MBE Advance Access published February 6, 2015

Temporal dynamics of intra-host molecular evolution for a plant RNA virus

José M. Cuevas[1,†], Anouk Willemsen[1], Julia Hillung[1], Mark P. Zwart[1,‡], Santiago F. Elena[1,2,*]

[1]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, Campus UPV CPI 8E, Ingeniero Fausto Elio s/n, 46022 València, Spain.

[2]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA.

[†]Present address: Institut *Cavanilles* de Biodiversitat i Biologia Evolutiva, Universitat de València, 46980 València, Spain.

[‡]Present address: Institute of Theoretical Physics, University of Cologne, 50937 Cologne, Germany.

[*]**Corresponding author:** santiago.elena@csic.es

Abstract

Populations of plant RNA viruses are highly polymorphic in infected plants, which may allow rapid within-host evolution. To understand tobacco etch potyvirus (TEV) evolution, longitudinal samples from experimentally evolved populations in the natural host tobacco and from the alternative host pepper were phenotypically characterized and genetically analyzed. Temporal and compartmental variability of TEV populations were quantified using high throughput Illumina sequencing and population genetic approaches. Of the two viral phenotypic traits measured, virulence increased in the novel host but decreased in the original one, and viral load decreased in both hosts, though to a lesser extent in the novel one. Dynamics of population genetic diversity were also markedly different among hosts. Population heterozygosity increased in the ancestral host, with a dominance of synonymous mutations fixed, while it did not change or even decreased in the new host, with an excess of nonsynonymous mutations. All together, these observations suggest that directional selection is the dominant evolutionary force in TEV populations evolving in a novel host while either diversifying selection or random genetic drift may play a fundamental role in the natural host. To better understand these evolutionary dynamics, we developed a computer simulation model that incorporates the effects of mutation, selection and drift. Upon parameterization with empirical data from previous studies, model predictions matched the observed patterns, thus reinforcing our idea that the empirical patterns of mutation accumulation represent adaptive evolution.

***Key words*: adaptation, experimental evolution, evolution of virulence, next generation sequencing, population dynamics, *Potyvirus*, virus evolution**

Introduction

RNA viruses are the causative agents of important diseases affecting humans, livestock, and crops. RNA viruses are characterized by rapid replication rates and huge population sizes, which in combination with their high mutation and recombination rates are expected to create highly polymorphic populations (Elena and Sanjuán 2007). A leading concept commonly used to describe the evolutionary dynamics of RNA viruses is that of quasispecies. Although many authors equate viral quasispecies to the theoretical model proposed by M. Eigen to explain the origin and evolution of pre-cellular molecular replicators (Eigen 1971), this is inappropriate since many of the assumptions of the original model do not hold for real viral populations, as already pointed out by Eigen (1996). Instead, to avoid confusions, the word "mutant swarm" should be used when referring to replicating viral populations. Such mutant swarms are defined as dynamic distributions of non-identical but closely related genomes subjected to a continuous process of genetic variation (mutation and recombination), competition and selection (Domingo 1999). A particularly critical yet key aspect of the mutant swarm is that the unit of selection is not the individual virus but the cloud of interconnected genotypes (Schuster and Swetina 1988). This prediction has been experimentally tested several times, with results that are largely consistent with the theoretical expectations (Miralles et al. 1997; Burch and Chao 2000; Codoñer et al. 2006; Sanjuán et al. 2007) yet the question still remains controversial.

The central component of the mutant swarm is the presence of many different mutant genotypes, whose frequency and fate depend on their fitness effects and on the mutational coupling with other genotypes. Using site-directed mutagenesis, it has been possible to evaluate the fitness effects associated with random single-nucleotide substitutions in RNA viruses. Such studies have shown that most mutations were either lethal or had strong deleterious fitness effects (Sanjuán et al. 2004; Carrasco et al. 2007b). An important side effect of this individual mutational hypersensitivity is that it creates robustness at the population level by facilitating the removal of deleterious alleles by purifying selection, especially if population size is large (Elena et al. 2006). Several studies have also shown that the deleterious fitness effects of random mutations tended to be larger alone than in combination (Bonhoeffer et al. 2004; Burch and Chao 2004; Sanjuán et al. 2004; Lalić and Elena 2012), which suggest that antagonistic epistasis is a characteristic feature of RNA viruses. This is in agreement with studies

showing that low levels of individual robustness are associated with antagonistic epistasis (Wilke and Adami 2011; Azevedo et al. 2006; Elena et al. 2006; Sanjuán et al. 2006). The low tolerance of RNA viruses to mutations is expected to yield few alternative adaptive responses to a given environmental change, which is consistent with the abundant examples of convergent evolution in bacteriophages (*e.g.*, Bull et al. 1997; Wichman et al. 1999), animal (*e.g.*, Cuevas et al. 2002; Remold et al. 2008) and plant RNA viruses (*e.g.*, Rico et al. 2006; Agudelo-Romero et al. 2008). Indeed, the number of mutations fixed during adaptation of RNA viruses is surprisingly small given their high mutation rates (Cuevas et al. 2002; Novella et al. 2004; Agudelo-Romero et al. 2008; Cabanilles et al. 2013). Hence, adaptation of RNA viruses seems to be a rapid, simple, and repetitive process, implying that it should be at least partially predictable. As a primary driver, the effect of mutations on viral fitness, alone or in combination, determines their evolutionary fate and the genetic composition of mutant swarms. However, the restrictions imposed by the highly structured tissue organization of eukaryotic hosts, by sampling events associated to transmission between distal host's organs or even between different hosts would clearly impact the genetic constitution and evolutionary dynamics of mutant swarms.

The fast evolution and small genomes of RNA viruses makes them excellent models for experimental evolution, and their rapid adaptation in laboratory (Holland et al. 1982; Elena et al. 1996; Bordería and Elena 2002) and natural (Duffy et al. 2008) environments has been shown in several studies. Our model system is *Tobacco etch virus* (TEV; genus *Potyvirus*, family *Potyviridae*). TEV has a moderately wide host range (Shukla et al. 1994). It has a positive sense single-strand RNA genome of 9.5 kb that encodes a large polyprotein, which is autocatalytically cleaved into ten multifunctional mature viral proteins (Riechmann et al. 1992). An additional peptide, P3N-PIPO, is translated from an overlapping ORF after +2 frame shifting of the P3 cistron (Chung et al. 2008). The genome replication in *Potyviridae* is performed by a virus-encoded RNA-dependent RNA polymerase (NIb) that lacks proofreading activity. TEV mutation rate is thus high, estimated to be around 10^{-5} to 10^{-6} mutations per site and per generation (Tromas and Elena 2010) and in the same range as recombination rate (Tromas et al. 2014). As an additional piece of important information, the distribution of mutational effects on fitness has also been characterized for TEV (Carrasco et al. 2007b). Consequently, TEV is an ideal candidate to experimentally test

the origin and evolutionary fate of genetic variability both in space (different plant tissues) and time (both as the plant grows and along experimental passages) in terms of molecular adaptation dynamics. In this study we undertook the characterization of TEV populations evolving either on their natural host, *Nicotiana tabacum*, or in a new one, *Capsicum annuum*. Two phenotypic viral traits have been tracked along the evolution experiment, the number of TEV genomes produced per a fixed amount of total RNA (*i.e.*, viral load) and the negative effect of infection in plant weight and size (*i.e.*, virulence). These two traits depend on the interaction between the virus and multiple cell factors. Mutant swarms from different tissues and evolutionary time points have been characterized using ultra-deep sequencing. Finally, these experiments were complemented with a computational approach that simulates the evolution experiments and allows relevant population genetic parameters to be inferred.

Results and Discussion

Viral load evolves towards lower values

The first viral phenotypic trait evaluated was viral load. Fig. 1A shows the changes in viral load along the serial passages for each lineage and host. For making the representation clearer, viral loads quantified for each leaf were added up into a single value per plant, although *LEAF* was still used as a factor in the analysis reported in table 1. Data were fitted to the linear model described in eq. (1) of the Materials and Methods by means of GLMM. An omnibus test for the goodness of fit shows that the model shown in eq. (1) fits the data significantly better than the simplest model containing only the interception (χ^2 = 1425.261, 96 d.f., $P < 0.001$). Significant overall differences exist among hosts; with the two lineages evolved in *N. tabacum* (NT1 and NT2) showing ~2.44 fold higher viral loads than the two lineages evolved in *C. annuum* (CA1 and CA2) (fig. 1A; significant effect of the *HOST* term in table 1). Indeed, significant differences have been generated during the experimental evolution among lineages evolved in a common host (fig. 1A; significant effect of the *LINEAGE*(*HOST*) term in table 1). Viral load shows a significant reduction with evolutionary time in all lineages and hosts (fig. 1A; significant effect of the *PASSAGE* term in table 1), although the rate of decline (slope of the regression line) is homogeneous among lineages evolved in *C. annuum* but not among the two lineages evolved in *N. tabacum* (fig. 1A;

significant effect of the *PASSAGE*×*LINEAGE*(*HOST*) term in table 1: lineage NT1 has a steeper slope). Finally, on average, significant differences in viral load exist among leaves sampled from the same plant (fig. 1A; significant effect of the *LEAF*(*PASSAGE*×*LINEAGE*(*HOST*)) term in table 1). All effects were large in magnitude ($\eta_P^2 > 0.15$ in all cases), except *LINEAGE*(*HOST*) and *PASSAGE*×*LINEAGE*(*HOST*), that were small ($\eta_P^2 < 0.15$) despite being significant.

