department of Pathology, Institute for Human Infections and Immunity, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0610, USA and [2]Department of Pharmacology and Toxicology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX 77555-0617, USA

*Corresponding author: E-mail: naforres@utmb.edu
†These authors contributed equally to this work.
‡Present address: Molecular Biology and Immunology, Intravacc, Antonie van Leeuwenhoeklaan 9, 3720 AL Bilthoven, Netherlands.
§http://orcid.org/0000-0003-3465-0848

Abstract

Viral diversity is theorized to play a significant role during virus infections, particularly for arthropod-borne viruses (arboviruses) that must infect both vertebrate and invertebrate hosts. To determine how viral diversity influences mosquito infection and dissemination *Culex taeniopus* mosquitoes were infected with the Venezuelan equine encephalitis virus endemic strain 68U201. Bodies and legs/wings of the mosquitoes were collected individually and subjected to multi-parallel sequencing. Virus sequence diversity was calculated for each tissue. Greater diversity was seen in mosquitoes with successful dissemination versus those with no dissemination. Diversity across time revealed that bottlenecks influence diversity following dissemination to the legs/wings, but levels of diversity are restored by Day 12 post-dissemination. Specific minority variants were repeatedly identified across the mosquito cohort, some in nearly every tissue and time point, suggesting that certain variants are important in mosquito infection and dissemination. This study demonstrates that the interaction between the mosquito and the virus results in changes in diversity and the mutational spectrum and may be essential for successful transition of the bottlenecks associated with arbovirus infection.

Key words: virus diversity; arbovirus; minority variants; bottlenecks; quasispecies.

1. Introduction

RNA viruses are comprised of a cloud of related viruses known as a quasispecies, or viral swarm (Lauring and Andino 2010; Rozen-Gagnon et al. 2014). This phenomenon is a direct result of the low fidelity attributed to the RNA-dependent RNA polymerase (RdRp), which lacks an exonuclease proofreading function resulting in error-prone replication (Elena and Sanjuan 2005). This lack of fidelity in replication is thought to allow RNA viruses to adapt rapidly to changing environments, which is particularly relevant for moving between different hosts and for transmitting between host compartments, as different organs

© The Author(s) 2018. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

or tissues will have different mutational spectra (Wagner et al. 2009; Sim et al. 2015). The presence of this cloud of mutations may also assist in the pathology and transmission of the virus, and transmission between different groups of cells within a host. Arthropod-borne viruses (arboviruses) have the added burden of infecting two disparate species, so it is theorized that this cloud of mutations allows viruses to sustain a transmission cycle between insect and vertebrate hosts, and is essential for maintenance of the virus within the transmission cycle.

During their infection cycle, arboviruses infect a wide variety of cell types in the invertebrate and vertebrate hosts. This host switching is hypothesized to require changes in the mutational spectrum to allow effective infection of these different host tissues. For example, the purifying selection observed during mammalian infections (Coffey et al. 2008) followed by restoring diversity upon infection of a mosquito (Ciota et al. 2007, 2009; Brackney et al. 2011). Significant bottlenecks also occur during the arbovirus infection of the mosquito vector (Ciota et al. 2012; Forrester et al. 2012). Such bottlenecks have been identified when the virus infects the mosquito midgut; escapes from the midgut into the hemocoel; and infects the mosquito salivary glands. Each bottleneck results in a significant reduction of diversity in the quasispecies population, which recovers after infection in a new tissue is established, although the timeline for recovery and barrier-specific mutations have not been completely understood (Grubaugh et al. 2016).

Venezuelan equine encephalitis virus (VEEV) is a New World alphavirus that causes encephalitis upon infection of equids and humans. Generally, the virus circulates endemically between small rodents and mosquitoes in the forests of Central and South America (Weaver et al. 2004). Occasionally, the virus undergoes specific mutations in the envelope glycoprotein that allow the virus to successfully infect horses to high titers (Brault et al. 2004; Anischenko et al. 2006). As a consequence, mosquitoes that are more promiscuous in their bloodmeal preferences can be infected (Brault et al. 2002; Greene et al. 2005) resulting in increased human transmission. The conjunction of these two events, the emergence of mutations and infection of equids or epidemic mosquitoes, can lead to an epidemic outbreak of the virus.

An increasing number of studies report the importance of minority variants in the infecting virus population (Brackney et al. 2011; Acevedo, Brodsky, and Andino 2014; Borderia et al. 2015; Grubaugh et al. 2016). Previous work has shown the presence of severe bottlenecks during escape of VEEV from the midgut and infection of the salivary glands (Forrester et al. 2012); however, these studies detected marked clones, not measures of diversity. This study aimed to determine if these bottlenecks resulted in significant shifts in the viral population diversity. To investigate the effect of diversity, particularly minority variants, next generation sequencing (NGS) of infected mosquitoes' midguts, bodies, legs and wings (hereafter legs/wings), and saliva was performed to collect data on viral sequences. The NGS platform is highly sensitive and allows for the proportion of even low frequency mutations (~1% frequency) to be identified. Analysis of the viral genomes allowed for the determination of the diversity and trends of viral sequences at different times and in different compartments of the mosquito vector as the virus transitions from the midgut, through the hemocoel to the salivary glands. The goal of these studies was to identify the role that diversity plays in sustaining a successful disseminated mosquito infection. In particular, to determine if the presence of known bottlenecks (Forrester et al. 2012) showed any evidence of reducing diversity in the subsequent tissue, if so, was reduced diversity accompanied by an increase in diversity to compensate for the reduction, and to determine if there was any evidence of selection with regard to the minority variants that arose during infection.

2. Materials and methods

2.1 Ethics statement

This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol 1309038A was approved by the Institutional Animal Care and Use Committee of the University of Texas Medical Branch.

2.2 Cell culture and viruses

African green monkey kidney (Vero) and baby hamster kidney (BHK) cells were obtained from the American Type Culture Collection (Bethesda, MD) and maintained in Dulbecco's minimal essential medium (DMEM) (Gibco, Carlsbad, CA) supplemented with 5% fetal bovine serum (FBS) and penicillin/streptomycin (100 U/ml). Using BHK cells, viruses were rescued from an infectious cDNA clone derived from an enzootic strain of VEEV IE, 68U201 as described previously and without further passage (Smith et al. 2008). The parental 68U201 virus (genomic sequence in GenBank accession no. U34999.1) was isolated from a sentinel hamster in a sylvatic Guatemalan focus of VEEV in 1968 and passaged once in infant mice and twice in BHK cells prior to cloning (Scherer, Dickerman, and Ordonez 1970).

2.3 Experimental infections

To establish experimental transmission cycles, CD-1 mice were infected subcutaneously with 3 $\log_{10}$ PFU of 68U201 virus rescued from an infectious clone as described above. Mice ($n = 6$) were allowed to recover from the injection for 24 h. At 24 h post-infection, mice were anesthetized using a mixture of ketamine and xylene before being placed on the screened lids of 0.5 l cardboard cartons containing cohorts of *Culex taeniopus* mosquitoes (colony originating from Chiapas, Mexico). Mosquitoes were allowed to feed on the mice for up to 1 h. Mice were bled via the retro-orbital route after presentation to mosquitoes, and again on Day 3 post-infection. On Day 6 post-infection, the mice were sacrificed and brains harvested.

Cohorts of *C. taeniopus* mosquitoes were sugar-starved for at least 16 h; then allowed to feed for up to one hour on mice infected with 68U201 as described above. Following feeding, engorged mosquitoes were incubated at 27°C and provided 10% sucrose *ad libitum*. On Day 1, mosquitoes were chilled and the midguts dissected out and the remaining carcass held separately. Midguts were cut in half and residual blood was washed out using PBS. On Days 4, 8, 12, and 19, additional mosquitoes were immobilized by cold to remove the legs/wings, and individuals were allowed to salivate for 45 min into a capillary tube containing FBS. The remaining carcasses were used as representation of the midgut, which has been demonstrated to be an effective substitute for midguts (Forrester et al. 2012).

