



OPEN

Evolutionary dynamics of *Tomato spotted wilt virus* within and between alternate plant hosts and thrips

Casey L. Ruark-Seward¹, Brian Bonville¹, George Kennedy¹ & David A. Rasmussen^{1,2}✉

Tomato spotted wilt virus (TSWV) is a generalist pathogen with one of the broadest known host ranges among RNA viruses. To understand how TSWV adapts to different hosts, we experimentally passaged viral populations between two alternate hosts, *Emilia sochifolia* and *Datura stramonium*, and an obligate vector in which it also replicates, western flower thrips (*Frankliniella occidentalis*). Deep sequencing viral populations at multiple time points allowed us to track the evolutionary dynamics of viral populations within and between hosts. High levels of viral genetic diversity were maintained in both plants and thrips between transmission events. Rapid fluctuations in the frequency of amino acid variants indicated strong host-specific selection pressures on proteins involved in viral movement (NSm) and replication (RdRp). While several genetic variants showed opposing fitness effects in different hosts, fitness effects were generally positively correlated between hosts indicating that positive rather than antagonistic pleiotropy is pervasive. These results suggest that high levels of genetic diversity together with the positive pleiotropic effects of mutations have allowed TSWV to rapidly adapt to new hosts and expand its host range.

Despite theoretical predictions that specialist pathogens should outcompete generalists, multi-host pathogens are abundant in nature¹. One extreme example of such generalism is provided by plant viruses, which unlike their animal infecting counterparts, often infect hundreds or even thousands of phylogenetically distant host species^{2,3}. In turn, generalist plant viruses are often transmitted by generalist plant-feeding insects such as aphids, thrips and whiteflies, which feed on a wide range of plants^{4–6}. Thus, highly polyphagous vectors routinely transmit viruses between different plant species, which may strongly favor generalists that can either readily adapt or remain adapted to a wide range of potential hosts.

Tomato spotted wilt virus (TSWV) is a negative-stranded RNA virus in the genus Orthotospovirus in the order *Bunyavirales*⁷. Even when considered among other plant viruses, TSWV is a generalist *par excellence*, with a described host range of over 1000 plant species distributed across more than 90 families of angiosperms^{8,9}. Important hosts include solanaceous crops like tomato, pepper and tobacco, of which TSWV is a major constraint on production worldwide¹⁰. Orthotospoviruses like TSWV also have the rather rare ability among plant viruses to persist and replicate in their insect vector, thrips¹¹. While little is known about the realized host range of any particular genotype of TSWV in nature, TSWV vectors like western flower thrips (*Frankliniella occidentalis*) feed on hundreds of plant species⁵, making it highly probable that a single viral lineage may move between multiple crop and wild host species over the course of a single growing season and then overwinter in another perennial host¹².

The TSWV genome is composed of three viral genomic RNAs or segments (small, medium and large) encoding for five proteins⁹. Three of these proteins appear to be conserved among all bunyaviruses: a nucleocapsid (N) gene, a glycoprotein composed of two domains (Gn/Gc) and the RNA-dependent RNA polymerase (RdRp) that replicates and transcribes the viral genome. The genome also encodes for two genes that likely reflect specific adaptations to plants and insects: the silencing suppressor NSs, which helps counteract host antiviral RNA silencing, and the movement protein NSm, which is involved in both cell-to-cell and long-distance movement in plants¹³. The TSWV genome is multi- or ambisense in that two of these proteins (NSs and NSm) are encoded

¹Department of Entomology and Plant Pathology, North Carolina State University, Ricks Hall 312, 1 Lampe Drive, Raleigh, NC 27607, USA. ²Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA. ✉email: drasmus@ncsu.edu

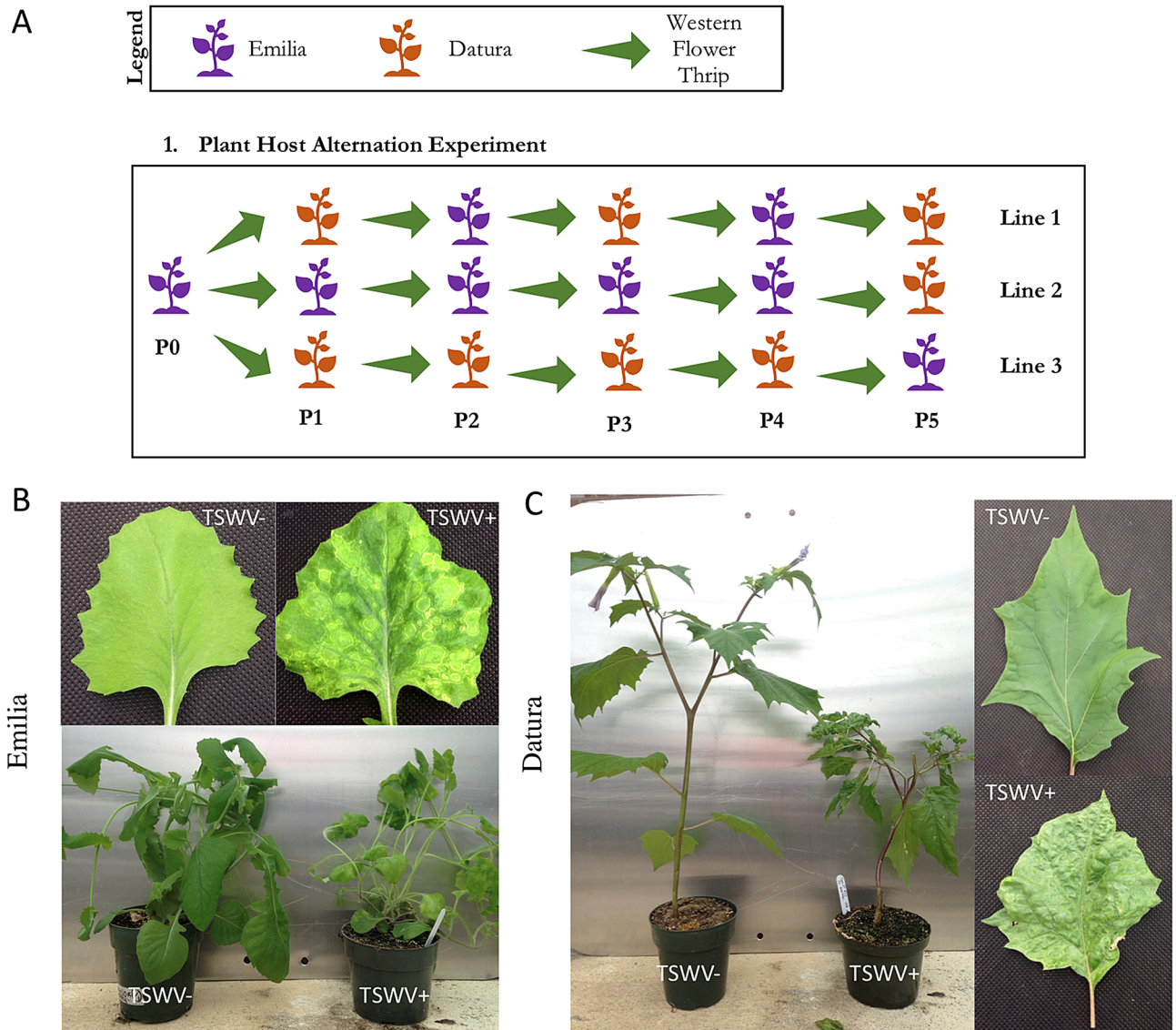


Figure 1. Schematic overview of the viral passing experiment. (A) Virus was passed for five generations alternately on *Emilia* and *Datura* (Line 1), passed exclusively on *Emilia* and then back passed to *Datura* (Line 2), or passed exclusively on *Datura* and then back passed to *Emilia* (Line 3). (B) Comparison of infected (TSWV+) and uninfected control (TSWV-) *Emilia* plants at passage P1. Infected *Emilia* leaves display the classic ringspot symptoms associated with TSWV. (C) Comparison of infected and uninfected *Datura* plants at passage P1.

in a positive-sense orientation while the other three are in a negative-sense orientation and must first be transcribed by RdRp into viral mRNAs before translation of the full set of proteins required for infection can occur.

Beyond these large-scale genomic features, what factors shape TSWV's broad host range remains little explored. Among RNA viruses more generally, fitness tradeoffs between alternative hosts are widely assumed to limit simultaneous adaptation to multiple hosts^{14–16}. Consistent with these predictions, experimental evolution studies in which viruses are passed between alternate hosts and/or vectors have provided evidence that mutations that increase fitness in one host often decrease fitness in another host, suggesting that antagonistic pleiotropy underlies fitness tradeoffs between hosts^{17,18}. Nevertheless, generalists with high fitness across multiple hosts often evolve in experimental evolution studies where microbes are serially passed between different host environments^{19–21}. Thus, the extent to which fitness tradeoffs actually limit adaptation to multiple hosts remains unclear, especially for generalist pathogens like TSWV whose past ecological success suggests that the virus may have evolved efficient strategies to circumvent fitness tradeoffs and readily adapt to new host environments.