Evolution of more virulent viruses in the novel host

Next, we evaluated virulence as the effect of infection in two plant traits: height and weight. Both traits are highly correlated (Spearman's $r_S = 0.963$, 58 d.f., $P < 0.001$) and thus we performed a principal components analysis to reduce these two variables into a new one that still explains the observed variability. The first principal component (PC1) explained 96.379% of observed variability and weighted equally both measures of virulence (0.982). The value of PC1 was thus taken as a general measure of virulence. Fig. 1B shows the temporal evolution of virulence for each lineage evolved on each host. Virulence data were fitted to the linear model described in eq. (2) of the Materials and Methods by means of GLMM. The fit to the model was significant (Omnibus test for the goodness of fit: $\chi^2 = 581.940$, 57 d.f., $P < 0.001$). All significant factors had also a large effect on virulence ($\eta_P^2 > 0.15$). Overall, significant differences exist among hosts; with lineages evolved in *N. tabacum* being ~145,565% more virulent than lineages evolved in *C. annuum* (fig. 1B; significant effect of the *HOST* term in table 2). Indeed, significant differences have been generated during the experimental evolution among lineages evolved in a common host (fig. 1B; significant effect of the *LINEAGE*(*HOST*) term in table 2). Virulence shows a non-linear pattern of evolution in all lineages and hosts (fig. 1B; significant effect of the *PASSAGE* term in table 2), with completely opposite dynamics on each host. While virulence increased in the novel host, it declined in the ancestral one (fig. 1B; significant effect of the *HOST*×*PASSAGE* term in table 2). No heterogeneity in slopes of the linear models exist among lineages evolved in the same host (non significant effect of the *PASSAGE*×*LINEAGE*(*HOST*) term in table 2). This being the case, however, while the increase in virulence observed for lineage CA2 was linear with time, it was better explained by a quadratic model for lineage CA1 (partial *F*-test: $F_{1,112} = 11.545$, $P = 0.001$), suggesting the existence of

maximum virulence value. Similarly, the reduction in virulence shown by lineage NT2 was linear with time, but better explained by a quadratic model for lineage NT1 (partial F-test: $F_{1,94} = 12.276$, $P = 0.001$), thus suggesting the existence of a minimum virulence value. This observation that TEV virulence in *C. annuum* increases upon serial passages while remaining constant or slightly declining in the natural host *N. tabacum* is in good agreement with previous observations (Agudelo-Romero et al. 2008).

Virulence does not depend on virus accumulation

Provided that virulence does not represent any clear advantage for viruses, explaining why most viruses induce symptoms is a relevant question. A common assumption is that virulence is an unavoidable consequence of virus' multiplication (Lenski and May 1994) and thus a positive association must exist between virulence and accumulation. In the case of plant viruses, proving this association has been difficult as results are contradictory. A positive association has been reported for TEV infecting pepper (Agudelo-Romero et al. 2008) and for *Cauliflower mosaic virus* (CaMV) infecting turnip (Doumayrou et al. 2012), but it was not found for TEV genotypes that differed in single point mutations infecting tobacco (Carrasco et al. 2007b), among TEV genotypes evolved in a set of different ecotypes of *Arabidopsis thaliana* (Hillung et al. 2014) nor among different necrogenic and non-necrogenic satellite RNAs of *Cucumber mosaic virus* (CMV) (Betancourt et al. 2011). Here, the situation is somehow more complex and interesting. Fig. 2 shows the association between viral load and virulence. At a first look, a positive association between these two phenotypic traits exists. However, this apparent association is entirely driven by the different allometric relationships observed on each plant host species (fig. 2): lineages evolved in tobacco plants are more virulent and also accumulate to higher loads than lineages evolved in pepper plants. Indeed, a partial correlation coefficient controlling for *HOST* found no significant association between virulence and viral load ($r = 0.191$, 21 d.f., $P = 0.383$). This apparent contradiction suggests that the positive association may be pathosystem-dependent and highly affected by other environmental variables. If a positive correlation does not exist, many other factors influencing the progression of viral infection would explain virulence. In particular, virulence would not depend on within-host replication if the extent of damage is not proportional to the amount of viral particles, as in the case of a hypersensitive response (Morel and Dangl 1997), if

expressing the systemic acquired resistance pathway is costly (Heidel et al. 2004), or if allocating resources to defense detracts from vegetative growth or reproductive effort (Heil 2001; Pagán et al. 2008; Bedhomme and Elena 2011).

Evolution of genetic variability among leafs and along evolutionary passages

Samples from three different leaves taken at passages 1, 4, 6, 9, 12, and 15 were sequenced by Illumina HiSeq2000 using paired-end libraries. The leaves sequenced are referred to as L0 (the inoculated leaf), L2 (the second leaf grown after L0) and L3 (the apical leaves grown after L2). After filtering and cleaning the sequence data, the mean per-site coverage of paired-end reads for TEV lineages evolved in *C. annuum* was 25840 fold and for lineages evolved in *N. tabacum* this was slightly higher, 32821 fold. In total, taking into account all passages and leaves that were properly sequenced, we detected 45 SNPs in lineage CA1, 56 in lineage CA2, 107 in lineage NT1 (passage 1 L0 was not sequenced) and 122 in lineage NT2 (passage 1 L0 was not sequenced and passage 15 L0 had too low coverage). Of these totals, 21 SNPs were unique in CA1, 15 in CA2, 48 in NT1, and 53 in NT2. Supplementary file S1 contains all these mutations as well as their population frequencies.

As a measure of TEV genetic diversity on different leaves sampled at sequential passages, we used the population heterozygosity index H (Li 1997). First, we sought to evaluate whether a trend exist in H among leaves from the same plant (fig. 3). No consistent change in H among leaves has been observed. Among all plants analyzed, 14 show no significant association between leaf age and the genetic diversity of the viral population replicating on each leaf (Spearman's $|r_S| = 0.5$, 1 d.f., $P = 0.667$); five show significant decreases (lineage NT1 at passages 1, 4 and 12; lineage CA1 at passages 6 and 15) and five show significant increases in H (lineage NT2 at passages 1 and 12; lineage CA1 at passage 4; lineage CA2 at passages 4 and 15) (in all cases, $|r_S| = 1$, 1 d.f., $P = 0$). Whether viral genetic diversity should increase as a function of the distance of new leaves from the inoculated leaf has not been well established, as reports show contradictory results. On the one hand, Li and Roossinck (2004) found that the number of mutants in successive leaves infected with CMV decreased as a function of distance from the source leaf. At the other hand, Gutiérrez et al. (2012) found a positive association between leaf age and CaMV diversity, at least for early stages of infection. Likewise, Dunham et al. (2014) observed a minor, yet significant, increase in the

genetic diversity of *Zucchini yellow mosaic virus* populations replicating in successive leaves of a *Cucurbita pepo* vine. These two reports thus support that the mutant swarm in the phloem sap may serve as a constant source of genetic diversity, collecting variants from older leaves that colonize new ones.

Next, we sought to evaluate whether an overall time trend exist for the amount of TEV genetic diversity along the evolutionary time (fig. 3). No overall changes in population heterozygosity were observed along passages in the two lineages evolved in the new host (r_S = –0.143, 4 d.f., P = 0.787 in both lineages). Either directional selection or random drift would reduce heterozygosity, though by different mechanisms. If directional selection is at work, the fixation of a beneficial mutation at a particular genomic site means that all other alleles at this site will be rare, the distribution of allele frequencies will be markedly uneven and heterozygosity will be unexpectedly low. Furthermore, variability at linked loci will also be reduced during the selective sweep. If drift is the dominant force, fixation will be a rapid process in small populations, for instance during early infection of new leaves, while its effect will be less and less important as the population expands to big numbers. This phenomenon is particularly obvious in lineage CA2, where the successive events of periodic selection of putatively beneficial alleles drop heterozygosity to low levels after each selective sweep (nicely illustrated by the corresponding Muller plot shown in fig. 4). In sharp contrast with the lineages evolved in the novel host, significant increases in genetic diversity were observed in the two lineages evolved in the original host (r_S = 0.829, 4 d.f., P = 0.042 for lineage NT1; r_S = 0.943, 4 d.f., P = 0.005 for lineage NT2). Under a strictly neutral model of molecular evolution, average heterozygosity is expected to increase with time. The same outcome is expected if diversifying selection or frequency dependent selection come into play. Virus interaction with adaptable plant immune responses (*e.g.*, RNA silencing-based immunity) may create the conditions for diversifying selection or frequency dependent selection to operate and thus explain the increase in heterozygosity observed in the original host (lineages NT1 and NT2 in fig. 3).

Lineages evolved in *N. tabacum* show 29.5% more synonymous than nonsynonymous mutations. In sharp contrast, lineages evolved in *C. annuum* have 40% less synonymous than nonsynonymous mutations. In other words, lineages evolved in tobacco are enriched in synonymous mutations whereas lineages evolved in pepper are enriched in nonsynonymous changes. This observation, combined with the lower

heterozygosity of lineages evolved in pepper support the idea that mutant swarms evolving in the new host are mostly driven by directional selection whereas either neutral accumulation of mutations or diversifying selection may be playing a more important role in TEV populations evolving in the natural host.