2.4 Processing of tissues

All mosquito tissues were placed into DMEM supplemented with 10% FBS, penicillin/streptomycin, and Fungizone (Sigma-Aldrich, St Louis, MO) and homogenized at 26 Hz for 5 min, then subjected to centrifugation at 3,820 × *g* for 10 min. All samples except saliva samples were tested for the presence of virus by a

Table 1. The number of samples sequenced from each day for both disseminated and non-disseminated samples.

	Disseminated					Non-disseminated				
	Day 1	Day 4	Day 8	Day 12	Day 19	Day 1	Day 4	Day 8	Day 12	Day 19
Midguts	7 (6)					5 (5)				
Bodies	6 (0)	6 (3)	16 (9)	10 (7)	23 (19)		5 (4)	2 (2)	3 (3)	1 (1)
Legs/Wings		6 (3)	15 (3)	10 (6)	23 (13)					
Saliva		6 (0)	15 (0)	10 (0)	22 (0)					

Samples in parentheses are the number included in the statistical analysis as they had greater than 100x coverage at more than 90% of the genome.

cytopathic effect (CPE) assay on Vero cells. Samples were counted as positive for virus if CPE was observed at 72 h post-infection (Li et al. 2013). Positive samples and their corresponding saliva samples were stored at −80°C for subsequent analysis. Saliva was not used for CPE assays because the entire sample was required for sequencing.

2.5 RNA extraction and sequencing

RNA was extracted from CPE positive samples, along with the saliva samples associated with CPE positive legs/wings. Samples were placed into Buffer AVL from the QIAamp Viral RNA Mini kit (Qiagen, Valencia, CA) and RNA was extracted as per the manufacturer's protocol.

2.5.1 Library construction

Viral RNA (0.05–1.7 μg) was quantified using a Qubit 2.0 fluorometer and was then fragmented by incubation at 94°C for 8 min in 19.5 μl of fragmentation buffer (Illumina, San Diego, CA). First and second strand synthesis, adapter ligation, and amplification of the library were performed using the Illumina TruSeq RNA Samplec Preparation kit as per the manufacturer's protocol.

2.5.2 Sequence determination

Cluster formation of the library DNA templates was performed using the TruSeq PE Cluster Kit v3 (Illumina) and the Illumina cBot workstation as per the manufacturer's protocol. Paired-end 50 base sequencing by synthesis was performed using TruSeq SBS kit v3 (Illumina) on an Illumina HiSeq 1500 as per the manufacturer's protocol.

2.6 Sequence assembly and analysis

2.6.1 Quality and filtration

The quality for each sample/dataset was assessed using FASTQC (Andrews 2015). The paired-end reads were merged for each sample and then filtered to exclude reads with unknown characters (anything other than A, T, C, G) and low quality (<10 quality score), so that only high-quality reads were used during analysis. Additionally, the first sixteen bases of each read were trimmed due to nucleotide bias.

2.6.2 Reference sequence

The analysis was performed using VEEV complete cds (GenBank accession No.: U34999.1; Oberste, Parker, and Smith 1996).

2.6.3 Variant analysis

To analyze the variant hotspots, regions of the genome with high levels of variation, of bodies, legs/wings, midgut, and saliva for each mosquito, each sample was run through a novel rare variant pipeline developed by the Fofanov laboratory (available upon request from Dr Forrester). The pipeline first maps each read to the reference VEEV genome with perfect match, then unmapped reads are re-mapped with one mismatch and added to the final map. The thirty-four base long reads used in the analyses were validated as viral sequences and not host sequences by analysis of the longest subsequences shared explicitly (no mismatches allowed) and longest similar (one mismatch allowed) between viral and host genomes. To avoid bias introduced by locations poorly covered by mapped reads, the diversity indices were then generated by body part, with a minimum per position nucleotide coverage threshold per base of 10, 100, and 1,000 (Supplementary Fig. S1). Diversity was calculated using Shannon entropy (Lin 1991). Positions with non-zero entropy were considered to be variant positions. Samples that did not have a minimum per position nucleotide coverage of 10 for >90% of the genome (reference sequence was 11,464 bp) were excluded. Once this threshold was reached no correlation was observed between the percentage of the genome covered and the number of positions that had variants present (Supplementary Fig. S2).

2.6.4 Statistics

All statistical analyses were performed with GraphPad Prism 6.0 (GraphPad Software, Inc.).

3. Results

3.1 Mosquito infections

After feeding cohorts of *C. taeniopus* mosquitoes on viremic CD-1 mice, seventy-eight of the mosquitoes were positive for virus infection by CPE. Mosquitoes sampled were from all six initial mouse infections, details of all the samples can be found in Supplementary Table S1. Of the seventy-eight mosquitoes, sixty-two had evidence of a disseminated infection, while the remaining sixteen failed to disseminate past the midgut. Of these, forty-four disseminated mosquitoes and fifteen non-disseminated mosquitoes had a sufficient number of viral reads to be included in further analysis (Table 1).

3.2 Viral diversity across minimum nucleotide coverage thresholds

To test for significant effects of bottlenecks on the diversity of the mosquitoes, the mutation spectrum for forty-four mosquitoes was determined for up to three tissues: bodies, legs/wings, and saliva. No difference was observed in diversity at minimum nucleotide coverage thresholds of 10, 100, and 1,000 (Supplementary Fig. S1) using the Mann-Whitney and Welch's *t*-test between the different compartments when sequences from mosquito samples on all days were pooled. However, to increase significance 100× coverage was used as a cut off for further analysis.

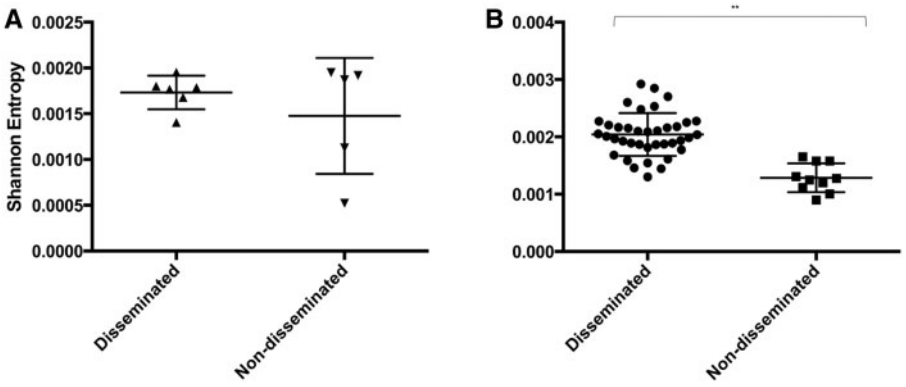

Figure 1. Amount of diversity is reduced in non-disseminated mosquitoes. ⊢ ⊤ Range of diversity was calculated for (A) Midguts with disseminated infections ($n = 6$) and non-disseminated infections ($n = 5$). (B) Bodies with disseminated infections ($n = 38$) and non-disseminated infections ($n = 10$). Statistics were calculated using Student's t-test with Welch correction. **$P < 0.0001$.

To prevent bias from insufficient coverage, sequences that had <90% of the genome covered at 100× sequencing coverage were excluded. This excluded all saliva sequences, as there were insufficient viral reads in all saliva samples. This is an inherent bias of the sampling method, as alphaviruses tend to deposit very few viral particles in the saliva (Smith et al. 2006). Details of all the samples sequenced and the amount of coverage can be found in Supplementary Tables S1 and S2 and the number of samples used in the final analysis can be found in Table 1.