To address these questions, we experimentally passed a field-collected isolate of TSWV between plants using western flower thrips as a vector (Fig. 1). In one line, we alternately passed the virus between two plant species: *Emilia sochifolia* (Asteraceae) and *Datura stramonium* (Solanaceae). In two other lines, we passed the virus exclusively on either *Emilia* or *Datura*. Hereafter, we refer to these as the *Alternating*, *Emilia* and *Datura* lines. During each passage cycle, the viral population was deep sequenced at multiple time points in plants and

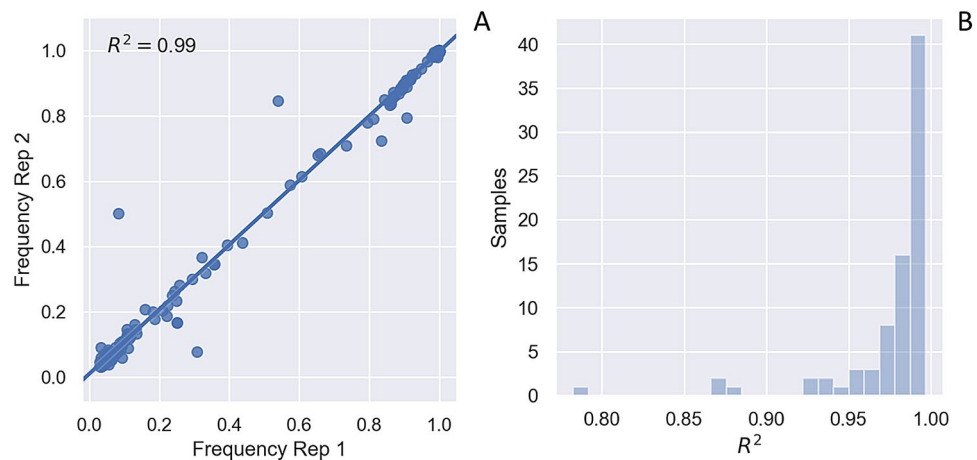


Figure 2. Single nucleotide variants (SNV) frequencies in paired sequencing replicates. (A) Frequency of individual SNVs at passage cycle P0 in two paired sequencing replicates obtained from independent reverse transcription (RT) reactions. The R^2 value for variant frequencies between each paired replicate was determined by least-squares regression. (B) Histogram showing the distribution of R^2 values between paired replicates at all time points sampled during the passing experiment.

once in thrips. This time-resolved deep sequencing data allowed us to track the evolutionary dynamics of viral populations both within and between hosts, and thereby quantify the fitness of genetic variants in different host environments.

Results

Deep sequencing of TSWV populations. Viral populations were deep sequenced twice at each sampling time point using paired sequencing replicates that originated from two independent reverse transcription reactions. Sequencing provided high coverage across all three segments of the TSWV genome, with a depth of coverage generally $> 1000\times$ in both sequencing replicates (Supp. Figure 1). To ensure that genetic variants in the sequence reads represented actual variants present in the viral population, a conservative variant calling scheme was used in which a variant needed to be present at a frequency of at least 3% in both sequencing replicates in order to be called a true variant (see “Methods”). Called variants are therefore unlikely to represent RT, PCR or sequencing errors. Furthermore, single nucleotide variant (SNV) frequencies were highly correlated between the paired sequencing replicates (Fig. 2), suggesting that our sequencing protocol introduced little sampling variance and that the sequence data accurately reflects the genetic composition of the original viral populations.

Within and between host patterns of viral genetic diversity. First, the initial viral population in the field-collected tomato fruit (TF2) was sequenced. Clear hotspots of genetic diversity can be seen in the TF2 population, especially in intergenic regions which contain a large number of SNVs and indels (Fig. 3). Protein coding regions exhibit less overall variability, but there is still considerable nonsynonymous variation. Of the 49 SNVs that fall within coding regions, 33 (67%) are predicted amino acid variants (AAVs) based on their translated sequences. There is also a hotspot of coding diversity at the 5' end of the movement protein NSm. Viral diversity was similar in a leaf sampled from the same tomato plant in the field (Supp. Figure 2).

TSWV was mechanically transferred from the field collected tomato sample to a single *Emilia* plant, P0, from which all three experimental lines were derived. Diversity within viral populations was quantified using average pairwise distance between viral sequences. In P0, the average pairwise distance (D) between viral sequences was 25 single nucleotide substitutions, or 1.6×10^{-3} mutations per site. Diversity tended to decrease slightly over time in all lines (Fig. 4; blue), but tended to decline more rapidly between sampling time points in *Datura* (Mean change in D = -1.08) than in *Emilia* (Mean change = -0.08), although the difference between hosts was not statistically significant (Welch's t-test: 1.11, DF = 37.86, p-value = 0.27). When passaged from thrips to plants, diversity also tended to decline slightly (Mean change = -0.11), although not significantly so (One-sided t-test: -0.17 p-value = 0.87). In contrast, genetic diversity tended to rebound whenever the virus was passaged from plants back through thrips (Mean change = 0.49), although again not significantly so (One-sided t-test: 0.60 p-value = 0.56). While increased diversity in thrips may indicate a lack of severe population bottlenecks in thrips, this may be an artefact of using multiple thrips to passage the virus between hosts and pooling these thrips for sequencing, but also may be due to TSWV's ability to replicate in thrips. In addition to diversity, we also quantified divergence between viral populations using the average pairwise distance between viral sequences sampled at different times points. All three viral lines diverged from P0 by about 5–10 mutations by P5, but divergence slowed after the first two passage cycles (Fig. 4; orange).

Within-host genetic diversity tended to mirror species-level diversity in TSWV samples collected around the world (Fig. 5). Although within-host diversity is lower than species-level diversity, hotspots of diversity can be seen both in intergenic regions as well as in certain protein coding regions, especially at the N-terminus of NSm.

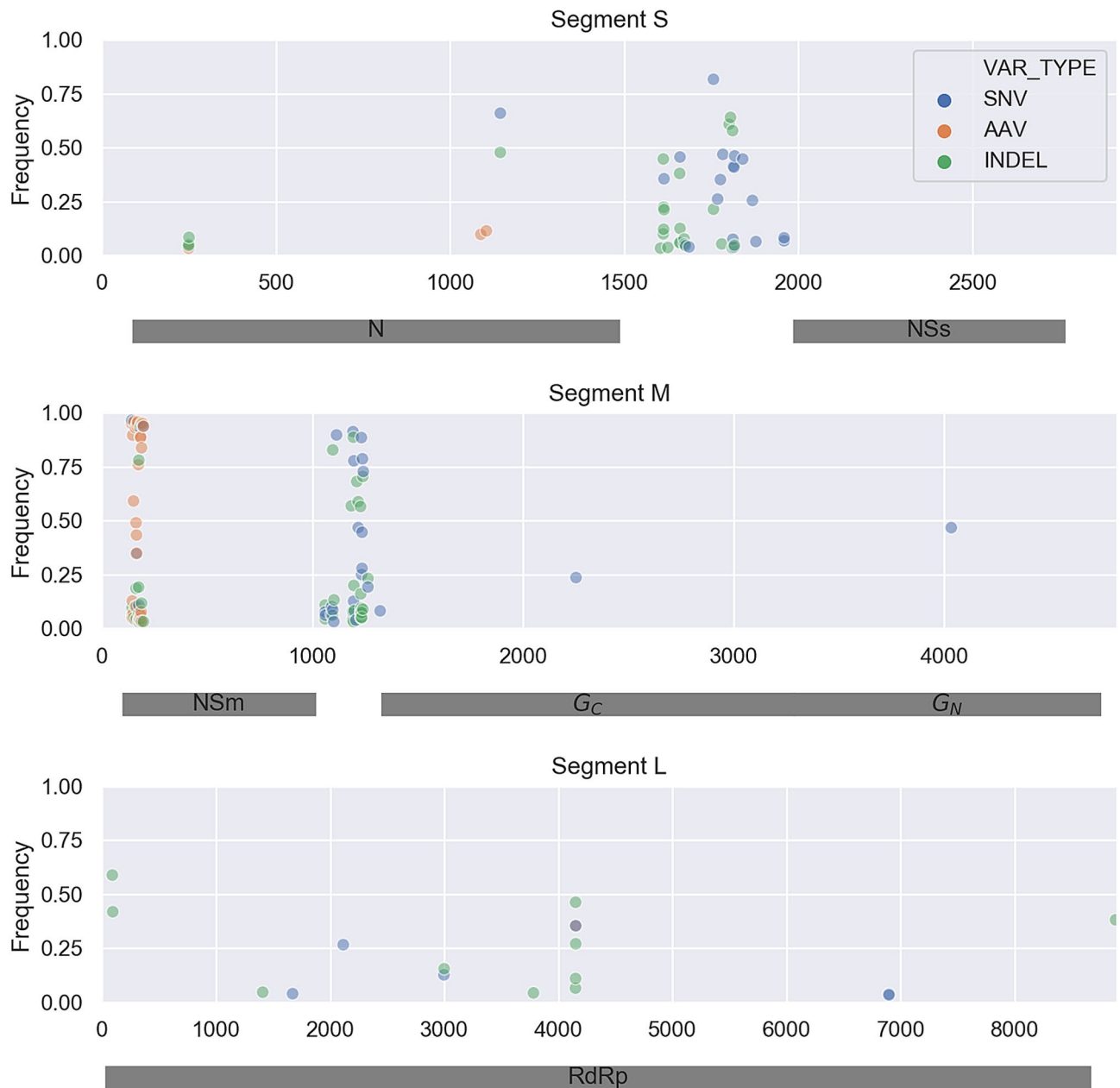


Figure 3. Viral genetic diversity in a naturally infected tomato fruit (TF2) collected in the field. All three passaging lines were derived from this TF2 isolate. The frequency of each single nucleotide variant (SNV), amino acid variant (AAV) and indel is plotted at its respective position along the three segments of TSWV's genome. Shaded grey rectangles represent protein coding regions.