Fingerprints of clonal interference in TEV populations passaged in the ancestral host?

The Muller plots (Barrick and Lenski 2013) shown in fig. 4 illustrate the temporal dynamics of different alleles. For now, let's focus on the dynamics observed in the ancestral host. Both replicates show qualitatively similar behavior. The ancestral sequence dominates the mutant swarm at early passages, but as new alleles appear, the frequency of the ancestral sequence decreases. Most of the new alleles observed have a transient existence; they appear within a plant, rise in frequency and then disappear at the next serial passage, as illustrated by the serrated pattern observed in both lineages, but specially during the first six passages of lineage NT1. After passage 9 of NT1, three mutations (C479U, U2763C, A7261G) appear on the ancestral genetic background and compete with each other and with the ancestral genome. Concomitant with the appearance of these three mutations (plus several others that never reach high frequency), the ancestral genotype declines in frequency, reaching a minimum around passage 15, where the population shows the largest variability. Likewise, after passage four of NT2, mutation C4384U appears on the ancestral genetic background and its frequency quickly increases, though the initial fast rise was likely associated to a transmission bottleneck event. At passage 6, mutation U5058C appears on the ancestral background and increases in frequency, survives the next transmission events (yet suffering a transitory reduction in frequency) and rises in frequency, now most likely competing against mutation C4384U. A third mutation, G273A, appeared during passage 9 in the still numerically abundant ancestral genome and also increases its frequency noticeably. At the end of the experiment, these three putative beneficial mutations dominate the mutant swarm, with the ancestral genotype practically being a minority variant.

We discussed above that the evolution of variability in the lineages evolved in the ancestral host could be explained by the steady accumulation of neutral mutations or, more likely, by diversifying selection. A third tantalizing possibility also exists: the

observed patterns of alternative beneficial mutations appearing, rising in frequency and coexisting in the population for considerable time are qualitatively similar to what is expected for clonal interference (Barrick and Lenski 2013). Clonal interference has been shown in the past to be an important modulator of the rate of RNA virus evolution (Miralles et al. 1999, 2000; Strelkowa and Lässig 2012; Cabanilles et al. 2013)

The most common mutations in NT1 are those at nucleotide positions C479U, U2763C and A7261G. Mutation C479U is nonsynonymous (T to M replacement) and occurs within the P1 protein, which is a highly multifunctional nucleocytoplasmic protein. It acts as a proteinase responsible for autocleavage at the C-terminal end of the polyprotein, is implicated in genome amplification and stimulates viral mRNA translation by associating to the 60S; indeed, it traffics into the granular component of the nucleolus where the final processing of preribosomal particles takes place (Urcuqui-Inchima et al. 2001; Martínez and Daròs 2014). Mutation C479U is around the middle of the P1 protein and does not affect the well-conserved proteolytic cleavage site, thus its effect may not directly be related to the protease activity but to diverting nucleolar proteins to perform novel roles in the virus infection cycle. Synonymous mutation U2763C occurs in the P3 protein. The function of this protein is not fully clear, but it has been suggested that P3 may play a role in virus movement and replication (Urcuqui-Inchima et al. 2001; Cui et al. 2010). Finally, nonsynonymous replacement A7261G (I by V) occurs within the NIb protein, which is the RNA-dependent RNA polymerase.

In NT2 the most common mutations are synonymous and affect nucleotide positions G273A, C4384U and U5058C. Mutation G273A occurs within the P1 protein. Mutations C4384U and U5058C occur within the multifunctional cylindrical inclusion body protein CI, which is implicated in cell-to-cell movement and has ATPase and RNA helicase activities (Urcuqui-Inchima et al. 2001), though neither of these synonymous mutations is expected to exert a direct effect on the function of these proteins but on RNA functions in translation, transcription or sensitivity to sequence-specific antiviral responses of the host (*e.g.*, RNA silencing).

TEV populations experience strong periodic selection in the novel host

The Muller plots in fig. 4 for the two lineages evolved in the novel host show quite different patterns among them and among those described above for the ancestral host.

In lineage CA1, the wild type virus remains the most abundant genotype along the entire duration of the evolution experiment. Different alleles appear, reached a noticeable transient frequency and then disappeared. In sharp contrast, lineage CA2 shows a pattern that corresponds to what is expected for strong periodic selection (Barrick and Lenski 2013), with two beneficial mutations sequentially sweeping in the population to fixation. The first beneficial mutation A5817G is synonymous and affects the cistron encoding for the VPg protein that is implicated in RNA translation and has been shown to be a determinant of host specificity (Urcuqui-Inchima et al. 2001). The strong beneficial effect observed for this mutation is, obviously, independent on VPg activity and may depend on other factors such as optimization of codon usage bias, alterations of secondary RNA structures, or avoidance of siRNA-mediated antiviral responses. This mutation is first detected within leaf L0 of passage 1 and was already fixed at passage 4 and remained fixed thereafter. The second beneficial mutation observed G404A, a nonsynonymous one (R by H) affecting the P1 protein, appeared on the A5817G genetic background in L0 of passages 9 and reached fixation at passage 12 (fig. 4). The potential implications of mutations in P1 on viral fitness have been already discussed in the previous section.

Parallel evolutionary dynamics among mutations: identifying coevolving groups of mutations

The short reads produced by Illumina sequencing do not allow inferring genome-wide haplotypes, which are longer than the read size, and thus linkage among mutations cannot be assessed, with the only exception of the double mutant G404A/A5817G fixed in lineage CA2. Thus, it is not at all obvious whether the several mutations discussed in the previous section arose independently on the wild type background or in other mutant genomes. An indirect way of assessing linkage among mutations is to check whether the frequency of pairs of mutations covaries along time. If two mutations have been observed together once and again, specially after transmission bottlenecks, and show a positive correlation in their frequencies, it can be concluded that they share a parallel evolutionary trajectory, thus enhancing the likelihood they are linked into the same haplotype. By contrast, if a pair of mutations is never observed together, observed just in one sample or their frequencies show no or even negative correlation, it can be concluded that they are not linked into the same haplotype. Fig. 5 shows the networks

of co-occurring mutations with parallel trajectories observed on each evolutionary lineage. Two mutations are linked by an edge if their frequencies show a significant positive correlation coefficient (Pearson's $r \geq 0.980$, FDR-corrected $P < 0.05$).

Three sets of covarying mutations have been observed in lineage NT1 (fig. 5). A large set formed by 10 mutations that covary in a pairwise manner, except synonymous mutations U2763C and U6165C that have three and four connections each. Since not all 10 mutations are interconnected, it is unlikely they may all belong to the same haplotype, though some connected pairs may be temporally linked. Likewise, the second group of three covarying mutations is not likely to be linked into a common haplotype. The third group of covarying mutations is formed by synonymous mutations A7119G and C7275U. These two mutations were first observed in L0 of passage 15 and afterwards, being likely part of the same haplotype.

We have inferred the existence of three covarying groups of mutations in lineage NT2 (fig. 5). The first group is formed by four mutations, with nonsynonymous mutation A1373C (Q by P replacement in the HC-Pro protein) being central and linked with the other three mutations, which are unlinked among them. This mutation and their three partners were only observed in passage 1 and disappeared upon the first transmission bottleneck. The second covariation group contains eight highly connected mutations, with nonsynonymous mutation G1699A (A by T replacement in the HC-Pro protein) being the most connected one. Finally, the third covariation group involves six mutations, with all but one being implicated in pairwise associations. Only synonymous mutation G273A shows a larger number of significant interactions (four).

Two significant covariation groups have been observed in lineage CA1 (fig. 5). The first one formed by nonsynonymous mutation G1430A (G to D replacement in HC-Pro) and synonymous mutation U2724C. Both mutations were first observed in L0 of passage 9 and were lost at the next serial transfer. The second covariation group is large and contains nine mutations, with the number of connections ranging between one and four. The most highly connected mutations (four edges each) are nonsynonymous mutations G1282A (V by I replacement in HC-Pro), C3713U (T by I replacement in CI) and C1868U (T by I replacement in HC-Pro), and synonymous mutation G2880A. These four mutations are all interconnected, thus likely being part of a single highly divergent haplotype.

Finally, a single covariation group has been identified in lineage CA2 (fig. 5). This large group consists of 10 mutations, including the beneficial pair G404A/A5817G already discussed above. Interestingly, mutation C9424U affecting the 3' UTR is the most highly connected one in the module. This mutation is significantly linked to beneficial mutation A5817G, but not to mutation G404A, as it only has a transient existence in the population in passage 6, before G404A was first observed.