Serum samples collected from the mice were also sequenced. Low coverage for these samples limited analysis to a consensus viral sequence, with no minority variants identified. The consensus sequence in the mice was identical to the plasmid sequence used to produce the virus stocks. The proportion of samples that was obtained from each mouse can be found in Supplementary Fig. S3.

3.3 Disseminated infections have higher diversity than non-disseminated infections

To determine if a lack of viral diversity correlates with non-disseminated infections in mosquitoes, both disseminated and non-disseminated midguts and disseminated and non-disseminated bodies were compared (Fig. 1). Midguts were determined to be non-disseminated if the CPE assay was only positive for the midgut and not the body for that mosquito. Bodies were determined to be non-disseminated if the CPE assay was only positive for the body and not the legs/wings for that mosquito. No significant difference was observed in midguts, which were processed on Day 1 (Fig. 1A). For the bodies, there is significantly less diversity for non-disseminated compared with the disseminated bodies ($P = 0.0070$) (Fig. 1B).

3.4 Variant frequency compared with diversity

To determine if the viral diversity was a result of a few mutations at high frequencies or large numbers of mutations at low frequencies, the number of positions along the genome that showed a variant was calculated. All samples that had <90% of the genome sequenced were excluded from these analyses to control for low coverage confounding the calculations of diversity and variant numbers. At <90% of the genome, the number of variants correlated with the percentage of genome covered. When >90% of the genome had a minimum sequencing coverage of 10, no correlation was observed (Supplementary Fig. S2), so all samples with fewer than 90% of the genome covered by sequencing at 10× coverage were discarded from the analysis. There was a significant drop in the number of variant positions between pooled bodies and pooled legs/wings across all days by Student's *t*-test ($P = 0.0043$) (Fig. 2A), and this was mirrored in the diversity results ($P = 0.0257$) (Fig. 2B) as determined by Welch's *t*-test.

3.5 Number of variant positions and diversity changes over the time of an infection

To determine if time was a factor in the generation of mutants and restoration of diversity, the samples were stratified by time. When the number of variant positions in the genome per sample was plotted on Days 1, 4, 8, 12, and 19 both bodies and legs/wings showed significant differences using Kruskal-Wallis test (Fig. 3A and B). Using *post hoc* Dunn's correction significant differences were detected between Days 8 and 12, and between Days 8 and 19 for the legs/wings (Fig. 3B), but no significant differences were observed between groups for the bodies. The pattern suggests that there is a reduction in the number of variant positions at Day 8 in the legs/wings compared with that observed in the bodies, which then expands dramatically by Day 12 and shows some reduction by Day 19, although this was not statistically significant. To determine if there was a relationship between the variation observed in bodies and legs/wings within the individual mosquitoes (i.e. whether the number of variant positions in the body correlated with the same effect in the legs/wings), the number of variant positions for bodies and legs/wings were plotted from the same mosquito (Fig. 3C). The ratio between bodies: legs/wings was calculated and shown in Supplementary Fig. S4. As in Fig. 2, there were significantly more variants in the bodies than legs/wings, and this was recapitulated for Day 8 ($n = 3$) ($P = 0.0258$), Day 12 ($n = 3$) ($P = 0.0315$), and Day 19 ($n = 13$) ($P = 0.0004$) as determined by Student's *t*-test. For Day 1 no legs/wings were sequenced, and on Day 4 only two mosquitoes had sufficient coverage in both bodies and legs/wings.

For comparison, the effect of time on the Shannon entropy measure of diversity was also calculated on Days 1, 4, 8, 12, and 19 for both bodies and legs/wings. For both the bodies and the legs/wings there was a significant difference between days by Kruskal-Wallis analysis (bodies $P = 0.0061$; legs/wings $P = 0.0049$) (Fig. 4A and B). Using the Dunn's *post hoc* correction significant differences were observed between Days 1 and 19 for the bodies (Fig. 4A), and Days 4 and 19 for the legs/wings (Fig. 4B), all other comparisons showed no significance. To determine if there was a relationship between the diversity as measured by Shannon entropy observed

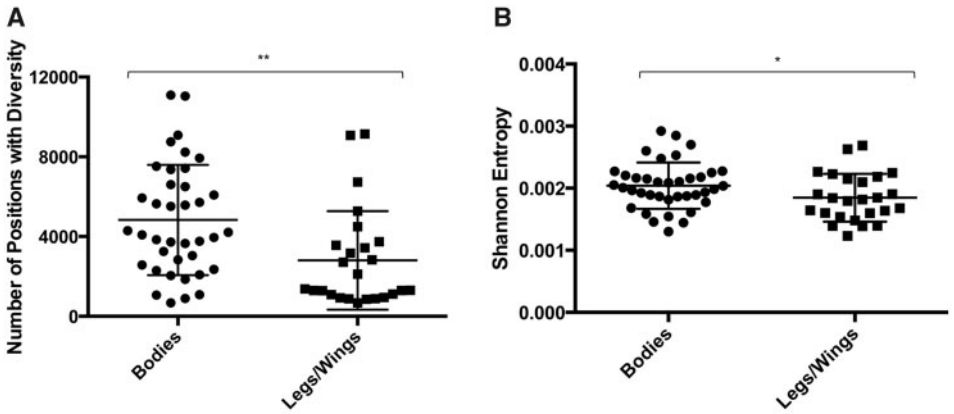

Figure 2. Bodies and legs/wings show differences in both the number of variant positions and Shannon entropy. The difference between bodies and legs/wings calculated for (A) the number of variant positions for all disseminated bodies ($n = 38$) and legs/wings ($n = 25$) with coverage of >90% of the genome, and (B) the diversity of all disseminated bodies ($n = 38$) and legs/wings ($n = 25$). Asterisk indicates statistically significant results by Student's *t*-test, $^{*}P < 0.05$, $^{**}P < 0.01$.