Evolutionary dynamics of individual variants. While divergence between viral populations was low at the genome-wide level, tracking the evolutionary dynamics of individual variants revealed rapid and often host-specific changes in variant frequencies over time. The evolutionary dynamics of all SNVs in the *Alternating* line through time is shown in Supp. Figure 3. Because variants that are differentially selected between hosts are of particular interest, we looked for variants that were enriched in either plant host or the vector. Here, a variant is considered to be enriched if its average frequency was >5% higher in one host than in an alternate host, where the average is computed over all sampled time points. The >5% threshold was chosen heuristically to focus attention on variants with the most dramatic frequency changes between hosts while filtering out variants with roughly constant frequencies.

To more easily visualize the evolutionary dynamics of individual variants, only amino acid variants (AAVs) are displayed for the *Alternating* line in Fig. 6. Figure 6A shows the evolutionary dynamics of variants enriched in one plant host (*Emilia* or *Datura*) relative to the other. Variants can be seen that increase in frequency in *Emilia* but decline in *Datura* (e.g. NSm V17G) as well as variants that increase in frequency in *Datura* but decline in *Emilia* (e.g. NSm N22S). Figure 6B shows AAVs that are enriched in plants versus thrips or in thrips versus plants. There are quite a few variants on NSs, G_N/G_C and RdRp that change dramatically in frequency whenever



Figure 4. Genetic diversity within and divergence between viral populations at each sampled time point. The average pairwise distance within populations (diversity) is shown in blue and the average pairwise distance between each population and the founder viral population (divergence) at P₀ is shown in orange. Subscripts in the sample names denote passage number, superscripts denote days post infection. Samples from plants (P) are colored according to whether they were from *Emilia* (green) or *Datura* (blue). Samples from thrips (T) are colored black.

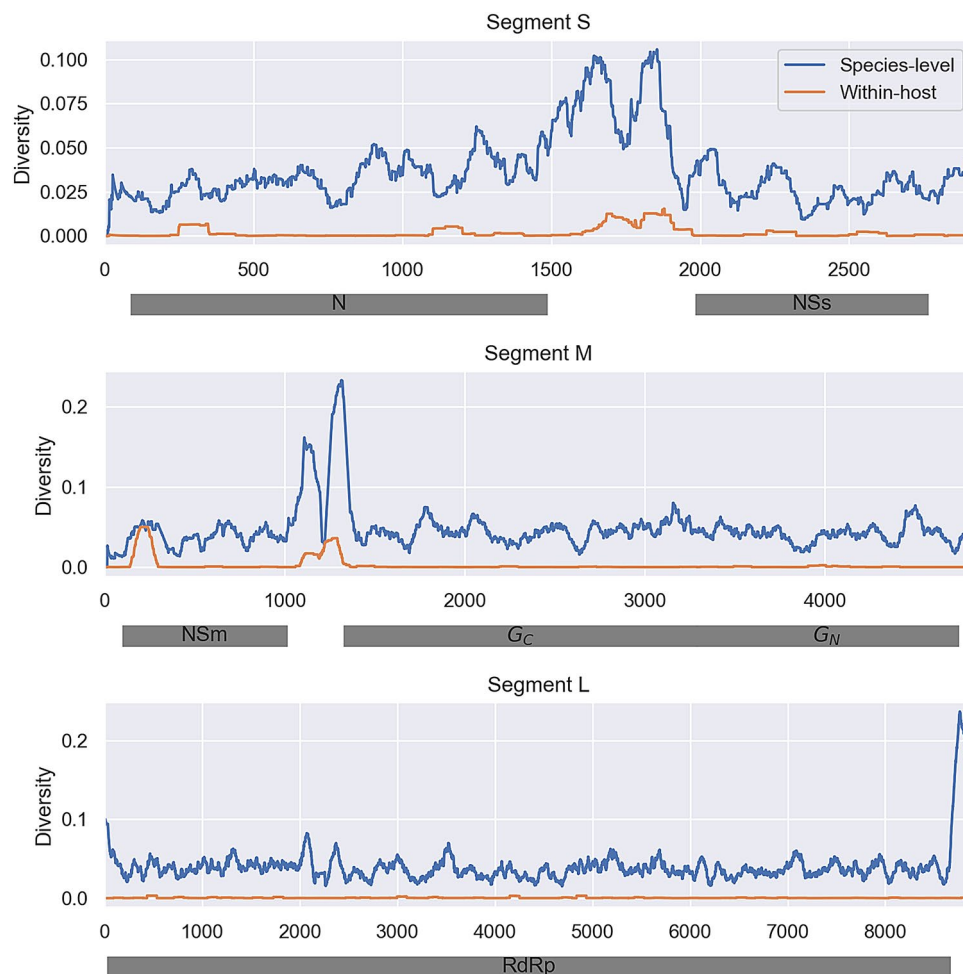


Figure 5. Comparison of within-host (orange) versus species-level (blue) genetic diversity in the TSWV genome. Genetic diversity is measured in terms of average pairwise distances between sequences. Within-host diversity values were averaged across all samples in the *Alternating* line. Species-level diversity was measured using a globally representative sample of publicly available TSWV samples (Supp. Table 1). Local fluctuations in diversity were smoothed by taking a running average across a 100 bp sliding window.

the virus is passed between plants and thrips. Of the 15 variants that are enriched between plants and thrips, five variants are consistently enriched across all three experimental lines (Fig. 7). Four of these five are enriched in plants versus thrips while only one variant (G_C E362D) was enriched in thrips versus plants.

Fitness effects between hosts and between plants and thrips. To more precisely quantify the fitness effects of variants in different hosts, the time-resolved deep sequencing data was used to estimate the growth rate of variants within both host plants and thrips. The growth rate of each variant can then be used as a proxy for the fitness effect of a variant relative to the reference type. The relative fitness effects we report here are the difference in growth rates between a variant and the reference type, such that a neutral variant will have a fitness of zero and a beneficial variant will have a positive fitness effect. We note, however, that these estimated fitness effects are potentially confounded by variants being linked to other mutations on the same viral genotype/haplotype. Nevertheless, quantifying fitness effects can provide general insights into how selection pressures vary between hosts.

First, the fitness effect of each variant was estimated in both *Emilia* and *Datura*. The joint distribution of fitness effects shows that only a small fraction of variants are estimated to be unconditionally deleterious (Fig. 8A). This result is likely due to an ascertainment bias against deleterious variants. Most strongly deleterious mutations were likely excluded since low frequency variants (<3%) were not considered and a variant must persist in the viral population between two or more time points in order to estimate its growth rate (see Methods). However, there are several mutations that are neutral or beneficial in one host but deleterious in the other, indicative of antagonistic pleiotropy. Some of the same amino acid variants seen to have different variant frequencies in *Emilia* versus *Datura* in Fig. 6A are again estimated to have fitness differences between hosts here (Table 1). For example, the N22S variant in the movement protein NSm has $\sim 3\times$ higher fitness in *Datura* than *Emilia*, although the NSm V17G appears only slightly deleterious in both hosts. More surprisingly, there is an overall positive correlation



Figure 6. Evolutionary dynamics of individual amino acid variants in the *Alternating* line. (A) Time series of variants enriched in either *Emilia* or *Datura* relative to the other plant host. (B) Time series of variants enriched in either plants or thrips. A variant was considered to be enriched if its average frequency in one host was > 5% higher than in the alternate host, where the average was computed over all sampled time points in each host. Sampling time points are colored by host; green = *Emilia*, black = thrips and blue = *Datura*.

(Pearson correlation coefficient $\rho = 0.12$) between fitness effects across hosts, suggesting that positive rather than antagonistic pleiotropy predominates between plant hosts.