Results of the simulation model

We developed a detailed simulation model of TEV molecular diversification and evolution, which predicts the mean number of mutations per genome that has accumulated over passages. Model selection was then performed over a set of four nested models, which varied only in the parameters being estimated from the experimental data. Model 1 assumes that the per nucleotide mutation rate, μ, is constant and all mutations all are neutral or deleterious ($\alpha = 1$). Model 2 also assumes that μ is constant but relaxes the constraint of fixed effects by allowing them to follow a truncated Weibull distribution of mutational fitness effects (DMFE) that covers the full range of effects, from lethal to beneficial. The upper limit of this DMFE is given by the parameter $\alpha > 1$. Model 3 allows for a variable μ yet mutational effects are assumed to be only neutral or deleterious. Finally, Model 4 allows for variable μ and DMFE. For the *N. tabacum* data, Model 2 was the best-supported model (table 3). This model clearly fits the data better than Model 1 (fig. 6A-F). When we fitted both μ and α (*i.e.*, Model 4), we noticed that whereas α values tend to strongly affect model fit, μ values over a one-order-of-magnitude range provide reasonable model fit (fig. 6P). On the other hand, estimating μ did not lead to an appreciable improvement in model fit, and reduced model support due to the addition of an extra free parameter (table 3). This result suggests that variation in mutation rate will not have a strong effect on the sampling of beneficial mutations, because μ values that are roughly similar to our empirical estimate for TEV (Tromas and Elena 2010) are high enough to ensure the occurrence of sufficient beneficial mutations to allow for adaptive evolution. On the other hand, the distribution of beneficial mutations appears to be critical to predicting the rate of mutation accumulation.

The idea that the distribution of beneficial mutations may be more important than the exact mutation rate for predicting the evolution of TEV in *N. tabacum* can be supported by considering the experimental system in more detail. The probability that any one specific single-nucleotide mutation will occur during the infection of one plant can be approximated as: $1-(1-\mu/3L)^x$, where L is the genome length and x is the cumulative number of genome copies over generations for the time period to be considered (calculated from a previously parameterized logistic function describing the number of genomes over time, see Material and Methods and also Zwart et al. (2012)). For 1 week (20 generations) of TEV infection in *N. tabacum*, the probability that any one single-nucleotide mutation will occur is approximately 0.19, and a rough approximation suggests over 5000 different mutations will occur ($0.19\times3L$). These numbers explain why the model fits the data for a wide range of μ values: there is a reasonable probability that some beneficial mutations will be sampled during each passage. On the other hand, if there are multiple beneficial mutations that might occur, this estimate would also explain why high levels of convergent evolution are only seen in longer passages in *N. tabacum* (Bedhomme et al. 2012; Tromas et al. 2014; Zwart et al. 2014); the occurrence of any one mutation is still not very likely and epistatic interactions (Lalić and Elena 2012) may limit the number of evolutionary trajectories once a beneficial mutation predominates in the mutant swarm.

For the *C. annuum* data, the model fitting results are similar (table 3). Model 4, which estimates both α and μ, was the best-supported, suggesting that an appreciably higher mutation rate also plays an important role in evolutionary dynamics in *C. annuum*. Compared to Model 1 (fig. 6G-I), Model 2 better predicts the higher number of mean mutations per genome observed (fig. 6J-L). For Model 4, the predicted mean number of mutations per genome is higher and – interestingly – the model also predicts periodic episodes of selection (fig. 6M-O), as clearly observed in lineage CA2 (fig. 4). Although model fit also appears to depend less on μ than on α for the *C. annuum* data (fig. 6Q), the effect was less pronounced than for the *N. tabacum* data.

Modelling of virus evolution therefore reinforces the idea that the observed patterns of mutation accumulation represent adaptive evolution. When we did not allow for mutations that increased fitness (Model 1), for both host species the model predicted the maintenance of the ancestral sequence, probably due to the bottlenecks in infection. Furthermore, the modelling results support the idea that the dynamics of molecular

evolution are different for the two host species, probably due to different bottleneck sizes during infection – which were measured empirically – and perhaps a higher mutation rate in *C. annuum*. The exact parameter estimates obtained from fitting the model to the *C. annuum* data must be approached with caution, however, since most model parameters other than bottleneck size in the inoculated leaf were not available for *C. annuum* and were assumed to be the same as for *N. tabacum.*

Conclusions

By using next generation sequencing techniques, we have described the temporal dynamics of molecular evolution of TEV in its natural host and in a novel one. We found that dynamics are markedly different between both hosts. Diversity increased in the natural host, as expected by the accumulation of neutral alleles or by the action of diversifying selection, while it was constant in the novel host, as expected by the combined action of periodic drift and strong periodic selection. These observations were qualitatively reproduced by a simulation model that was parameterized using empirical data, thus providing further support to our main conclusions: drift and neutral mutations are the dominant evolutionary forces at play in the natural host whereas positive selection was so in the new host. These differences in the dynamics of molecular evolution were correlated to phenotypic evolution. While in all cases viral load was reduced, the magnitude of the observed reduction was larger and more diverse in lineages evolved in the natural host. Selective pressures on virulence also varied among hosts. While virulence was reduced in the natural host, it increased in the novel one.

Materials and methods

Virus and plants

Plasmid pMTEV contains the TEV genome (Bedoya and Daròs 2010). The TEV genome used to generate this clone has been isolated from *N. tabacum* (Carrington et al. 1993), and its sequence is published elsewhere (Carrasco et al. 2007a). To generate a virion stock, infectious RNA was obtained by *in vitro* transcription after *Bgl*II linearization of the plasmid containing the infectious clone as described previously

(Carrasco et al. 2007a). Two *N. tabacum* plants were infected with 4 µg RNA each in the third true leaf, and the symptomatic systemically infected tissue was collected at 10 days post inoculation (dpi) and used for purification of virions (Carrasco et al. 2007a). Virions were resuspended in 200 mL of 0.05 M borate buffer (pH 8.0, with 5 mM EDTA) with 20% glycerol. Virion concentration of the stock obtained was estimated to be around 5×10^5 virions/mL according to titration assays in *Chenopodium quinoa* (Kleczkowski 1950).

Host species *N. tabacum* cv. Xanthi and *C. annuum* cv. Marconi (*Solanacea* family) were used, where TEV produces systemic symptoms. For all the experimental steps, plants were maintained in a greenhouse at 25 °C and a 16 h photoperiod. We have previously observed that viral load reached a plateau at 7 dpi for both plant hosts (Bedhomme et al. 2012). Differences in viral accumulation were observed (2.2×10^5 viral genomes/ng of total RNA for *N. tabacum* versus 3.2×10^4 viral genomes/ng of total RNA for *C. annuum*) and consequently taken into account to equalize the size of the inoculum, even though we still expect a narrower transmission bottleneck in the *C. annuum* due to lower infectivity of virions for this host (Zwart et al. 2012).

Experimental evolution

The experimental evolution design included two different hosts, TEV being better adapted to *N. tabacum* (Carrington et al. 1993; Agudelo-Romero et al. 2008) than to *C. annuum* (Agudelo-Romero et al. 2008). This design consisted of two replicates for each plant host (indicated as NT1 and NT2 for *N. tabacum* and CA1 and CA2 for *C. annuum*). To initiate the experimental evolution, virion stock was diluted in inoculation buffer (100 mg/mL Carborundum, 0.5 M K_2HPO_4). For each host, two plants/replicates were mechanically inoculated with 5 mL in the third true leaf. For convenience, the inoculated leaf will be indicated as L0, whereas L1, L2 and L3 are the subsequent upper leaves, from the oldest to the youngest. L1 corresponds to the leaf opposed to L0 and it remains uninfected for the duration of this experiment. L3 corresponds to the small newly born leaves in the apical meristem. For the subsequent passages, at 7 dpi, the aerial part of the four plants (two replicates per host) was collected. For each plant, the four leaves described above (L0, L1, L2, and L3) were separately ground in mortar with liquid nitrogen. An equal amount of ground material for each of the four leaves from a

given plant was subsequently mixed in a single tube and used to initiate the next passage. To do so, a sap was prepared with about 100 mg of mixed tissue in 1 mL of inoculation buffer if the infected tissue was from *N. tabacum* and 100 mg of infected tissue in 200 mL of inoculation buffer for *C. annuum,* so that each infection was started with similar amounts of viral RNA. Fifteen serial passages were performed.

Measure of viral accumulation

The viral load of each plant (4 lineages × 15 passages = 60 plants) was quantified by RT-qPCR. For this, similarly to the evolution experimental procedure, a mix of the ground material (four leaves per plant) was obtained. Each mix was subsequently separated in two tubes, one for total RNA extraction (and viral load estimation) and a second one for phenotyping assays, as explained below. RNA extraction from 100 mg tissue per plant was performed using Spin Plant RNA Mini Kit (Invitek) following manufacturer's instructions. The concentration of total plant RNA extracts was adjusted to 50 ng/mL for each sample and the quantification of viral load was done with real time RT-PCR (RT-qPCR), using primers and methods previously described (Lalić et al. 2011). Amplifications were done using the ABIPRISM Sequence Analyzer 7500 (Applied Biosystems), according to the following profile: 5 min at 42 °C, 10s at 95 °C following 40 cycles of 5 s at 95 °C and 34 s at 60 °C. RT-qPCR reactions were performed in triplicate for each sample. Quantification results were further examined using SDS7500 software v. 1.2.3 (Applied Biosystems).