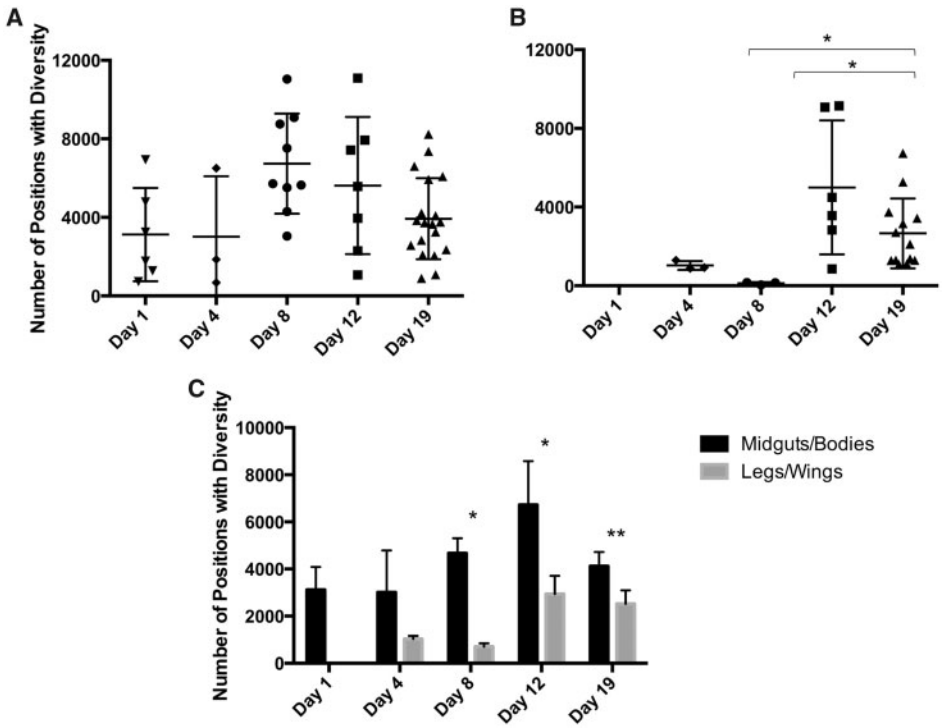

Figure 3. The number of variant positions varies over time and between body compartments. To determine the effect of time on the number of sites showing variation, the number of variant positions was calculated for (A) bodies at Days 1, 4, 8, 12, and 19, significant differences were determined by Kruskal-Wallis test $P = 0.0461$, but no differences between groups were determined by Dunn's multiple comparisons test, and (B) legs/wings at days 4, 8, 12, and 19, significant differences were detected between days by the Kruskal-Wallis test $P = 0.0067$, significant differences between groups were determined by Dunn's multiple comparisons test and are marked on the graph. (C) To determine the effect of a bottleneck on the number of variant sites, samples from the same mosquitoes were compared for Days 4, 8, 12, and 19. Asterisks indicate statistically significant results, $^{*}P < 0.01$, $^{**}P < 0.001$.

in bodies and legs/wings within the individual mosquitoes (i.e. whether the diversity was related between body compartments), all samples for which a paired sample was not available were excluded (Fig. 4C). There was a significantly higher diversity in the bodies compared with the legs/wings measured by Shannon entropy on Days 8 ($n = 3$) ($P = 0.0349$) and 19 ($n = 13$) ($P = 0.0045$) as determined by Student's *t*-test. For Day 1 no legs/wings were sequenced, and on Day 4 only two mosquitoes had sufficient coverage in both bodies and legs/wings. On Day 12 ($n = 3$) no significance was observed. Ratios of the diversity for bodies: legs/wings were also calculated and are shown in Supplementary Fig. S4.

3.6 Distribution of mutations across the genome

Previous studies have demonstrated that there are diversity hotspots, i.e. areas within the genome where increased diversity is found, along the virus genome (Brackney, Beane, and Ebel 2009; Sim et al. 2015). To investigate this further the average diversity was plotted for every base in the genome to determine if there were any regions of increased diversity associated with either time or tissue (Fig. 5). For all the samples, there was a low level of variation across the genome. However, a few spots showed increased diversity compared with the rest of the genome. Two of these sites with increased variation only occurred in the

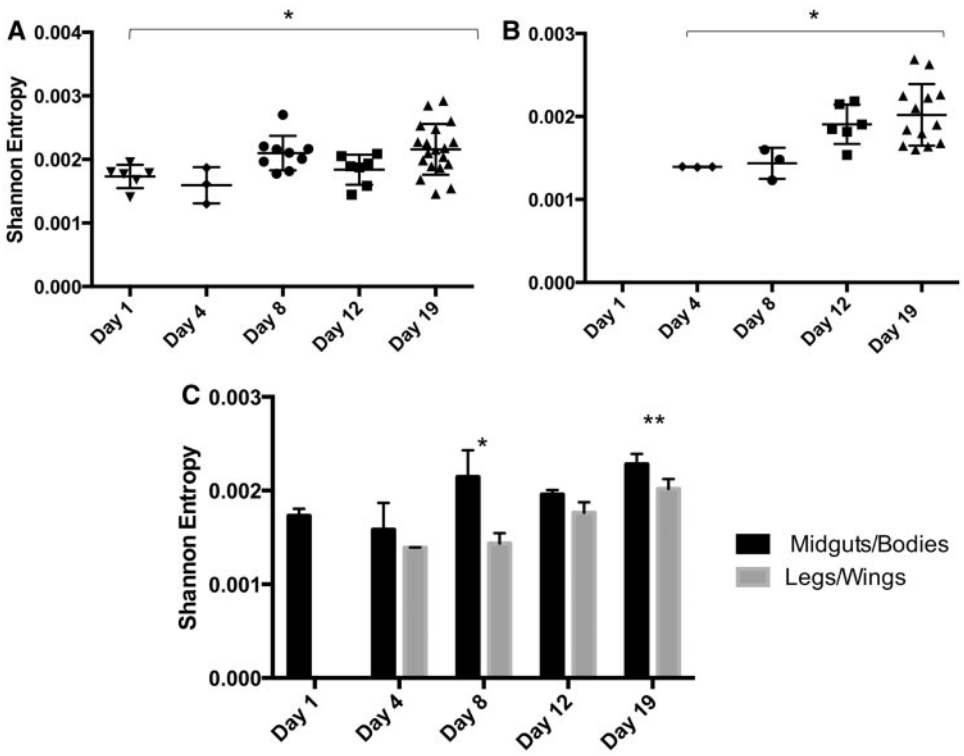

Figure 4. Shannon entropy measure of diversity increases varies over time and between body compartments. Samples were stratified by day and the measure of diversity was calculated for (A) bodies at Days 1, 4, 8, 12, and 19 with statistical significance as determined by Kruskal-Wallis $P = 0.0061$, significant differences between means was determined by Dunn's multiple comparison test and are marked on the graph, and (B) legs/wings at Days 4, 8, 12, and 19, with statistical significance as determined by Kruskal-Wallis $P = 0.0049$, significant differences between means was determined by Dunn's multiple comparison test and are marked on the graph. (C) The diversity between the bodies and legs/wings using paired samples from the same mosquitoes were calculated and plotted for Days 4, 8, 12, and 19, statistical significance of the raw results was performed using a paired *t*-test. Asterisks indicate statistically significant results, $^{*}P < 0.05$, $^{**}P < 0.001$.

negative strand for the bodies (Fig. 5A–E), and the same was not observed for the legs/wings (Fig. 5F–I). These two mutations, 10082 and 10083 are associated with a deletion in the 6K gene, which had resulted in an abnormally high diversity. The deletion has already been described for VEEV (Forrester et al. 2011), although previous studies did not differentiate between the positive and negative strand of the virus. The presence of a known deletion would increase the potential entropy for that position by reducing the coverage at this point, and therefore, it is likely that this peak, as indicated on Fig. 5A by an arrow is a result of the deletion and reduced coverage.

The patterns of diversity change at different time points. On Days 1 and 4 in the bodies, and Days 4 and 8 in the legs/wings, the variation is higher at many positions along the genome (Fig. 5A, B, F, and G). However, it is important to note that at early time points for the bodies (Fig. 5A and B) and legs/wings (Fig. 5F and G) the majority of the genome positions have no variation (Fig. 3A and B), and lower average diversity across the genome (Fig. 4A and B). Many positions of high variance are reduced by Days 8, 12, and 19 (Fig. 5C–E). This scenario was recapitulated for the legs/wings, although the high levels of individual position diversity were observed at Days 4 and 8 (Fig. 5F and G), and many positions with diversity have been reduced by Days 12 and 19 (Fig. 5H and I).

3.7 Consistent selection of minority variants associated with midgut escape

Eight minority variants were found in a vast majority of bodies with disseminated infection at Days 8, 12, and 19. The mutations were not found in consensus at earlier time points. Three of these minority variants were also found at Days 8, 12, and 19 in the legs/wings (Fig. 6; Table 2). Two of these mutations were insertions. The first insertion truncates the RdRp (7,403 ins T/G) while the second insertion truncates the structural polyprotein (9,980 ins T). Both of these mutations result in frameshifts, which are usually considered deleterious, but are found to be conserved in 97% (7,403 ins T/G) (range of Shannon entropy: 0.010499–0.046342), and 77% of mosquito bodies (9,980 ins T) (range of Shannon entropy: 0.0117982–0.0704658) (Table 2). In addition, four non-synonymous changes were observed: an Asp to Ala at Position 5,184 and an Ile to Phe at Position 7,397, both in the RdRp, a Phe to Gln at Position 7,868 in the capsid protein, as well as an Ala to Asp at Position 9,989 in the 6K protein. Although these mutations are only present at low levels, they appear consistently in the bodies and the legs/wings except the Ile to Phe at Position 7,397, which only appears in the bodies. The other mutations were non-coding (T11061G), and a deletion in the capsid at Position 7,874, which results in a truncated structural protein. Again, the deletion that results in a frameshift was surprising as this would likely be a deleterious mutation. These mutations were not identified in the consensus sequence of the virus collected from the mice used for feeding mosquitoes.