The joint distribution of fitness effects between plants and thrips is shown in Fig. 8B. While many variants have beneficial fitness effects in both hosts, fitness effects are largely uncorrelated between plants and thrips ($\rho = 0.003$). There are also a rather large number of AAVs that are beneficial in plants but deleterious in thrips, indicating potential fitness conflicts between plants and thrips in certain regions of the genome. Several of these AAVs are strongly deleterious in thrips and occur in RdRp (Table 1), and correspond to some of the same AAVs that fluctuate in frequency between plants and thrips in Figs. 6 and 7, including RdRp variants R290S, N289S, R495Q, K566R, K863R and A1799T. To get a better sense of where these fitness conflicts occur, fitness differences between plants and thrips were mapped onto the TSWV genome (Fig. 9). Many of the largest fitness differences between plants and thrips are localized on RdRp, and to a lesser extent NSs and G_N/G_C . Fitness differences between *Emilia* and *Datura* are distributed over the entire genome, although there are several localized at the N-terminus of NSm.

Discussion

Experimental evolution studies have become a standard approach in virology to investigate how viruses adapt to novel host environments^{17,22–24}. A large number of these studies have focused on arboviruses or other multi-host pathogens, as virologists have long been interested in how viruses overcome the constraints imposed by

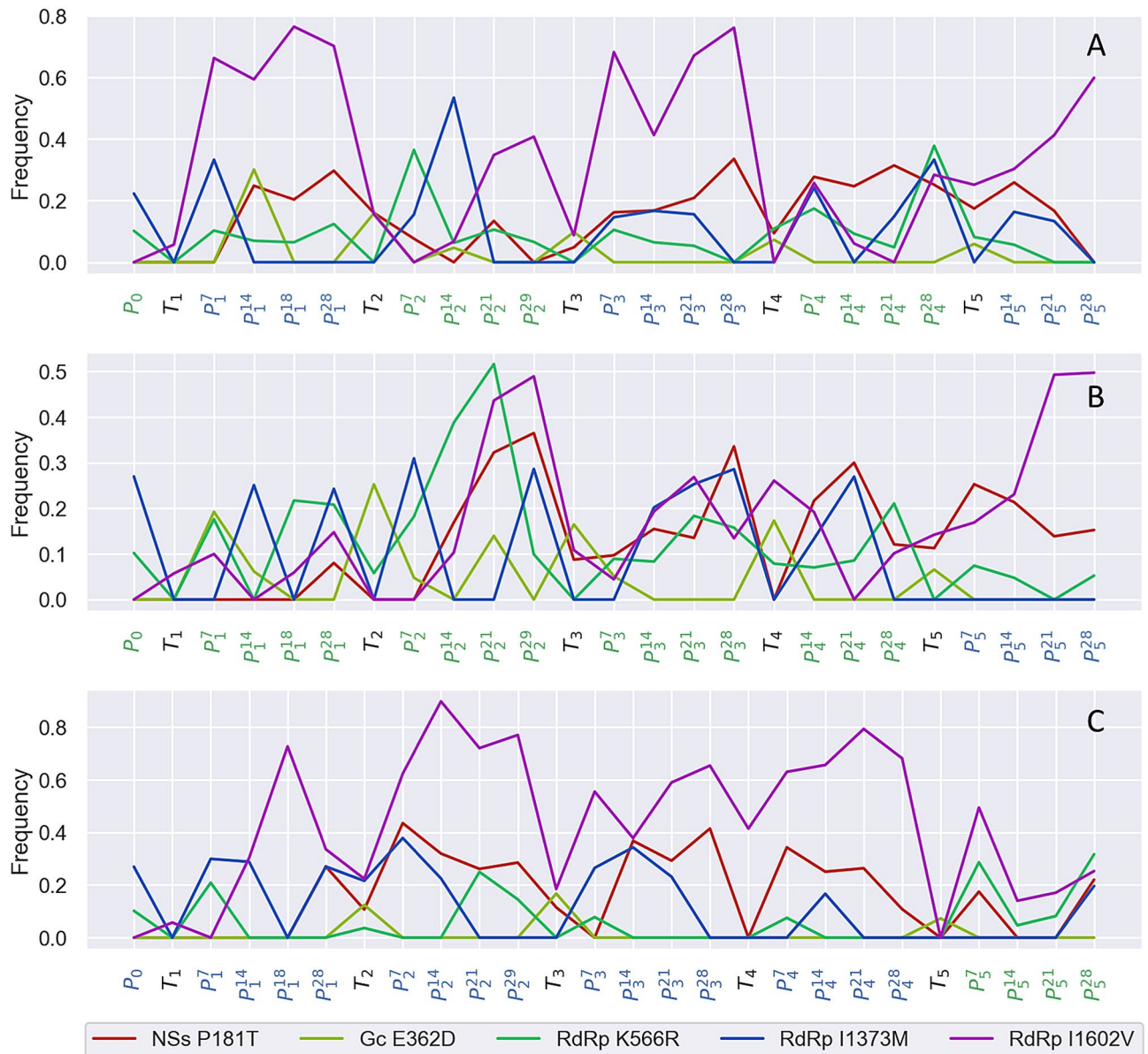


Figure 7. Evolutionary dynamics of individual amino acid variants consistently enriched in plants or thrips across all three lines. The three lines are: (A) *Alternating*, (B) *Emilia* and (C) *Datura*.

alternating between hosts. These experimental evolution studies have often yielded results that challenge long-held assumptions in evolutionary theory. For example, while evolutionary theory largely assumes that performance or fitness tradeoffs will limit simultaneous adaptation to more than one environment, experimental studies have repeatedly demonstrated that viruses can adapt to new hosts with little or no fitness cost in alternate hosts^{19,21,24}. Likewise, while antagonistic pleiotropy has long been assumed to underlie fitness tradeoffs between environments, recent experimental work has shown that mutations often have positive pleiotropic effects between hosts^{25,26}, especially in hosts that are phylogenetically closely related²⁷. In light of this work, we sought to explore how an extreme generalist like TSWV adapts to alternate plant hosts and thrips.

Deep sequencing TSWV populations revealed that much of the genetic diversity present in the initial founder population persisted for multiple passage cycles, with little evidence for bottlenecks in diversity at transmission events. Genetic diversity tended to increase when the virus was passed through thrips but decrease over the course of a single infection in plants; although this was more evident in *Datura* than *Emilia*. This loss of diversity may be due to the fact that leaves sampled at later time points were more distal from the site of infection. Although TSWV moves systemically through plants, only a subset of the viral population may undergo long-distance transport to new leaves, resulting in distinct founder populations with lower diversity.

Although one might expect to see similar bottlenecks within thrips as the virus must traverse through the midgut to the salivary glands before transmission can occur²⁸, we found that genetic diversity tended to increase slightly in thrips, although not to a statistically significant level. We may have failed to detect bottlenecks in thrips as multiple insects were sampled and then pooled at a single time point to obtain enough RNA for sequencing.

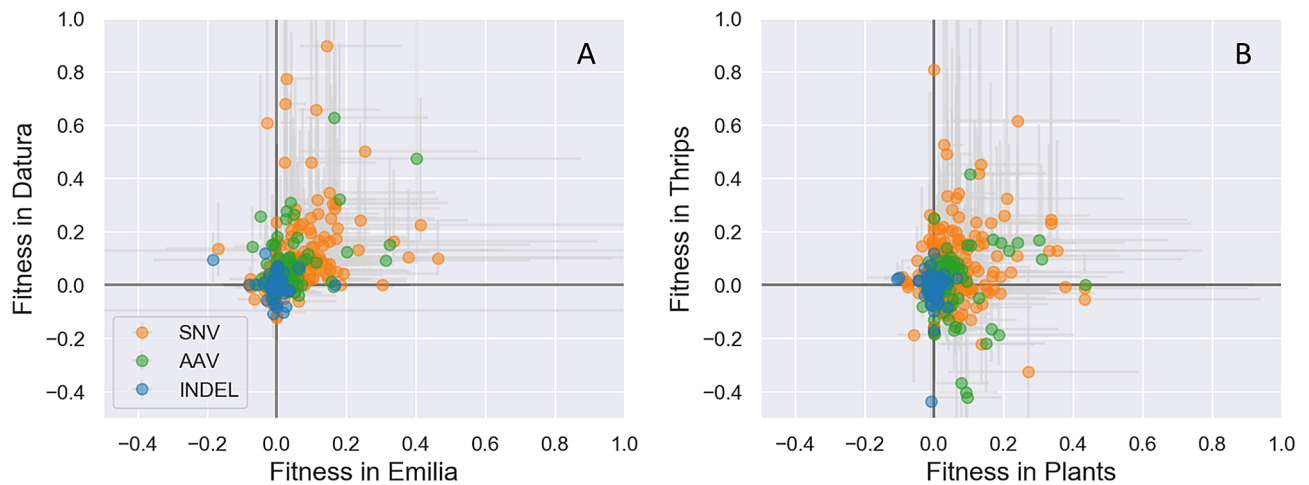


Figure 8. The joint distribution of fitness effects in *Emilia* versus *Datura* (A) and in plants versus thrips (B). The fitness effects of individual variants were estimated based on their relative growth rate in each host. The circles mark the median and the light grey lines indicate the 95% credible intervals of the inferred posterior distribution of the fitness effect in each host. Fitness estimates are colored according to the type of mutation: orange = single nucleotide variant; green = amino acid variant; blue = indel.