Viral load data were analyzed using generalized linear mixed models (GLMM) with a Normal distribution and an identity link function. The model has one fixed factor, the host species wherein evolution took place (*HOST*), and three random factors: the virus evolutionary lineage (*LINEAGE*), which is nested within *HOST*, the passage (*PASSAGE*), and the leaf sampled for which the viral load was evaluated (*LEAF*). Since *LEAF* is evaluated for each plant, it was considered as nested within the interaction between *PASSAGE* and *LINEAGE* within *HOST*. The model equation was:

$$L_{ijklm} = \mu + HOST_i + PASSAGE_j + (HOST\times PASSAGE)_{ij} + LINEAGE(HOST)_{ik} + (PASSAGE\times LINEAGE(HOST))_{ijk} + LEAF(PASSAGE\times LINEAGE(HOST))_{ijkl} + \varepsilon_{ijklm}, \quad (1)$$

where μ is the grand mean value and ε_{ijklm} is the error associated with individual measure m. The statistical significance of each factor was evaluated using a likelihood ratio test (*LRT*) that asymptotically follows a χ^2 distribution.

The magnitude of the different effects included in the model was evaluated using the η_P^2 statistic that represents the proportion of the total variability attributable to a given factor. Conventionally, values $\eta_P^2 < 0.05$ are considered as small, $0.05 \leq \eta_P^2 < 0.15$ as medium and $\eta_P^2 \geq 0.15$ as large effects.

Virulence assays

Virulence assays were performed using the samples described in the previous section to characterize how the interaction between the plant and the evolving TEV mutant swarm changed along experimental evolution. To do so, the viral load quantification previously obtained was used to prepare saps of equal viral RNA concentration. Each sap was used to infect 7-8 plants of *N. tabacum* or *C. annuum*, depending on the original host. At 21 dpi, the aerial part of each plant was measured with a precision of 0.5 cm and weighted with a precision of 10 mg. We then calculated the virulence as the reduction in these two traits due to infection (Bedhomme et al. 2012).

Virulence data were analyzed using GLMM with a Normal distribution and an identity link function. The model has one fixed factor, *HOST*, and two random ones: *LINEAGE*, which is nested within *HOST*, and *PASSAGE*. The model equation was:

$$V_{ijklm} = \mu + HOST_i + PASSAGE_j + (HOST \times PASSAGE)_{ij} + LINEAGE(HOST)_{ik} + (PASSAGE \times LINEAGE(HOST))_{ijk} + \varepsilon_{ijkl}, \quad (2)$$

where μ is the grand mean value and ε_{ijkl} is the error associated with individual measure l. Significance and magnitude of effects were evaluated as above.

To evaluate whether virulence changed in a linear manner throughout the evolution experiment or it reached a maximum or minimum value after some passage, the virulence data measured for each lineage were fitted, by the least squares method, to a null linear model ($V = \beta_0 + \beta_1 PASSAGE$) and to a quadratic model ($V = \beta_0 + \beta_1 PASSAGE + \beta_2 PASSAGE^2$). A partial *F*-test was used to test whether addition of an additional parameter (β_2) to the model was justified in the basis of a significant reduction in the proportion of unexplained variance (Quinn and Keough 2002).

Estimation of TEV bottleneck sizes

We estimated the effective population size (N_e) for TEV infection of *N. tabacum* and *C. annuum* plants, by counting the number of primary infection foci. Plants were inoculated with TEV-eGFP (Zwart et al. 2011) using the same protocol as during serial passages. The ground tissue used to prepare the inoculum was infected tissue collected from *N. tabacum* or *C. annuum*, respectively, so as to mimic passage conditions. Fluorescent primary infection foci were quantified by microscopy at 3 dpi (see Zwart et al. (2011) for details). The Poisson and negative binomial distributions were both fitted to the data using the fitdistr() function in R 3.0.1 (The R Foundation, 2013), and the Akaike information criterion (*AIC*) was used to determine which model was better supported. We subsequently refer to the mean number of foci in the inoculated leaf as λ_3, as this is bottleneck size for infection in the third true leaf, in which infection results from mechanical inoculation.

Preparation of samples for Illumina NGS

Only samples belonging to leaves L0, L2 and L3 from passages 1, 4, 6, 9, 12, and 15, were chosen for ultra-deep sequencing analysis. For lineages NT1 and NT2, samples from L0 for passage 1 were lost during the process and not used for sequencing. For each sample, RNA extraction from 100 mg ground tissue was performed using Spin Plant RNA Mini Kit (Invitek) and the concentration of total RNA extracts was adjusted to 50 ng/mL. Viral load was also obtained for these samples as previously described, showing high titers in all cases, ranging 6.9×10^4 - 1.4×10^6 viral genomes/ng total RNA for *N. tabacum* samples and 10^4 - 6.4×10^5 viral genomes/ng total RNA for *C. annuum* samples.

The TEV genome was amplified in three overlapping fragments, performing first a retrotranscription (RT) with AccuScript (Agilent Technologies) following manufacturer's instructions. To minimize the potential introduction of mistakes during this reaction, seven independent RTs (followed by seven parallel PCRs) were performed for each sample. Ninety ng total RNA were added in each reaction. PCR amplification was subsequently performed with the high-fidelity Phusion polymerase (Finnzymes) following manufacturer's instructions. The conditions used for PCR were one cycle of

98 °C for 3 min followed by 30 cycles of denaturation for 8 s at 98 °C, annealing of primers for 30 s at 57 °C, and extension for 2 min, with a final 5 min extension at 72 °C. The three pairs of primers used are the following ones: (PC2-10f 5'-GCAATCAAGCATTCTACTTC, PC40-45r 5'-ATCCAACAGCACCTCTCAC), (PC45-48f 5'-TTGACGCTGAGCGGAGTGATGG, PC67-77r 5'-AATGCTTCCAGAATATGCC) and (PC73-80f 5'-TCATTACAAACAAGCACTTG, PC97-101r 5'-CGCACTACATAGGAGAATTAG), with the reverse primer used for the RT in each case. The seven PCR products obtained for each sample were pooled together, purified from agarose gels using the Gene JET PCR Purification Kit (Fermentas) and quantitated by PicoGreen fluorescence (Invitrogen). After equimolar pooling of all three amplicons for each sample, sequencing was performed at GenoScreen (Lille, France). Illumina HiSeq2000 2×100bp paired-end libraries with multiplex adapters were prepared along with an internal PhiX control. Sequencing quality control was performed by GenoScreen, based on PhiX error rate and Q30 values.

Plasmid pMTEV was used as a control of the ultra-deep sequencing. To do so, the plasmid was diluted in total RNA from healthy plants of *N. tabacum* or *C. annuum* (50 ng/ml), respectively, to simulate the natural conditions of plant RNA extracts. After quantifying viral load in these two samples by RT-qPCR, three serial dilutions were performed for both control samples maintaining total RNA concentration at 50 ng/mL and modifying viral titer to 10^4, 10^5 and 10^6 viral genomes/ng total RNA.

Bioinformatic analysis

Artifact filtering and read quality trimming (3' minimum Q20 and minimum read-length of 50 bp) was done using FASTX-Toolkit v.0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/index.html). De-replication of the reads and 5' quality trimming requiring a minimum of Q20 was done using PRINSEQ-lite v.0.20.3 (Schmieder and Edwards 2011). Reads containing undefined nucleotides (N) were discarded. Before mapping, reads left without a mate were separated into a different file. Mapping was done against the reference genome TEV-7DA (accession: DQ986288) using Bowtie v.2.1.0 (Langmead and Salzberg 2012). Mapping was done separately for the paired reads, single reads and for the single and paired reads merged. For every sample, single nucleotide polymorphisms (SNPs) were identified using

SAMtools' mpileup (Li et al. 2009) and VarScan v.2.3.6 (Koboldt et al. 2012). For SNP calling maximum coverage was set to 20000 and SNPs with a frequency < 1% were discarded.

The absolute nucleotide frequency of A, U, G, C, and - (gap) was calculated with a costume Perl script and used to compute the per site heterozygosity in each sample, using the formula $h_l = 1 - \sum_{i=1}^{m} f_{li}^2$, where f_{li} stands for the frequency of the $m = 5$ states on each given position l (Li 1997). States with a ratio lower than 0.01 were ignored. Statistically significant values for h were determined using a two-tailed z-score test. The average heterozygosity of a sample was thus calculated as $H = \frac{1}{L}\sum_{l=1}^{L} h_l$, where $L = 9539$ is the length of TEV genome (Li 1997).

Networks of co-occurring mutations were drawn using Cytoscape 3.0 (Su et al. 2014).

Simulation model of TEV molecular evolution

We generated a detailed simulation model of TEV molecular evolution, with the main goal of predicting the mean number of mutations that accumulate in the whole genome over serial passages, as well as the within-host competitive fitness of the virus variants present in the population. The model includes the following processes: (*i*) population bottlenecks following inoculation in leaf L0 (Zwart et al. 2011), (*ii*) bottlenecks during the establishment of systemic infection in the L2 and L3 leaves (L1 is not modeled, since there is no detectable infection in this leaf) (Tromas et al. 2014), (*iii*) growth of the number of virus genomes within the host over time (Zwart et al. 2012) and how the virus population is distributed over different leaves over time, (*iv*) single-nucleotide mutations (Tromas and Elena 2010) and their DMFE (Carrasco et al. 2007b), and (*v*) displacement of virus variants based on their competitive fitness.