4. Discussion

Viral diversity plays a significant role in the infectivity and successful transmission of RNA viruses (Domingo et al. 2005;

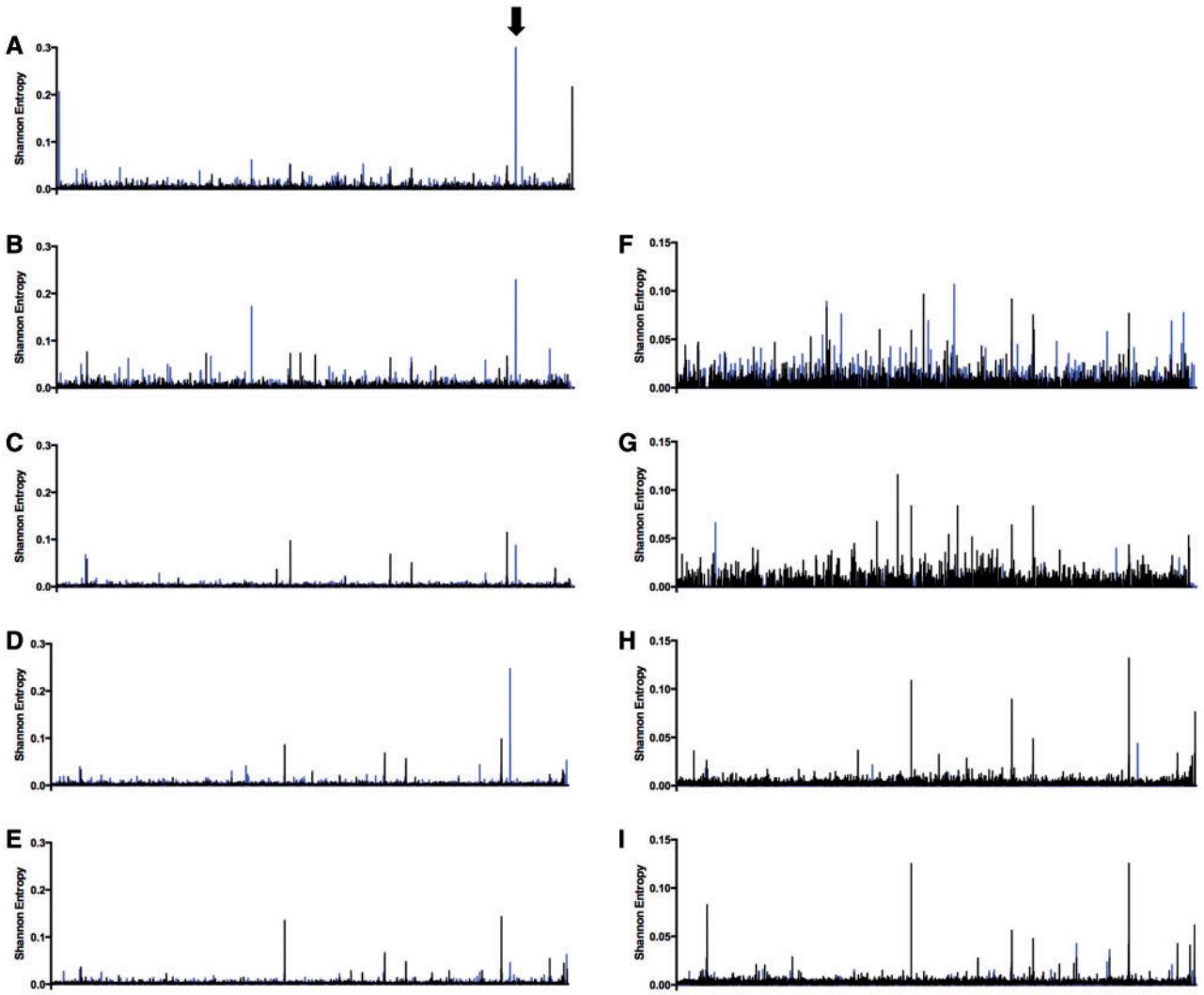

Figure 5. Specific sites in the bodies and legs/wings show evidence of changing diversity over time. All bodies and legs/wings were analyzed for diversity at each position along the genome and the average was determined for each genome position. The average diversity was plotted for bodies on (A) Day 1 ($n = 6$), (B) Day 4 ($n = 3$), (C) Day 8 ($n = 9$), (D) Day 12 ($n = 7$), and (E) Day 19 ($n = 19$), and for legs/wings on (F) Day 4 ($n = 3$), (G) Day 8 ($n = 3$), (H) Day 12 ($n = 6$), and (I) Day 19 ($n = 13$). Black lines indicate the variants on the positive strand and blue lines indicate the variants on the negative strand. The arrow on Fig. 5A shows an area with known deletion mutants.

Manrubia et al. 2005; Escarmis, Lazaro, and Manrubia 2006; Aguirre and Manrubia 2008; Manrubia, Domingo, and Lázaro 2010). In fact, numerous studies have now demonstrated that individual vector tissues exhibit different diversity profiles (Jerzak et al. 2005; Metzner et al. 2009; Wagner et al. 2009; Poon et al. 2010; Wang et al. 2010; Brackney et al. 2011; Parameswaran et al. 2012; Stapleford et al. 2014; Sim et al. 2015; Grubaugh et al. 2016; Lequime et al. 2016). Most RNA viruses replicate without a proofreading enzyme and, therefore, accumulate mutations at a high rate. This gives the viruses a wide range of genetically diverse progeny, and a plasticity in the mutant cloud that has been demonstrated to be important in RNA viruses (Woelk and Holmes 2002). Recently, studies have demonstrated that altering RNA virus fidelity has significant implications for viral infectivity and transmission (Vignuzzi et al. 2006; Coffey et al. 2011; Van Slyke et al. 2012), as a small shift resulting in either increased or decreased fidelity can lead to a significant attenuation of the virus. For arboviruses, which are subject to repeated bottlenecks, it is theorized that there is a need for diversity to compensate for the reduction in the number of virus particles associated with each bottleneck (Manrubia et al. 2005). This study confirms this theory by demonstrating that virus populations presenting lower levels of diversity are less likely to be successful in disseminating through the host. This reduction in diversity may be a result of a smaller number of viral particles that the mosquito ingested; a lack of replication in the midgut, or short-lived infections in the hemocoel that were neutralized. Sampling on Day 1 may have also been too early to allow the virus to disseminate. In any of these cases there is evidence that a baseline of diversity is required for successful dissemination, and overcoming the midgut escape barrier.