Protein	AAV	Emilia	Datura	Plant	Thrips
NSs	I79M	0.004	0.104	0.044	-0.084
NSs	P181T	-0.004	0.009	0.002	-0.186
NSm	V17G	-0.014	-0.028	-0.018	-0.036
NSm	N22S	0.036	0.091	0.053	0.063
Gc	E362D	-0.228	NA	-0.228	NA
Gn	V141F	0.0	0.0	0.0	0.0
RdRp	I226S	0.048	0.0	0.0	0.0
RdRp	R290S	0.025	0.094	0.061	-0.159
RdRp	N389S	0.084	0.097	0.095	-0.421
RdRp	R495Q	0.075	0.091	0.091	-0.402
RdRp	K566R	0.0	0.006	0.0	-0.178
RdRp	N813D	0.0	NA	0.0	0.0
RdRp	K863R	0.0451	0.078	0.078	-0.367
RdRp	I1373M	NA	NA	NA	NA
RdRp	I1602V	0.0	0.0	0.0	0.0
RdRp	A1799T	0.048	0.159	0.057	-0.171
RdRp	C2148Y	-0.069	0.144	0.086	NA

Table 1. Amino acid variant fitness estimates in plant hosts and thrips. Fitness values are the estimated posterior median fitness effects relative to the reference amino acid.

However, thrips larvae were inoculated by feeding on a single infected leaf from the source plant, such that the viral diversity transferred to thrips should reflect the diversity in a single leaf. This suggests that viral diversity may actively increase in thrips, and we note that viral diversity was previously shown to increase in thrips (J. Brown, unpublished), consistent with our results here. Thus, unlike in other plant viruses where vector-borne transmission leads to extreme bottlenecks in viral population sizes and genetic diversity²⁹, the ability of TSWV to replicate persistently in thrips may largely preserve diversity.

Several amino acid variants rapidly fluctuated in frequency between plant hosts and vectors. The evolutionary dynamics of these variants may provide clues into the selection pressures imposed by different hosts and how the virus adapts to them. In plants, different amino acid variants in the NSm protein were found to be differentially enriched in either *Emilia* and *Datura*. NSm functions as a viral movement protein that is necessary for both short and long distance movement¹³, and previous reports have implicated NSm in host range determination³⁰. Functional analysis in tobacco plants indicated that amino acid mutations in the N-terminus of NSm abolish tubule formation and cell-to-cell movement, but not long distance movement³¹. Interestingly, the first 50 amino acids of the N-terminus are hypervariable at the species level^{30,31} and hypervariable within hosts, as shown here. Furthermore, we found amino acid variants V17G and N22S are differentially enriched in *Emilia* versus *Datura*

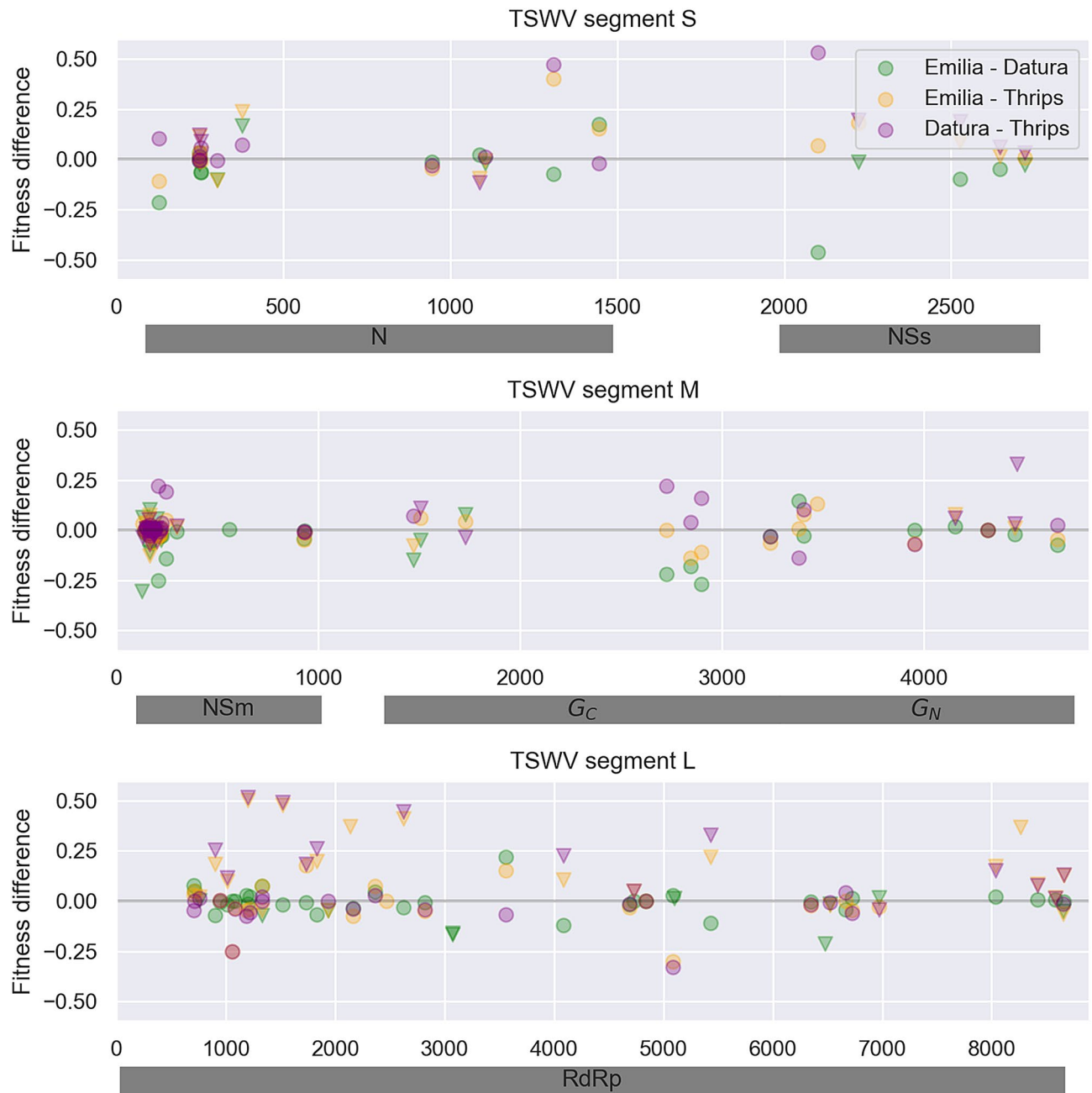


Figure 9. Fitness differences between hosts mapped onto the TSWV genome. Fitness differences between hosts are quantified as the relative growth rate of a variant in the first host minus the growth rate in the second host, such that positive values indicate higher fitness in the first host in each pair. Inverted triangles indicate instances of sign pleiotropy in which the variant had a positive fitness effect in one host and a negative effect in the other host.

(Fig. 6), all of which suggests that host-specific changes in NSm may be required for TSWV to move efficiently through different plants.

Several amino acid variants were also found to be enriched in plants versus thrips in a consistent manner across lines. These variants include single amino acid mutations in the silencing suppressor NSs and the glycoprotein G_C , as well as several in the viral RNA-dependent RNA polymerase (RdRp). As NSs is involved in suppressing RNA silencing in both plants and thrips³², it is perhaps not surprising that different variants may be favored in plants versus thrips. Consistent with this, a recent analysis of genomic diversity among Orthotospoviruses showed that NSs contained the most codon sites under positive selection among protein coding genes³³. In the glycoprotein G_C , the amino acid variant E362D appears to be very strongly selected for in thrips but only observed at very low frequencies in plants (Fig. 7). Arboviruses have repeatedly been found to adapt to their insect vectors through single amino acid mutations in viral glycoproteins^{34,35}, and in TSWV, G_C likely acts as a viral fusion protein that along with G_N is essential for transmission in thrips³⁶. But it is less clear why the

E362D mutation would be so strongly selected against in plants, since G_N and G_C are thought to be dispensable in plants^{9,37}. Finally, several amino acid variants in RdRp appear to be beneficial in plants but strongly deleterious in thrips. Based on structural homology to other bunyavirus RdRp proteins, one of these mutations occurs in the endonuclease domain involved in host mRNA cap-snatching and two others occur within the central catalytic domain responsible for RNA synthesis^{38,39}. Both of these domains are under predominantly purifying selection at the TSWV species level³⁹. We speculate that alternative amino acid variants are required to optimize replication and transcription in plants versus thrips due to interactions with different, host-specific cellular factors. Such host-specific interactions between viral polymerases and cellular factors have been shown to be a key determinant of host adaptation in other RNA viruses^{40,41}.