Infection begins with a bottleneck in the inoculated leaf of size λ_3, the number of virions infecting the inoculated leaf, which for all our experiments was the third true leaf. For each genotype present in the inoculum, the number of infecting individuals is determined by a random draw from a negative binomial distribution using the R version 3.0.1 rnbinom() function, with a mean $\lambda_3 f_i$, where f_i is the frequency of the i^{th} variant in

the inoculum, and a shape parameter σ, where λ_3 and σ are both obtained from experimental data for the total number of primary infection foci per inoculated leaf.

We assumed there were 2.91 virus generations per day (Martínez et al. 2011), and there are therefore approximately 20 generations during a virus infection lasting one week, the length of single serial passage. The total number of viral genomes in the plant over time follows a logistic function, such that at *t* hours post-inoculation there are

$$N(t) = \frac{\lambda_3 \kappa}{\lambda_3 + (\kappa - \lambda_3)e^{-r_0 t}}$$

viral genomes, where κ is the carrying capacity and r_0 is the initial growth rate. However, the virus population is divided over the different infected leaves. We followed the approach of Tromas et al. (2014), where the frequency of infected cells in different leaves was tracked over time and a meta-population susceptible-infectious (SI) model was then fitted to the cell-level data. The SI model predicts that the rate of change of the fraction of infected cells (*I*) in the k^{th} leaf is:

$$\frac{dI_k}{dt} = \beta I_k S_k + \chi_k S_k \sum_{j=1}^{k-1} I_j \,,$$

where *S* is the fraction of susceptible (*i.e.*, uninfected) cells, β is the constant for cell-to-cell transmission of infection within a leaf (assumed to be the same over all leaves), and χ is the constant for between-leaf transmission to the k^{th} leaf. The model only allows virus infection to be transmitted from lower to higher leaves, as is typically the case for infection by a phloem-transported virus. Moreover, there is leaf-dependent aggregation of infected cells (*i.e.*, infected cells not randomly distributed over the leaf but are likely to be found together due to cell-to-cell movement of the virus within a leaf). The constant ψ_k determines the strength of spatial aggregation:

$$S_k = \begin{cases} (1 - I_k/\psi_k), & I_k < \psi_k \\ 0, & I_k \geq \psi_k \end{cases}.$$

Using this SI meta-population model, the number of virus genomes in the k^{th} leaf at time *t* can be estimated to be

$$N_k(t) = N(t) I_k(t) / \sum_{j=1}^{k} I_j(t).$$

The population initiating infection in each systemically infected leaf is, however, also subject to an additional leaf-dependent bottleneck (Tromas et al. 2014). The population size in each leaf after one virus generation ($N_k(t = 24/2.91)$) is estimated to be larger than the leaf-dependent bottleneck by the fitted SI model. During the first virus generation, we therefore sampled λ_k individuals from the inoculated leaf to be the

founders of systemic infection in the k^{th} leaf. We assume that λ_k also follows a negative binomial distribution, with a mean λ_k and using the same shape parameter σ as for the inoculated leaf. The negative binomial distribution was assumed to be zero truncated (*i.e.*, the same leaves always become infected when a plant is inoculated under the conditions used). We then determined the frequency of all i variants in the founding population, and assigned $N_k(t = 24/2.91)$ number of viral genomes present in the leaf their identity accordingly.

To make the model computationally tractable, for each of the N viral genomes present in an infected plant, we modeled only the number of single-nucleotide mutations (m) that have occurred since the start of the experiment, and its within-host competitive fitness (W) relative to the ancestral wild-type virus. Here μ is the mutation rate per base per generation, L is the genome length and ξ is the fraction of mutations that is not lethal. We assume $\mu L \xi$ follows a binomial distribution for each genome per generation. Given the growth of the population in each leaf is constrained by the logistic growth and SI models, we ignore the occurrence of lethal mutations as their impact will be negligible (their only effect could be to slightly lower the frequency of a particular virus variant). We then assume that the DMFE follows a Weibull distribution, with a shape parameter γ and scale parameter τ, and use the R version 3.0.1 rweibull() function to make random draws from this distribution. Moreover, this distribution is truncated at a maximum value of α. We assume that mutational effects are additive, so that fitness of a mutant (W) will be the product of the fitness effects of all previous mutations. For computational reasons, we rounded off W values to the nearest the $1/100^{th}$, thereby limiting the number of possible fitness values to 100γ values, and only simulated the first 15 generations of infection, the exponential growth phase. Finally, during each virus generation in each leaf, the new genomes generated were partitioned over the existing genotypes based on their frequency and competitive fitness, so that for a genotype c:

$$N_{c,k}(t) = \left[N_{c,k}(t-1) + \left(N_k(t) - N_k(t-1)\right) \frac{f_c W_c}{\sum_{h=1}^{i} f_h W_h} \right].$$

After an infection of g generations, virus populations from all infected leaves are pooled into one population, and this mixed population is used to again draw λ_3 individuals to initiate the next round of infection. Note that during the first virus generation, only the inoculated leaf is infected. Variants that occur *de novo* during this first round of infection can be sampled when infection is founded in systemically infected leaves. All

de novo variation that subsequently occurs will be limited to the leaf in which it occurs, until pooling of tissues occurs at the end of infection.

Model parameter estimates and fitting

The model was implemented in R version 3.0.1, and all model parameters are given in Table 4. For *N. tabacum*, parameters λ_3 and σ were estimated from the primary-infection-foci data. Zwart et al. (2012) provide estimates of κ and r_0, using data obtained under highly similar conditions to the ones used here for passaging. Model estimates for the SI model (I_0, β, χ_5, χ_6, ψ_3, ψ_5, and ψ_6) were obtained from Tromas et al. (2014), who again used similar conditions to those we used for our experiments, including the size of the plants and inoculation in the same leaf. From the same study we could also obtain estimates of the leaf-dependent bottleneck size for the two systemically infected leaves (χ_5, χ_6), the fraction of cells infected in the inoculated leaf at the immediately after inoculation (I_0), and estimates of the spatial aggregation of infected cells in each leaf (ψ_3, ψ_5 and ψ_6). Note that the virus cannot be detected in the leaf L1 under the conditions used here (Tromas et al. 2014); verified by RT-qPCR with specific primers for the coat protein here), and hence we do not model infection in this leaf. The single-nucleotide mutation rate (μ) has been estimated in Tromas and Elena (2010), and frequency of lethal mutations ($1 - \zeta$) has been reported in Carrasco et al. (2007b). To obtain parameters for the Weibull distribution describing DMFE (γ and τ), we used the within-host competitive fitness data reported in Carrasco et al. (2007b), restricting our analysis to only non-lethal mutations. Parameter estimates were made by using the fitdistr() function in R version 3.0.1 to fit a Weibull distribution. The only parameter for which we do not have an estimate *a priori* is α, the value at which the Weibull distribution for the DMFE is truncated.

For *C. annuum,* parameters λ_3 and σ were estimated from the primary-infection-foci data. For all other parameters we assumed the same values as for *N. tabacum*, as estimates are not available. Note that there are DMFE data available for *C. annuum*, but these data concern virus accumulation and not within-host fitness (Lalić et al. 2011). For TEV infection of *N. tabacum*, within-host fitness and virus accumulation were not correlated (Zwart et al. 2014), suggesting we cannot infer meaningful information on the DMFE of TEV within-host fitness from accumulation data.

To fit the model to the experimental data simultaneously for all four models described in the main text, we performed a grid search over large parameter spaces of μ and α. The model was initially run for evolution of 25 independent lineages in an initial rough fitting, and 100 independent lineages were used for fitting in a more restricted, higher resolution grid space. Model predictions for the mean number of single-nucleotide mutations per viral genome (μL) at a given serial passage were compared to the NGS data by determining the sum of squares, and then calculating the corresponding negative log likelihood (*NLL*) (Johnson and Omland 2004).

Supplementary material

Supplementary file S1 contains the allelic frequencies measured throughout the experiments. This file is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org).

Acknowledgements

We thank Francisca de la Iglesia and Paula Agudo for excellent technical assistance, our labmates for useful discussions and suggestions and Dr. José A. Daròs for gifting us the pMTEV infectious clone. This work was supported by grants BFU2009-06993 and BFU2012-30805 from the Spanish Ministry of Economy and Competitiveness (MINECO), grant PROMETEOII/2014/021 from Generalitat Valenciana, and by the European Commission 7th Framework Programme (FP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project to SFE. JMC was supported by a JAE-doc postdoctoral contract from CSIC. AW was supported by the EvoEvo project. JH was recipient of a predoctoral contract from MINECO. MPZ was supported by a Juan de la Cierva postdoctoral contract from MINECO.

References

Agudelo-Romero P, de la Iglesia F, Elena SF. 2008. The pleiotropic cost of host-specialization in tobacco etch potyvirus. *Infect Genet Evol.* 8:806-814.

Azevedo RB, Lohaus R, Srinivasan S, Dang KK, Burch CL. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* 440:87-90.

Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. *Nat Rev Genet.* 14:827-839.

Bedhomme S, Elena SF. 2011. Virus infection suppresses *Nicotiana benthamiana* adaptive phenotypic plasticity. *PLoS ONE* **6**:e17275.