This study sought to test the hypothesis that there would be a reduction of viral diversity in the hemocoel following the midgut escape bottleneck, simply due to the reduction in the number of particles that transition across the barrier. This is an example of a physical bottleneck that would reduce the number of variants generated during dissemination of the virus from the mosquito body to the legs and wings. It has been previously shown that the midgut acts as a major bottleneck for VEEV and severely restricts the number of viral particles (from one to fifty)

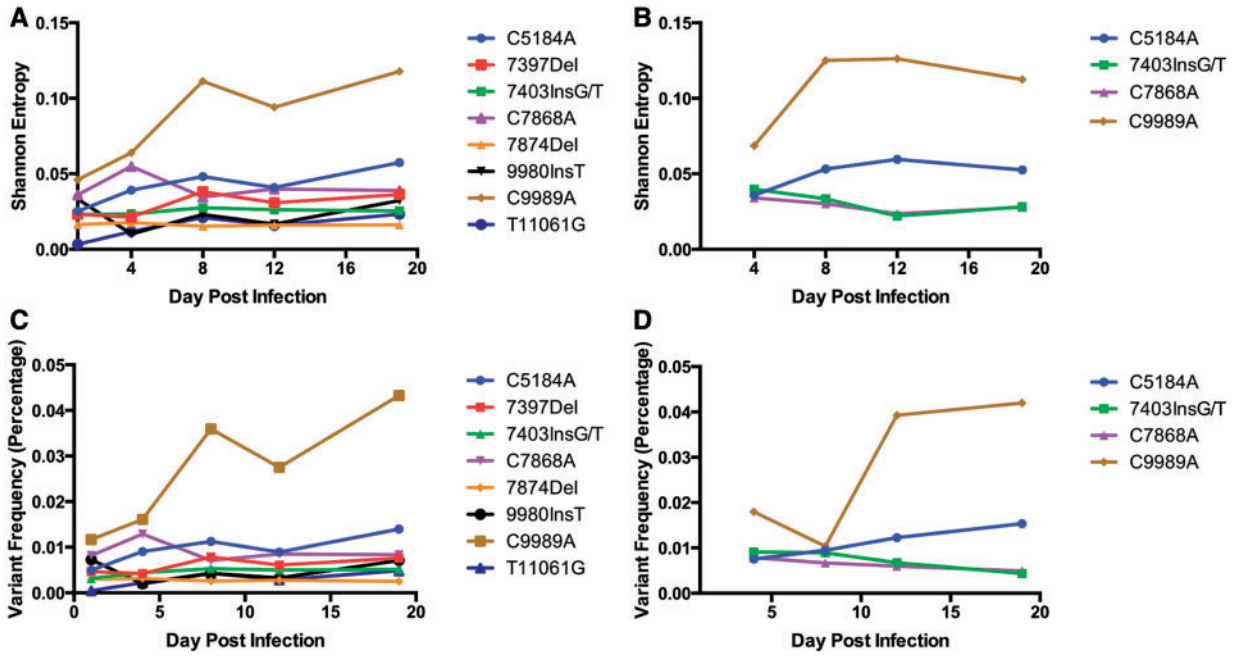

Figure 6. Individual variants are repeatedly found in different body compartments. The top twenty mutations were identified for each day and body compartment. Those that were consistent across every day were selected and the amino acid change and diversity measure identified. The average Shannon Entropy measure for minority variants at Days 4, 8, 12, and 19 in (A) the bodies, and (B) the legs/wings. The average percentage frequency of each variant at Days 4, 8, 12 and 19 in (C) the bodies and (D) the legs/wings.

Table 2. Common minority variants identified in mosquito samples.

Positions	Mutation	Protein	% Bodies	% Legs/Wings
5,184	C-A	Asp-Ala (RdRp)	97	100
7,397	A-T	Ile-Phe (RdRp)	95	–
7,403	Ins T/G	Truncated RdRp	97	96
7,868	C-A	Phe-Gln (Capsid)	97	84
7,874	Del	Truncated Capsid	97	–
9,980	Ins T	6K/Truncated E1	77	–
9,989	C-A	Ala-Asp (6K)	93	72
11,061	T-G	None	77	–

The position and protein changes of the most conserved mutations found on each day, and the percent of mosquitoes' tissues the mutations were found in.

that escape from the midgut. This particle number is dependent on the starting titer of the bloodmeal, and increases in correlation with bloodmeal titer (Forrester et al. 2012), which is consistent with other studies (Smith et al. 2008; Ciota et al. 2012). Conversely, there are studies that did not find bottlenecks in similar systems (Brackney et al. 2011; Gutierrez et al. 2015). When the results from grouped bodies were compared with legs/wings, there was a significant decrease in both viral diversity and the number of variant positions in the legs/wings compared with that observed in the bodies. This result confirms that the presence of a bottleneck decreases the amount of diversity and the number of variant positions in the genome, which is consistent with the midgut escape bottleneck in *C. taeniopus* that has been previously demonstrated (Forrester et al. 2012). Due to the lack of saliva samples with sufficient coverage it was not possible to accurately determine if the saliva showed an increase in diversity that is observed in other viruses (Brackney et al. 2011; Lequime et al. 2016). The limited data that was generated suggested a trend to higher amounts of diversity in the saliva compared with the other tissues (Supplementary Fig. S1). This would be expected, as the increased diversity would facilitate successful transmission of the virus to the vertebrate host (Grubaugh et al. 2017). However, as all saliva samples were excluded due to having <90% of the genome covered, this will need to be investigated further.

When the results were stratified by date of collection similar patterns in bodies and legs/wings emerged. A pattern of initial low variant positions for both bodies (Days 1 and 4) and legs/wings (Days 4 and 8) was observed, with subsequent expansion of the number of variant positions (by Day 8 in bodies and Day 12 in legs/wings), although these changes were only significant in the legs/wings. This increase in diversity and variant positions demonstrates the ability of the virus to restore diversity and the mutant spectrum that is thought to be so crucial to virus propagation and survival. At Day 19 there was a reduction in the number of variant positions for both bodies and legs/wings, but not diversity as measured by Shannon entropy. It is possible that differences in titer is responsible for this change, but previous data collected suggests that this is not the case, as titers tend to be between 3 and 5 $\log_{10}$ PFU/ml (Supplementary Table S2). Rather, the reduction suggests that the virus is undergoing purifying selection and that a few mutations are present at higher frequencies. It should be mentioned, that there is the potential for sampling bias associated with these results, as mosquitoes could not be repeatedly sampled. However, the biological patterns reflect the prevailing hypothesis that positive selection is the driver of Alphavirus evolution and are, therefore, an accurate representation of what is going on in the mosquito.

The observed shifts in diversity that are seen, particularly in the legs/wings could be due to many different factors. One of the most likely factors is the mosquito immune system, in particular the RNAi response, which is a sequence specific response that is the most important antiviral response in the

mosquito (Bronkhorst and van Rij 2014; Piatek and Werner 2014). Previous research has shown that specific regions in the genome are more frequently targeted by the RNAi response and these regions exhibit increased genetic diversity (Brackney et al. 2009). The purifying selection observed by Day 19 could be the result of the RNAi response in the mosquito. In the future, the extent of the relationship between virus genome variation and the genetic regions targeted by the RNAi response will be explored. Whichever mechanism is associated with selection, the fact remains that there is a progressive refinement of the mutational spectrum over time, and it is likely that this has implications for transmission as this may impact the diversity in the mosquito salivary glands and, therefore, the initial diversity in the vertebrate host.

Other reported analyses of viruses using NGS have shown the presence of regions of higher diversity in the viral genome (Brackney et al. 2009; Myles, Morazzani, and Adelman 2009; Sim et al. 2015; Stapleford et al. 2016). These hotspots are unique for each virus and each tissue. With VEEV-infected *C. taeniopus*, there were differences between the hotspots in the bodies compared with the hotspots in the legs/wings, which was expected. However, there were also several sites of variation along the genome that were shared between bodies, and legs/wings. This suggests that diversity hotspots are specific to tissue compartments, although there is some overlap. These mutations are either perpetuating by individual mutants making their way through the midgut barrier or by being regenerated following the midgut bottleneck. However, because the midgut bottleneck is so severe (Forrester et al. 2012), this is likely due to regeneration of the mutant swarm after the bottleneck. It was also observed that unlike the bodies, the legs/wings had fewer conserved mutations, suggesting that the bottleneck does play a role in shaping the mutant spectrum. It could also be interpreted that there are perhaps fewer pressures associated with the legs/wings, which may also result in fewer conserved mutations. Regardless, the origin of this diversity is still recognized as a product of the infidelity of the viral polymerase and the number of viral replication events that allows the virus to increase the diversity following replication. Further work is required to determine if these mutations are a result of selection pressures or random stochastic generation following the escape from the midgut.