We therefore found some evidence for antagonistic pleiotropy between plants and thrips, and to a lesser extent between *Emilia* and *Datura*, which may place constraints on TSWV's ability to simultaneously adapt to multiple plant hosts and thrips. Nevertheless, beyond a few sites of apparent conflict in the genome, the fitness effects estimated between hosts show that positive pleiotropy is common. Consistent with the positive correlation in fitness effects between plant hosts, we did not see major changes in the viral population after passaging the virus back to the alternate host in the final passage of the *Emilia* and *Datura* only lines. It is therefore tempting to speculate that this tendency towards positive pleiotropy endows TSWV with the ability to find beneficial mutations in new hosts without a concomitant loss of fitness in previous hosts, allowing TSWV to rapidly expand its host range. Moreover, even if antagonistic pleiotropy does arise at particular sites in the genome, the ability to maintain extensive genetic diversity between transmission events may allow for variants that are deleterious in the current host to be maintained, possibly at low frequency, long enough to be transmitted to another host in which the variant may become beneficial. Thus, the ability to persistently replicate and thereby avoid a narrow transmission bottleneck may allow TSWV to more readily adapt to new hosts than other viruses.

While we were able to estimate the relative fitness effects of variants between hosts, one serious limitation of our study is that the absolute fitness of viral populations between hosts were not directly measured. Our study also lacked true biological replicates of each experimental line, limiting our ability to draw conclusions about the repeatability of the evolutionary changes we observed in each host. However, most of our major results are replicated across multiple individuals and three independent lines. General patterns of diversity and divergence were consistent across all three lines. Inference of fitness effects in each host were based on variant growth rates in multiple different plants of the same species. Furthermore, many of the amino acid variants found to fluctuate in frequency in plants versus thrips were consistent across all three lines. These results suggest that our main findings, including estimates of fitness effects and the sign of pleiotropy, are highly repeatable.

Furthermore, the present study only considered fitness in alternate plant hosts and not between different thrips species. Like many other plant viruses, TSWV is considered to be a plant host generalist but a vector specialist⁴². Indeed, only 9 of more than 7000 described thrips species are known to be competent vectors of TSWV^{5,28}, and particular genetic isolates of TSWV appear to be intimately adapted to local thrips populations⁴³. Future work by our group will therefore look at differences in absolute fitness between hosts and whether it is more difficult for TSWV to adapt to new plant hosts or new vector species.

Materials and methods

Experimental passaging. A TSWV-infected tomato plant (var. Celebrity) was collected from a field near Apex, North Carolina in August of 2018. The fruit tissue was immediately used for mechanical inoculation onto a 20 day-old *Emilia sonchifolia* plant (referred to as P0 above). We did not screen the initially collected tissue for the presence of other viruses, although no other thrips-vectored viruses have been reported to infect tomato in North Carolina. Both the fruit and leaf tissue from the same plant were preserved at $-80\text{ }^{\circ}\text{C}$ for later RNA extraction and sequencing. The mechanically inoculated *Emilia* plant was used as source material for viral passaging via thrips and maintained under greenhouse conditions within an insect cage.

Western flower thrips (*Frankliniella occidentalis*) were used as the vector species for all passages. Thrips were obtained from a laboratory colony maintained at $27\text{ }^{\circ}\text{C}$, ca. 55% RH and under continuous light on insecticide-free cabbage (*Brassica oleracea* var. *capitata* L.) foliage in 0.35 L plastic food containers (Fabri-Kal Corp., Kalamazoo, MI) ventilated with thrips-proof screen (81×81 mesh; Bioquip Products, Inc., Rancho Dominguez, CA). At each transmission cycle, approximately 100 adult females from the colony were confined in a rearing container and allowed to oviposit for 24 h through a stretched Parafilm membrane into a 3% sucrose solution contained in a 9 cm Petri dish. Following oviposition, the eggs were collected by filtering the sucrose solution through filter paper and rinsing any eggs attached to the membrane onto the filter paper with distilled water. To obtain viruliferous adults, the filter paper was positioned on top of a single excised TSWV-infected leaf from the designated source plant such that the eggs were sandwiched between the filter paper and the abaxial surface of the infected leaf, which was maintained on moistened filter paper in a sealed rearing container at $27\text{ }^{\circ}\text{C}$. After four days all eggs had hatched and the larvae were shaken onto an excised upper leaf from a non-infected plant of either *Emilia* or *Datura*, depending on the treatment, where they completed development to adults.

At each transmission cycle, groups of eight viruliferous adults (3–7 days post-eclosion) were aspirated onto each *Emilia* or *Datura* seedling (three to four-true leaves). Seedlings were grown separately in 296 ml plastic cups (Solo Cup Company, Lake Forest, IL, USA) with a 25 mm diameter fine mesh screen on the bottom. Thrips were contained on the seedlings by inverting a plastic cup with a screened bottom over the seedling and sealing it to the cup containing the plant using Parafilm. After approximately 48 h, each seedling was sprayed with spinetoram (Radiant SC; Corteva Agriscience, Indianapolis, USA) to kill the thrips. TSWV infected plants were maintained in a growth chamber under a 16-h photoperiod, $27\text{ }^{\circ}\text{C}$ and ca. 50% relative humidity for approximately one month after inoculation.

Name	Sequence [5' to 3']
TSWV_S_R	AGAGCAATTGTGCAA
TSWV_M_R	AGAGCAATCAGTGCAAAC
TSWV_L_4474R	GAGTGCACAATCCATCTAG
TSWV_L_R	AGAGCAATCAGGTACAAC

Table 2. Primers used for segment-specific cDNA synthesis.

Name	Sequence [5' to 3']
TSWV_SF_tail	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGCAAT TGTGTCNN
TSWV_SR_tail	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGCAAT TGTGTCAA
TSWV_MF_tail	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGCAAT CAGTGCCNN
TSWV_MR_tail	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGCAAT CAGTGCAAAC
TSWV_LF_tail	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGCAAT CAGGTAACAACG
TSWV_L4474R_tail	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGCAAT CATCCATCTAG
TSWV_L4120F_tail	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAAT TCGAAATGGTC
TSWV_LR_tail	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGCAAT CAGGTACAAC

Table 3. Primers used for segment-specific PCR amplification. 5' end of primers are adapter sequences specific to Illumina DNA Flex Nextera kit. Binding portion of primers denoted in bold.

Three separate experimental lines were developed in which the virus was either alternated (Line 1) between plant hosts (*Emilia sonchifolia* and *Datura stramonium*) or maintained on *Emilia* (Line 2) or *Datura* (Line 3). Approximately 21 days after inoculation, tissue was collected from the plant lines and used to initiate the next passing round by feeding to western flower thrips. At the final passage cycle, the single host lines were passaged back to the alternate plant species.

Sample collection. Plant lines were sampled at four time points following virus transmission (at approximately 7, 14, 21, and 28 days post-infection). The time of infection was defined as when the viruliferous thrips were rendered inactive on the host plant. A sterilized 8-mm diameter cork borer was used to collect tissue from the three most recently emerged leaves. Five leaf disks were sampled in total: 2 disks from the two larger leaves and 1 disk from the smallest leaf. Disks from each plant were pooled and immediately frozen at -80°C for later RNA extraction.

At each transmission cycle, approximately 40 thrips from the cohort of viruliferous adults used to inoculate test plants were collected into a 1.5 ml microcentrifuge tube at the time that transmission was initiated. These thrips were immediately frozen at -80°C for later RNA extraction.

Total RNA extraction. For plant tissue, five 8 mm diameter leaf disks were placed into a 1.5 ml microcentrifuge with three-3 mm Pyrex glass beads (Corning). Sample tubes were then placed in liquid nitrogen followed by bead beating on the Silamat S6 (Ivoclar Vivodent) for 20 s. Contrastingly, 40 thrips in a 1.5 ml microcentrifuge were placed into liquid nitrogen then ground via motorized pestle.

Following tissue destruction, TRI Reagent (Zymo Research) was immediately added and vortexed on high with 600 μl TRI Reagent added to plant tissue samples and 300 μl added to thrips samples. Samples were incubated for 5 min in TRI reagent at room temperature before following manufacturer's protocol for RNA extraction kits. For plant tissue, the Direct-zol RNA MiniPrep Plus kit (Zymo Research) was utilized and RNA resuspended in 60 μl . For thrips, the Direct-zol MicroPrep kit (Zymo Research) with resuspension in 15 μl . RNA quality was assessed via electrophoresis and on a Nanodrop 1000. All RNA samples were stored at -80°C .

cDNA synthesis. For synthesis of cDNA from total RNA extracted from plant and thrips tissue samples, approximately 500 ng of total RNA was used for a 10 μl cDNA synthesis with ProtoScript II (NEB). 15 μM of the appropriate strand-specific primer (IDT; Coralville, IA) (Table 2), 10 mM dNTP, total RNA, and sterile water (up to 5 μl total volume) were incubated at 65°C for 5 min then placed on ice. Next, 2 μl 5 \times ProtoScript II Buffer, 0.1 M DTT, 4 U RNase Inhibitor, 100 U ProtoScript II RT, and 1.4 μl sterile water were added. The samples were first incubated at 25°C for 5 min, 42°C for 1 h, then 65°C for 20 min before storing at -20°C . All passaging samples were duplicated for two independent sequencing replicates beginning at the cDNA step.

PCR. PCR was used to amplify viral cDNAs to enrich viral representation. 50 μl PCR reactions were set up with approximately 1 μg cDNA and Phusion High-Fidelity DNA Polymerase (NEB) was used. The manufacturer's protocol was followed and the addition of 1.5 μl DMSO was included. Primers (IDT) utilized were genome segment-specific (Table 3), and the 5' end included a tail sequence to preferentially bind the Illumina Nextera

DNA Flex adapter (to increase terminal end coverage of viral genome segments). PCR reactions were amplified with a BioRad C1000 Thermocycler on the following settings: 98 °C–30 s; 30X: (98 °C–10 s, 52 °C–30 s, 72 °C–4 m); 72 °C–5 m; infinite hold at 12 °C. Expected PCR product sizes were verified via electrophoresis before proceeding to sample purification step.

Sample purification. Products for the four genome fragments were combined (approximately 200 µl volume when pooled). Combined PCR products were purified with gDNA Clean and Concentrator kit-10 (Zymo Research) via two sequential elutions of 10 µl each (20 µl total volume). Sample concentrations were measured on a Nanodrop 1000.

Illumina library preparation and sequencing. For deep sequencing, 500 ng of the purified and pooled PCR genome amplicons were prepared for sequencing via Nextera DNA Flex Kit (Illumina, #20018705) with 96 indexes (Illumina, #20018708) according to the manufacturer's protocol. Library quality was analyzed via Agilent 2200 TapeStation at the NCSU GSL. All sequencing was performed on an Illumina MiSeq instrument at North Carolina State University's Genomic Science Laboratory to obtain paired end reads of approximately 300 base pairs.

Sequence analysis. Paired end reads were obtained from two sequencing replicates at each sampling time point from the MiSeq runs. After trimming adapter sequences from the raw reads, sequences were mapped to the TSWV reference genome assembly on GenBank with accession number GCA_000854725.1⁴⁴ using Bowtie2⁴⁵. In order to minimize the potential for misalignment due to using a divergent reference genome, we then assembled a new consensus genome sequence from our TF2 field-collected isolate. All paired end reads from our passaging experiments were then realigned against the TF2 reference genome using the 'sensitive-local' preset parameters in Bowtie2. Alignments for each sample were converted into SAM and BAM files for further processing using SAMtools⁴⁶.

To call genetic variants in each viral population, the *mpileup* routine in SAMtools was used to identify single nucleotide variants (SNVs) and indels relative to the TF2 reference in both paired sequencing replicates from each sample. Variants at primer binding sites were first filtered out. The remaining variants were subsequently filtered in *ivar* using the criteria proposed by the authors⁴⁷. Using their criteria, a variant needed to be present at a frequency of at least 0.03 and obtain an Illumina/Phred quality score of 20 (i.e. a 0.01 sequencing error probability) in both paired sequencing replicates in order to be considered a true variant. Thus, even for sites with a relatively low coverage (< 100×), the probability of a variant caused by a sequencing error reaching our threshold frequency of 0.03 is extremely unlikely, with a probability of 10⁻⁶. Furthermore, while it is possible that an error introduced at the RT or PCR stage could reach a frequency of 0.03, it is extremely unlikely that such an error would occur in both sequencing replicates independently. Our variant calling strategy therefore ensures that all called variants were actually present in the viral population.

We used the frequency of SNVs at each site in the genome to compute diversity within and divergence between viral populations. Diversity was computed as the average pairwise distance between viral sequences in the same population. Divergence was computed as the average pairwise distance between viral sequences in two different populations. In both cases, the pairwise distance *D* between viral sequences at each site was computed using the frequency *q_i* of each variant *i* present at the site:

$$D = \sum_i q_i (1 - q_i).$$

From the variants called at each individual sampling time, we created a master list tracking how the frequency of each variant changed over time. We also categorized SNVs as either amino acid variants (AAVs) based on whether their translated sequence was predicted to cause a nonsynonymous substitution in the reference sequence.

Global TSWV diversity. Within-host genetic diversity was compared to species-level diversity among a global collection of TSWV isolates sampled from different hosts. For this analysis, the same set of sequences as Lian et al.⁴⁸ was used which included 53 S, 57 M and 17 L full-length segment sequences. To this collection, we added 23 L segment sequences that have been deposited in GenBank since 2013. GenBank accession numbers for all sequences are provided in Supp. Table 1.

Estimating fitness effects. The fitness effect of each variant was estimated based on changes in variant frequencies over time within hosts. Following the strategy of Illingworth et al.^{49,50}, it is assumed that each variant's frequency changes over time according to a model of deterministic exponential growth:

$$q_i(t_{k+1}) = q_i(t_k) \exp(\sigma_{i,h} \Delta_{k,k+1}) / \sum_j q_j(t_k) \exp(\sigma_{j,h} \Delta_{k,k+1}) \quad (1)$$

Here, *q_i*(*t_{k+1}*) is the predicted frequency of variant *i* at time *t_{k+1}* given its observed frequency *q_i*(*t_k*) at time *t_k*. The term $\Delta_{k,k+1}$ is the time elapsed between a pair of sequential samples taken at times *t_k* and *t_{k+1}*. The host-specific growth rate of variant *i* in host *h* is given by $\sigma_{i,h}$. Note that the growth rate of each variant is estimated relative to the reference type since absolute growth rates cannot be estimated because only changes in variant frequencies

are observed through time. Fitness effects are reported as the difference in growth rates between a variant and the reference type.

The growth rate $\sigma_{i,h}$ therefore reflects variant i 's relative fitness in a particular host, and we seek to estimate these values from observed frequency changes over time. Let n_{k+1} be a vector holding the number of observed sequence reads representing each variant at time t_{k+1} . Given the expected variant frequencies $q(t_{k+1})$ predicted under the exponential growth model, we compute the likelihood of observing n_{k+1} assuming a multinomial sampling process:

$$L(n_{k+1}|q(t_{k+1})) = \text{Multinom}(n_{k+1}|N_{k+1}, q(t_{k+1})), \quad (2)$$

where $N_{k+1} = \sum_i n_i$, the total depth of coverage at the site of variant i .

To obtain a maximum likelihood estimate for the fitness effects, we can then find the value of $\sigma_{i,h}$ that maximizes the product of the individual multinomial likelihood terms (Eq. 2) across all pairs of time points k and $k+1$ for which we have observed variant frequencies, using Eq. (1) above to compute $q_i(t_{k+1})$ whenever we need to evaluate the likelihood function. All three lines were used to estimate fitness differences between plant and thrips. All samples from *Emilia* in lines one and two were used to estimate fitness in *Emilia* and likewise, all samples from *Datura* in lines one and three were used to estimate fitness in *Datura*. We exclude all pairs of time points where the initial frequency of the variant or reference allele was zero at time t_k because in this case Eq. (2) is not defined. We also exclude all pairs of time points where $N_{k+1} < 100$ to minimize variability in our estimates due to a low total depth of coverage at a given sampled time point.

Maximum likelihood estimates were obtained by numerically optimizing the likelihood with SciPy's *minimize* function using Sequential Least Squares Programming. To evaluate the uncertainty surrounding these estimates, we also estimated the Bayesian posterior distribution $p(\sigma_{i,h})$ of each fitness effect:

$$p(\sigma_{i,h}) \sim \prod_k L(n_{k+1}|q(t_{k+1}))g(\sigma_{i,h}), \quad (3)$$

where the first term on the right hand side is the product of the multinomial likelihoods across all paired time points and $g(\sigma_{i,h})$ is the prior distribution. An uninformative, uniform prior was specified for all fitness effects. A Metropolis–Hastings MCMC sampler was then used to sample parameter values from the posterior distribution in (3), from which the posterior median and 95% credible intervals were computed.

We tested the statistical performance of our fitness inference methods using simulated time series of variant frequencies. Random fitness effects were drawn uniformly from between -0.2 and 0.2 , and then the frequency of each variant was simulated forward through time using Eq. (1) for 10 time steps. At each time step, a random number of sequence reads n_{k+1} were drawn from a multinomial distribution with probabilities proportional to the simulated variant frequencies. We then used the MCMC sampler to estimate the posterior median and 95% credible intervals of the fitness effects for 100 simulated time series. The estimated fitness effects were highly correlated with the true fitness effects with no detectable bias and good posterior coverage (Supp. Figure 4).

Data availability

Python scripts automating the alignment, variant calling and fitness estimation procedures are available at <https://github.com/davidrasm/DeepTSWVSeq>. A master list providing the frequency of each variant through time in all three lines is also available in this repository. Raw sequence reads have been submitted to NCBI's Sequence Read Archive as a single BioProject with SRA accession number PRJNA627851.