Bottlenecks are theorized to have the potential to shift the population due to selection of just a few viral particles. Six mutations were observed in >90% of the bodies with a successfully disseminated infection (Table 2). Determining exactly how this may impact the successful dissemination in the mosquito will require further investigation especially as only four of these mutations were present in the legs/wings as well. In addition, none of these mutations were found in >70% of non-disseminated bodies, suggesting that they may play a role in successful dissemination of the virus in the mosquito (data not shown). Two more mutations were observed in a large proportion of the bodies studied (>75%). Of the eight mutations found in a large number of mosquitoes, two result in frameshifts that would compromise the structure of the virion. Although these sequences would not be viable on their own, defective particles can accompany functional particles and have been identified in samples collected in mosquitoes with suspected sustained transmission (Aaskov 2006). These mutations have the potential to influence the infection and dissemination of the virus in the mosquito. The use of six mice to feed mosquitoes and generate the samples may influence the resulting diversity and mutations. Unfortunately, low-coverage limited analysis beyond comparison to a consensus sequence and further work will need to be done to determine if the individual mouse diversity resulted in changes in the mosquito diversity. This may be a confounding factor whose effect is still to be determined; however, it is unlikely that all commonly found mutations were generated and transmitted in the mice. Mosquitoes ingest between 1 and 100 virus particles at 10^5 PFU/ml, which for Alphaviruses equates to a maximum of 500,000 RNA genomes (Forrester et al. 2012). Using the percentage of reads at which these particles were sequenced, the probability that all eight mutations would be picked up from an individual mouse is 3×10^{-12}. Thus, it is more likely that these are generated during the initial replication in the mosquito and take several days for detection in a high proportion of mosquitoes.

Previous studies with WNV have shown the generation of mutations following bottlenecks in multiple mosquito species (Grubaugh et al. 2016). This study used a single mosquito species, and also found there was the generation of mutations following bottlenecks (Day 8). This is likely due to regeneration of the mutants and may also reflect the presence of molecular memory as has previously been described for both foot and mouth disease virus (FMDV) and human immunodeficiency virus (HIV) (Ruiz-Jarabo et al. 2000; Garcia-Arriaza, Domingo, and Briones 2007). Subsequent purifying selection was also observed, as exemplified by a slight reduction in diversity in both bodies and legs/wings on Day 19 compared with Day 12. These results are similar to that observed in DENV; although for DENV, this was only in the midgut and the reduction was observed at Day 7 compared with the initial diversity at Day 4 (Lequime et al. 2016). It is likely that these results are all mosquito-virus pairing specific because previous literature studying WNV and DENV demonstrated that even small changes in mosquito genetics can alter the diversity spectrum (Grubaugh et al. 2016; Lequime et al. 2016). The current study was unable to address this aspect of variation, as only one colony of *C. taeniopus* was used. However, by sampling repeatedly and frequently, changes were observed that were associated with the presence of bottlenecks that had been previously quantified by Forrester et al. (2012). Additionally, the increase in the number of overall samples allowed for the identification of specific mutations that are present in numerous tissues. The greater number of samples sequenced (bodies $n = 38$ and legs/wings $n = 25$) increases the potential of observing minority variants that play a role in viral replication and dissemination. Randomly selecting a subset of the variants did not provide the same robustness in identification of these minority variants (data not shown).

Using VEEV as a model system, this study provides a base of evidence to aid studies showing the importance of minority variants in the dissemination and successful infection of RNA viruses in the mosquito. This study also shows that there is a significant reduction in the number of variant positions in the legs/wings following the bottleneck associated with escape from the midgut, which is restored 4 days later. Interestingly, there was no reduction in diversity measure seen for the legs/wings, unlike the number of variant positions, which suggests that measures of diversity by themselves do not fully capture the changes associated with viral diversity. Several of the mutations detected were consistently selected for in many samples, suggesting that at least some variants are not caused by random generation of mutations, but instead because of directed selection, and this needs to be validated further. Consequently, these mutations may play an important role in the successful infection and dissemination of the virus in the mosquito. How these mutations interact with the mosquito immune pathways remains to be determined. This study confirms that the

presence of bottlenecks does play a role in modulating the minority variants present in different compartments, but this is not necessarily captured fully by measuring the diversity across the entire genome. However, the presence of minority variants appears to play an important role and either the selection or regeneration of these variants facilitates the infection and dissemination of VEEV in *C. taeniopus*.

Data availability

Data will be available at NCBI BioProject ID: PRJNA390970.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Funding

This work was funded by a grant to NLF from the NIH R01 - AI095753-01A1.

Conflict of interest: None declared.

References

Aaskov, J. (2006) 'Long-Term Transmission of Defective RNA Viruses in Humans and Aedes Mosquitoes', *Science (New York, N.Y.)*, 311: 236–8.

Acevedo, A., Brodsky, L., and Andino, R. (2014) 'Mutation and Fitness Landscapes of an RNA Virus Revealed through Population Sequencing', *Nature*, 505: 686–90.

Aguirre, J., and Manrubia, S. C. (2008) 'Effects of Spatial Competition on the Diversity of a Quasispecies', *Physical Review Letters*, 100:

Andrews, S. (2015), 'FastQC: A Quality Control Tool for High Throughput Sequence Data', https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anishchenko, M. et al. (2006) 'Venezuelan Encephalitis Emergence Mediated by a Phylogenetically Predicted Viral Mutation', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 4994–9.

Borderia, A. V. et al. (2015) 'Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype', *PLoS Pathogens*, 11: e1004838.

Brackney, D. E., Beane, J. E., and Ebel, G. D. (2009) 'RNAi Targeting of West Nile Virus in Mosquito Midguts Promotes Virus Diversitification', *PLoS Pathogens*, 5: e1000502.

——, ——, and —— et al. (2011) 'West Nile Virus Genetic Diversity Is Maintained during Transmission by *Culex pipiens Quinquefasciatus* Mosquitoes', *PLoS One*, 6: e24466.

Brault, A. C. et al. (2002) 'Positively Charged Amino Acid Substitutions in the E2 Envelope Glycoprotein Are Associated with the Emergence of Venezuelan Equine Encephalitis Virus', *Journal of Virology*, 76: 1718–30.

—— et al. (2004) 'Venezuelan Equine Encephalitis Emergence: Enhanced Vector Infection from a Single Amino Acid Substitution in the Envelope Glycoprotein', *Proceedings of the National Academy of Sciences of the United States of America*, 101: 11344–9.

Bronkhorst, A. W., and van Rij, R. P. (2014) 'The Long and Short of Antiviral Defense: Small RNA-Based Immunity in Insects', *Current Opinion in Virology*, 7: 19–28.

Ciota, A. et al. (2012) 'Quantification of Intrahost Bottlenecks of *West Nile Virus* in *Culex pipiens* Mosquitoes Using an Artificial Mutant Swarm', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 12: 557–64.

—— et al. (2009) 'Experimental Passage of St. Louis Encephalitis Virus in Vivo in Mosquitoes and Chickens Reveals Evolutionarily Significant Virus Characteristics', *PLoS One*, 17: e7876.

—— et al. (2007) 'Cell-Specific Adaptation of Two Flaviviruses following Serial Passage in Mosquito Cell Culture', *Virology*, 357: 165–74.