Received: 9 June 2020; Accepted: 4 September 2020

Published online: 25 September 2020

References

1. Woolhouse, M. E., Taylor, L. H. & Haydon, D. T. Population biology of multihost pathogens. *Science* **292**, 1109–1112 (2001).
2. Dawson, W. O. & Hilf, M. E. Host-range determinants of plant viruses. *Annu. Rev. Plant Biol.* **43**(1), 527–555 (1992).
3. Moury, B., Fabre, F., Hébrard, E. & Froissart, R. Determinants of host species range in plant viruses. *J. Gen. Virol.* **98**, 862–873 (2017).
4. Nault, L. R. Arthropod transmission of plant viruses: A new synthesis. *Ann. Entomol. Soc. Am.* **90**(5), 521–541 (1997).
5. Jones, D. R. Plant viruses transmitted by thrips. *Eur. J. Plant Pathol.* **113**(2), 119–157 (2005).
6. Gilbertson, R. L., Batuman, O., Webster, C. G. & Adkins, S. Role of the insect super-vectors *Bemisia tabaci* and *Frankliniella occidentalis* in the emergence and global spread of plant viruses. *Annu. Rev. Virol.* **2**, 67–93 (2015).
7. Abudurexiti, A. *et al.* Taxonomy of the order Bunyavirales: Update 2019. *Arch. Virol.* **164**, 1949–1965 (2019).
8. Parrella, G., Gognalons, P., Gebre-Selassie, K., Vovlas, C. & Marchoux, G. An update of the host range of tomato spotted wilt virus. *J. Plant Pathol.* **85**, 227–264 (2003).
9. Oliver, J. E. & Whitfield, A. E. The genus Tospovirus: Emerging bunyaviruses that threaten food security. *Annu. Rev. Virol.* **3**, 101–124 (2016).
10. Pappu, H. R., Jones, R. A. C. & Jain, R. K. Global status of tospovirus epidemics in diverse cropping systems: Successes achieved and challenges ahead. *Virus Res.* **141**, 219–236 (2009).
11. Whitfield, A. E., Ullman, D. E. & German, T. L. Tospovirus-thrips interactions. *Annu. Rev. Phytopathol.* **43**, 459–489 (2005).
12. Groves, R. L., Walgenbach, J. F., Moyer, J. W. & Kennedy, G. G. The role of weed hosts and tobacco thrips, *Frankliniella fusca*, in the epidemiology of tomato spotted wilt virus. *Plant Dis.* **86**, 573–582 (2002).
13. Lewandowski, D. J. & Adkins, S. The tubule-forming NSm protein from tomato spotted wilt virus complements cell-to-cell and long-distance movement of tobacco mosaic virus hybrids. *Virology* **342**, 26–37 (2005).
14. Elena, S. F., Agudelo-Romero, P. & Lalić, J. The evolution of viruses in multi-host fitness landscapes. *Open Virol. J.* **3**, 1–6 (2009).
15. Garcia-Arenal, F. & Fraile, A. Trade-offs in host range evolution of plant viruses. *Plant Pathol.* **62**, 2–9 (2013).

16. Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J. & Jiggins, F. M. The evolution and genetics of virus host shifts. *PLoS Pathog.* **10**, e1004395 (2014).
17. Duffy, S., Turner, P. E. & Burch, C. L. Pleiotropic costs of niche expansion in the RNA bacteriophage phi 6. *Genetics* **172**, 751–757 (2006).
18. Elena, S. F. *et al.* The evolutionary genetics of emerging plant RNA viruses. *Mol. Plant. Microbe. Interact.* **24**, 287–293 (2011).
19. Weaver, S. C., Brault, A. C., Kang, W. & Holland, J. J. Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J. Virol.* **73**, 4316–4326 (1999).
20. Kassen, R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**, 173–190 (2002).
21. Bedhomme, S., Lafforgue, G. & Elena, S. F. Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Mol. Biol. Evol.* **29**, 1481–1492 (2012).
22. Cuevas, J. M., Willemsen, A., Hillung, J., Zwart, M. P. & Elena, S. F. Temporal dynamics of intrahost molecular evolution for a plant RNA virus. *Mol. Biol. Evol.* **32**, 1132–1147 (2015).
23. Elena, S. F. *et al.* Experimental evolution of plant RNA viruses. *Heredity* **100**, 478–483 (2008).
24. Turner, P. E. & Elena, S. F. Cost of host radiation in an RNA virus. *Genetics* **156**, 1465–1470 (2000).
25. McGee, L. W. *et al.* Payoffs, not tradeoffs, in the adaptation of a virus to ostensibly conflicting selective pressures. *PLoS Genet.* **10**, e1004611 (2014).
26. Moreno-Pérez, M. G., García-Luque, I. & Fraile, A. Mutations that determine resistance breaking in a plant RNA virus have pleiotropic effects on its fitness that depend on the host environment and on the type, single or *J. Virol.* **90**(20), 9128–9137 (2016).
27. Lalić, J., Cuevas, J. M. & Elena, S. F. Effect of host species on the distribution of mutational fitness effects for an RNA virus. *PLoS Genet.* **7**, e1002378 (2011).
28. Chen, Y., Dessau, M., Rotenberg, D., Rasmussen, D. A. & Whitfield, A. E. Entry of bunyaviruses into plants and vectors. *Adv. Virus Res.* **104**, 65–96 (2019).
29. Gutiérrez, S., Michalakis, Y. & Blanc, S. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr. Opin. Virol.* **2**, 546–555 (2012).
30. Silva, M. S. *et al.* Sequence diversity of NS M movement protein of tospoviruses. *Adv. Virol.* **146**, 1267–1281 (2001).
31. Li, W., Lewandowski, D. J., Hilf, M. E. & Adkins, S. Identification of domains of the tomato spotted wilt virus NSm protein involved in tubule formation, movement and symptomatology. *Virology* **390**, 110–121 (2009).
32. Hedil, M., Sterken, M. G., de Ronde, D., Lohuis, D. & Kormelink, R. Analysis of tospovirus NSs proteins in suppression of systemic silencing. *PLoS ONE* **10**, e0134517 (2015).
33. Nigam, D. & Garcia-Ruiz, H. Variation profile of the orthotospovirus genome. *Pathogens* **9**, 521 (2020).
34. Brault, A. C. *et al.* Venezuelan equine encephalitis emergence: Enhanced vector infection from a single amino acid substitution in the envelope glycoprotein. *Proc. Natl. Acad. Sci. USA.* **101**, 11344–11349 (2004).
35. Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E. & Higgs, S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog.* **3**, e201 (2007).
36. Sin, S.-H., McNulty, B. C., Kennedy, G. G. & Moyer, J. W. Viral genetic determinants for thrips transmission of tomato spotted wilt virus. *Proc. Natl. Acad. Sci. USA.* **102**, 5168–5173 (2005).
37. Nagata, T., Inoue-Nagata, A. K., Prins, M., Goldbach, R. & Peters, D. Impeded Thrips transmission of defective tomato spotted wilt virus isolates. *Phytopathology* **90**, 454–459 (2000).
38. Reguera, J., Weber, F. & Cusack, S. Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS Pathog.* **6**, e1001101 (2010).
39. Terret-Welter, Z., Bonnet, G., Moury, B. & Gallois, J.-L. Analysis of tomato spotted wilt virus RNA-dependent RNA polymerase adaptive evolution and constrained domains using homology protein structure modelling. *J. Gen. Virol.* **101**, 334–346 (2020).
40. Long, J. S. *et al.* Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature* **529**, 101–104 (2016).
41. Boivin, S., Cusack, S., Ruigrok, R. W. H. & Hart, D. J. Influenza A virus polymerase: Structural insights into replication and host adaptation mechanisms. *J. Biol. Chem.* **285**, 28411–28417 (2010).
42. Lefeuvre, P. *et al.* Evolution and ecology of plant viruses. *Nat. Rev. Microbiol.* **17**, 632–644 (2019).
43. Jacobson, A. L., Booth, W., Vargo, E. L. & Kennedy, G. G. Thrips tabaci population genetic structure and polyploidy in relation to competency as a vector of tomato spotted wilt virus. *PLoS ONE* **8**, e54484 (2013).
44. De Haan, P., Wagemakers, L., Peters, D. & Goldbach, R. Molecular cloning and terminal sequence determination of the S and M RNAs of tomato spotted wilt virus. *J. Gen. Virol.* **70**(Pt 12), 3469–3473 (1989).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
47. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrioSeq and iVar. *Genome Biol.* **20**, 8 (2019).
48. Lian, S. *et al.* Phylogenetic and recombination analysis of tomato spotted wilt virus. *PLoS ONE* **8**, e63380 (2013).
49. Illingworth, C. J. R., Fischer, A. & Mustonen, V. Identifying selection in the within-host evolution of influenza using viral sequence data. *PLoS Comput. Biol.* **10**, e1003755 (2014).
50. Illingworth, C. J. R. Fitness inference from short-read data: Within-host evolution of a reassortant H5N1 influenza virus. *Mol. Biol. Evol.* **32**, 3012–3026 (2015).

Acknowledgements

The authors would like to thank Dr. Tim Sit at NC State for generously providing lab space.

Author contributions

C.R.S., D.A.R. and G.K. conceived and designed the experiments. C.R.S. collected data and performed all lab work including sequencing. B.B. assisted with collecting data. D.A.R. analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-72691-3>.

Correspondence and requests for materials should be addressed to D.A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020