Coffey, L. L. et al. (2011) 'Arbovirus High Fidelity Variant Loses Fitness in Mosquitoes and Mice', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 16038–43.

—— et al. (2008) 'Arbovirus Evolution *in Vivo* Is Constrained by Host Alternation', *Proceedings of the National Academy of Sciencesof the United States of America*, 105: 6970–5.

Domingo, E. et al. (2005) 'Quasispecies Dynamics and RNA Virus Extinction', *Virus Research*, 107: 129–39.

Elena, S. F., and Sanjuan, R. (2005) 'Adaptive Value of High Mutation Rates of RNA Viruses: Separating Causes from Consequences', *Journal of Virology*, 79: 11555–8.

Escarmis, C., Lazaro, E., and Manrubia, S. C. (2006) 'Population Bottlenecks in Quasispecies Dynamics', *Quasispecies: Concept and Implications for Virology*, 299: 141–70.

Forrester, N. L. et al. (2011) 'Analysis of Intrahost Variation in Venezuelan Equine Encephalitis Virus Reveals Repeated Deletions in the 6-Kilodalton Protein Gene', *Journal of Virology*, 85: 8709–17.

—— et al. (2012) 'Vector-Borne Transmissions Imposes a Severe Bottleneck on an RNA Virus Population', *PLoS Pathogens*, 8: e1002897.

Garcia-Arriaza, J., Domingo, E., and Briones, C. (2007) 'Characterization of Minority Subpopulations in the Mutant Spectrum of HIV-1 Quasispecies by Successive Specific Amplifications', *Virus Research*, 129: 123–34.

Greene, I. P. et al. (2005) 'Envelope Glycoprotein Mutations Mediate Equine Amplificiation and Virulence of Epizootic Venezuelan Equine Encephalitis Virus', *Journal of Virology*, 79: 9128–33.

Grubaugh, N. D. et al. (2016) 'Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching', *Cell Host and Microbe*, 19: 481–92.

—— et al. (2017) 'Mosquitoes Transmit Unique West Nile Virus Populations during Each Feeding Episode', *Cell Reports*, 19: 709–18.

Gutierrez, S. et al. (2015) 'Demographics of Natural Oral Infection of Mosquitoes by Venezuelan Equine Encephalitis Virus', *Journal of Virology*, 89: 4020–2.

Jerzak, G. (2005) 'Genetic Variation in West Nile Virus from Naturally Infected Mosquitoes and Birds Suggests Quasispecies Structure and Strong Purifying Selection', *Journal of General Virology*, 86: 2175–83.

Lauring, A. S., and Andino, R. (2010) 'Quasispecies Theory and the Behavior of RNA Viruses', *PLoS Pathogens*, 6: e1001005.

Lequime, S. et al. (2016) 'Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-Host Genetic Diversity in Mosquitoes', *PLoS Genetics*, 12: e1006111.

Li, Y.-G. et al. (2013) 'Chikungunya Virus Induces a More Moderate Cytopathic Effect in Mosquito Cells than in Mammalian Cells', *Intervirology*, 56: 6–12.

Lin, J. (1991) 'Divergence Measures Based on the Shannon Entropy', *IEEE Transactions on Information Theory*, 37: 145–51.

Manrubia, S. C., Domingo, E., and Lázaro, E. (2010) 'Pathways to Extinction: Beyond the Error Threshold', *Philosophicals*', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365: 1943–52.

——, ——, and —— et al. (2005) 'High Mutation Rates, Bottlenecks, and Robustness of Viral RNA Quasispecies', *Gene*, 347: 273–82.

Metzner, K. J. et al. (2009) 'Minority Quasispecies of Drug-Resistant HIV-1 That Lead to Early Therapy Failure in Treatment-Naive and -Adherent Patients', *Clinical Infectious Diseases*, 48: 239–47.

Myles, K. M., Morazzani, E. M., and Adelman, Z. N. (2009) 'Origins of alphavirus-Derived Small RNA's in Mosquitoes', *RNA Biology*, 6: 387–91.

Oberste, M. S., Parker, M. D., and Smith, J. F. (1996) 'Complete Sequence of Venezuelan Equine Encephalitis Virus Subtype IE Reveals Conserved and Hypervariable Domains within the C Terminus of nsP3', *Virology*, 219: 314–20.

Parameswaran, P. et al. (2012) 'Genome-Wide Patterns of Intrahuman Dengue Virus Diversity Reveal Associations with Viral Phylogenetic Clade and Intrahost Diversity', *Journal of Virology*, 86: 8546–58.

Piatek, M. J., and Werner, A. (2014) 'Endogenous siRNAs, Regulators of Internal Affairs', *Biochemical Society Transactions*, 42: 1174–9.

Poon, A. F. Y. et al. (2010) 'Phylogenetic Analysis of Population-Based and Deep Sequencing Data to Identify Coevolving Sites in the *Nef* Gene of HIV-1', *Molecular Biology and Evolution*, 27: 819–32.

Rozen-Gagnon, K. et al. (2014) 'Alphavirus Mutator Variants Present Host-Specific Defects and Attenuation in Mammalian and Insect Models', *PLoS Pathogens*, 10: e1003877.

Ruiz-Jarabo, C. M. et al. (2000) 'Memory in Viral Quasispecies', *Journal of Virology*, 74: 3543–7.

Scherer, W. F., Dickerman, R. W., and Ordonez, J. V. (1970) 'Discovery and Geographic Distribution of Venezuelan Encephatlitis Virus in Guatemala, Honduras, and British Honduras during 1965-68, and Its Possible Movement to Central America and Mexico', *The American Journal of Tropical Medicine and Hygiene*, 19: 703–11.

Sim, S. et al. (2015) 'Tracking Dengue Virus Intra-Host Genetic Diversity during Human-to-Mosquito Transmission', *PLoS Neglected Tropical Diseases*, 9: e0004052.

Smith, D. R. et al. (2008) 'Venezuelan Equine Encephalitis Virus in the Mosquito Vector Aedes Taeniorhynchus: Infection Initiated by a Small Number of Susceptible Epithelial Cells and a Population Bottleneck', *Virology*, 372: 176–86.

—— et al. (2006) 'Venezuelan Equine Encephalitis Virus Transmission and Effect on Pathogenesis', *Emerging Infectious Diseases*, 12: 1190–6.

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.

—— et al. (2014) 'Emergence and Transmission of Arbovirus Evolutionary Intermediates with Epidemic Potential', *Cell Host and Microbe*, 15: 706–16.

Van Slyke, G. A. et al. (2012) 'Point Mutations in the West Nile Virus (Flaviviridae; Flavivirus) RNA-Dependent RNA Polymerase Alter Viral Fitness in a Host-Dependent Manner *in Vitro* and *in Vivo*', *Virology*, 427: 18–24.

Vignuzzi, M. et al. (2006) 'Quasispecies Diversity Determines Pathogenesis through Co-Operative Interactions in a Viral Population', *Nature*, 439: 344–8.

Wagner, T. A. et al. (2009) 'Increased Mutations in Env and Pol Suggest Greater HIV-1 Replication in Sputum-Compared to Blood-Derived Viruses', *Aids*, 23: 923–8.

Wang, G. P. et al. (2010) 'HCV Transmission Bottlenecks Analyzed by Deep Sequencing', *Journal of Virology*, 84: 6218–28.

Weaver, S. C. et al. (2004) 'Venezuelan Equine Encephalitis', *Annual Review of Entomology*, 49: 141–74.

Woelk, C. H., and Holmes, E. C. (2002) 'Reduced Positive Selection in Vector-Borne RNA Viruses', *Molecular Biology and Evolution*, 19: 2333–6.