Bedhomme S, Lafforgue G, Elena SF. 2012. Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Mol Biol Evol.* 29:1481-1492.

Bedoya LC, Daròs JA. 2010. Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149:234-240.

Betancourt M, Fraile A, García-Arenal F. 2011. Cucumber mosaic virus satellite RNAs that induce similar symptoms in melon plants show large differences in fitness. *J Gen Virol*. 92:1930-1938.

Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ. 2004. Evidence for positive epistasis in HIV-1. *Science* 306:1547-1550.

Bordería AV, Elena SF. 2002. *r*- and *K*-selection in experimental populations of *Vesicular stomatitis virus*. *Infect Genet Evol.* 2:137-143.

Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ. 1997. Exceptional convergent evolution in a virus. *Genetics* 147:1497-1507.

Burch CL, Chao L. 2000. Evolvability of an RNA virus is determined by its mutational neighborhood. *Nature* 406:625-628.

Burch CL, Chao L. 2004. Epistasis and its relationship to canalization in the RNA virus ϕ6. *Genetics* 167:559-567.

Cabanilles L, Arribas M, Lázaro E. 2013. Evolution at increased error rate leads to the coexistence of multiple adaptive pathways in an RNA virus. *BMC Evol Biol*. 13:11.

Carrasco P, Daròs JA, Agudelo-Romero P, Elena SF. 2007a. A real-time RT-PCR assay for quantifying the fitness of *Tobacco etch virus* in competition experiments. *J Virol Meth.* 139:181-188.

Carrasco P, de la Iglesia F, Elena SF. 2007b. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J Virol.* 81:12979-12984.

Carrington JC, Haldeman R, Dolja VV, Restrepo-Hartwig MA. 1993. Internal cleavage and trans-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus *in vivo*. *J Virol.* 67:6995-7000.

Chung BY, Miller WA, Atkins JF, Firth AE. 2008. An overlapping essential gene in the Potyviridae. *Proc Natl Acad Sci USA.* 105:5897-5902.

Codoñer FM, Daròs JA, Solé RV, Elena SF. 2006. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* 2:e136.

Cuevas JM, Elena SF, Moya A. 2002. Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162:533-542.

Cui X, Wei T, Chowda-Reddy RV, Sun G, Wang A. 2010. The *Tobacco etch virus* P3 protein forms mobile inclusions via the early secretory pathway and traffics along actin microfilaments. *Virology* 397:56-63.

Domingo E. 1999. Quasispecies. In Granoff A, Webster RG, editors. Encyclopedia of Virology. London (UK): Academic Press. p. 1431-1436.

Doumayrou J, Avellan A, Froissart R, Michalakis Y. 2012. An experimental test of the transmission-virulence trade-off hypothesis in a plant virus. *Evolution* 67:477-486.

Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet.* 9:267-276.

Dunham JP, Simmons HE, Holmes EC, Stephenson AG. 2014. Analysis of viral (*Zucchini yellow mosaic virus*) genetic diversity during systemic movement through a *Cucurbita pepo* vine. *Virus Res*. 191:172-179.

Eigen M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465-523.

Eigen M. 1996. On the nature of virus quasispecies. *Trends Microbiol*. 4:216-218.

Elena SF, Carrasco P, Daròs JA, Sanjuán R. 2006. Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* 7:168-173.

Elena SF, González-Candelas F, Novella IS, Duarte EA, Clarke DK, Domingo E, Holland JJ, Moya A. 1996. Evolution of fitness in experimental populations of *Vesicular stomatitis virus*. *Genetics* 142:673-679.

Elena SF, Sanjuán R. 2007. Virus evolution: Insights from an experimental approach. *Annu Rev Ecol Evol Syst*. 38:27-52.

Gutiérrez S, Yvon M, Pirolles E, Garzo E, Fereres A, Michalakis Y, Blanc S. 2012. Circulating virus load determines the size of bottlenecks in viral populations progressing within a host. *PLoS Pathog*. 8:e1003009.

Heidel AJ, Clarke JD, Antonovics J, Dong X. 2004. Fitness costs of mutations affecting the systemic acquired resistance pathway in *Arabidopsis thaliana*. *Genetics* 168:2197-2206.

Heil M. 2001. The ecological concept of costs of induced systemic resistance (ISR). *Eur J Plant Pathol.* 107:137-146.

Hillung J, Cuevas JM, Valverde S, Elena SF. 2014. Experimental evolution of an emerging plant virus in host genotypes that differ in their susceptibility to infection. *Evolution* 68:2467-2480.

Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S. 1982. Rapid evolution of RNA genomes. *Science* 215:1577-1585.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. *Trends Ecol Evol.* 19:101-108.

Kleczkowski A. 1950. Interpreting relationships between the concentrations of plant viruses and numbers of local lesions. *J Gen Microbiol.* 4:53-69.

Quinn GP, Keough MJ. 2002. Experimental design and data analysis for biologists. Cambridge: Cambridge University Press. p. 133.

Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, Miller C, Mardis E, Ding L, Wilson R. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 22:568-576.

Lalić J, Cuevas JM, Elena SF. 2011. Effect of host species on the distribution of mutational fitness effects for an RNA virus. *PLoS Genet.* 7:e1002378.

Lalić J, Elena SF. 2012. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. *Heredity* 109:71-77

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth*. 9:357-359.

Lenski RE, May RM. 1994. The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *J Theor Biol.* 169:253-265

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25:2078-2079.

Li H, Roossinck MJ. 2004. Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J Virol*. 78:10582-10587.

Li WH. 1997. Molecular Evolution. Sunderland (MA): Sinauer Associates Inc.

Martínez F, Daròs JA. 2014. *Tobacco etch virus* protein P1 traffics to the nucleolus and associates with the host 60S ribosomal subunits during infection. *J Virol*. 88:10725-10737.

Martínez F, Sardanyés J, Elena SF, Daròs JA. 2011. Dynamics of a plant RNA virus intracellular accumulation: Stamping machine vs. geometric replication. *Genetics* 188:637–646.

Miralles R, Gerrish PJ, Moya A, Elena SF. 1999. Clonal interference and the evolution of RNA viruses. *Science* 285:1745-1747.

Miralles R, Moya A, Elena SF. 1997. Is group selection a factor modulating the virulence of RNA viruses? *Genet Res*. 69:165-172.

Miralles R, Moya A, Elena SF. 2000. Diminishing returns of population size in the rate of RNA virus adaptation. *J Virol*. 74:3566-3571.

Morel JB, Dangl JL. 1997. The hypersensitive response and the induction of cell death in plants. *Cell Death Differ.* 4:671-683.

Novella IS, Zárate S, Metzgar D, Ebendick-Corpus BE. 2004. Positive selection of synonymous mutations in *Vesicular stomatitis virus*. *J Mol Biol.* 342:1415-1421.

Pagán I, Alonso-Blanco C, García-Arenal F. 2008. Host responses in life-history traits and tolerance to virus infection in *Arabidopsis thaliana*. *PLoS Pathog.* 4:e1000124.

Remold SK, Rambaut A, Turner PE. 2008. Evolutionary genomics of host adaptation in *Vesicular stomatitis virus*. *Mol Biol Evol.* 25:1138-1147.

Rico P, Ivars P, Elena SF, Hernández C. 2006. Insights into the selective pressures restricting *Pelargonium flower break virus* genome variability: Evidence for host adaptation. *J Virol.* 80:8124-8132.

Riechmann JL, Laín S, García JA. 1992. Highlights and prospects of potyvirus molecular biology. *J Gen Virol.* 73:1-16.

Sanjuán R, Cuevas JM, Furió V, Holmes EC, Moya A. 2007. Selection for robustness in mutagenized RNA viruses. *PLoS Genet*. 3:e93.

Sanjuán R, Elena SF. 2006. Epistasis correlates to genomic complexity. *Proc Natl Acad Sci USA.* 103:14402-14405.

Sanjuán R, Moya A, Elena SF. 2004. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc Natl Acad Sci USA.* 101:15376-15379.

Sanjuán R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci USA.* 101:8396-8401.

Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863-864.

Schuster P, Swetina J. 1988. Stationary mutant distributions and evolutionary optimization. *Bull Math Biol.* 50:635-660

Shukla DD, Ward CW, Brunt AA. 1994. The Potyviridae. Wallingford: CABI.

Strelkowa N, Lässig M. 2012. Clonal interference in the evolution of influenza. *Genetics* 192:671-682.

Su G, Morris JH, Demchak B, Bader GC. 2014. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics* 47:8.13.1-8.13.24.

Tromas N, Elena SF. 2010. The rate and spectrum of spontaneous mutations in a plant RNA virus. *Genetics* 185:983-989.

Tromas N, Zwart MP, Lafforgue G, Elena SF. 2014. Within-host spatiotemporal dynamics of plant virus infection at the cellular level. *PLoS Genet.* 10:e1004186.

Tromas N, Zwart MP, Poulain M, Elena SF. 2014. Estimation of the *in vivo* recombination rate for a plant RNA virus. *J Gen Virol.* 95:724-732.

Urcuqui-Inchima S, Haenni AL, Bernardi F. 2001. Potyvirus proteins: a wealth of functions. *Virus Res.* 74:157-75.

Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. 1999. Different trajectories of parallel evolution during viral adaptation. *Science* 285:422-424.

Wilke CO, Adami C. 2001. Interaction between directional epistasis and average mutational effects. *Proc R Soc B.* 268:1469-1474.

Zwart MP, Daròs JA, Elena SF. 2011. One is enough: *in vivo* effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog.* 7:e1002122.

Zwart MP, Daròs JA, Elena SF. 2012. Effects of *Potyvirus* effective population size in inoculated leaves on viral accumulation and the onset of symptoms. *J Virol.* 86:9737–9747.

Zwart MP, Willemsen A, Daròs JA, Elena SF. 2014. Experimental evolution of pseudogenization and gene loss in a plant RNA virus. *Mol Biol Evol.* 31:121-134.

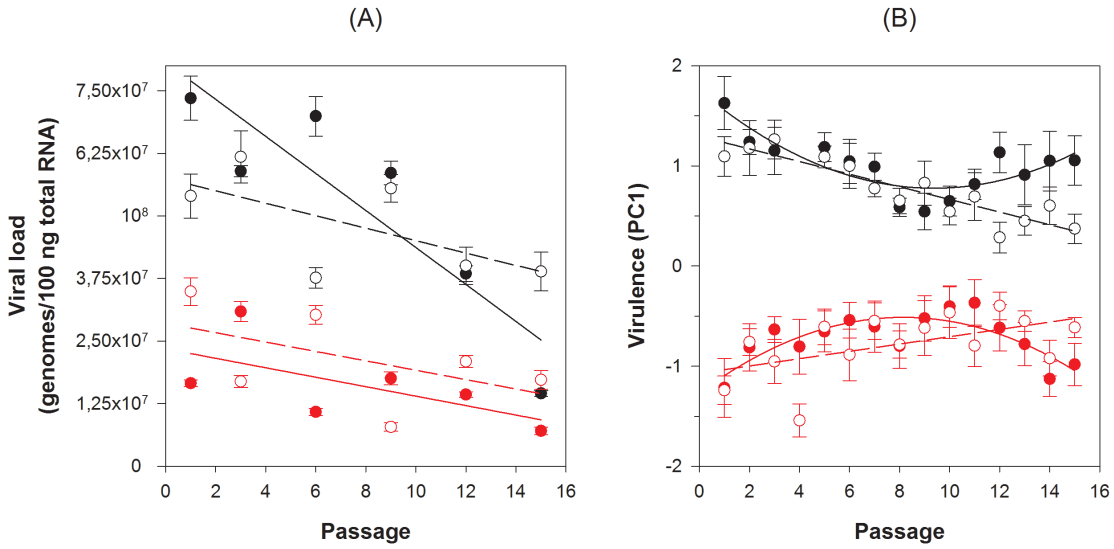

Fig. 1. Evolution of viral load (A) and of virulence (B) for the lineages serially passaged on each of the two experimental hosts. Black symbols and lines represent lineages NT1 (solid dots and continuous line) and NT2 (open dots and dashed line) evolved in *N. tabacum*; red symbols and lines represent lineages CA1 (solid dots and continuous line) and CA2 (open dots and dashed line) evolved in *C. annuum*. Error bars represent ±1 SEM. Linear or quadratic regressions were fitted to each lineage. Decision of using a quadratic model instead of a linear one was based on a partial *F*-test.

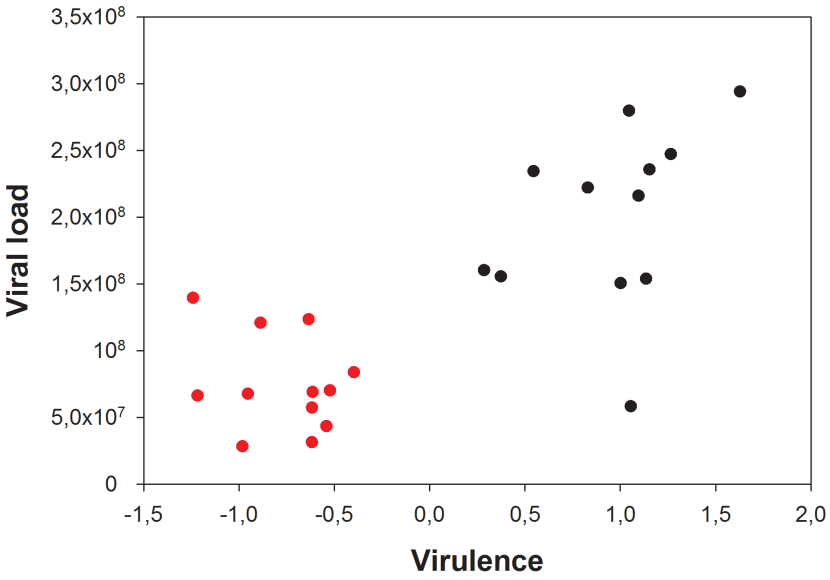

Fig. 2. Association between viral load and virulence. Red dots correspond to viral lineages evolved in *C. annuum*, whereas black dots correspond to lineages evolved in the ancestral host *N. tabacum*.

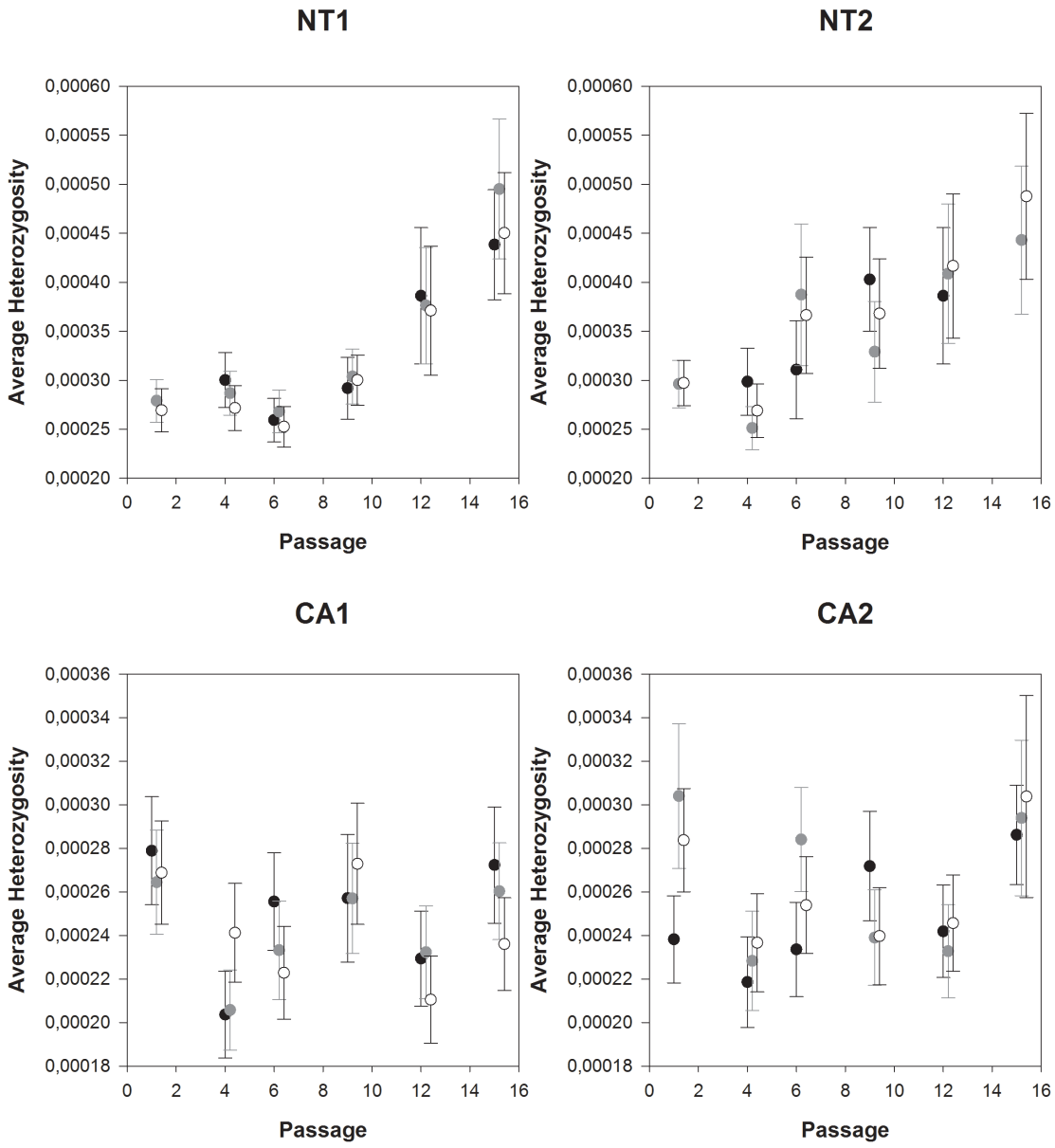

Fig. 3. Evolution of genetic diversity, measured as population heterozygosity (*H*), along the evolution experiment on each host for each lineage. Black circles represent *H* in L0 (the inoculated leaf), gray circles in L2 and white circles in L3. Error bars represent ±1 SEM.

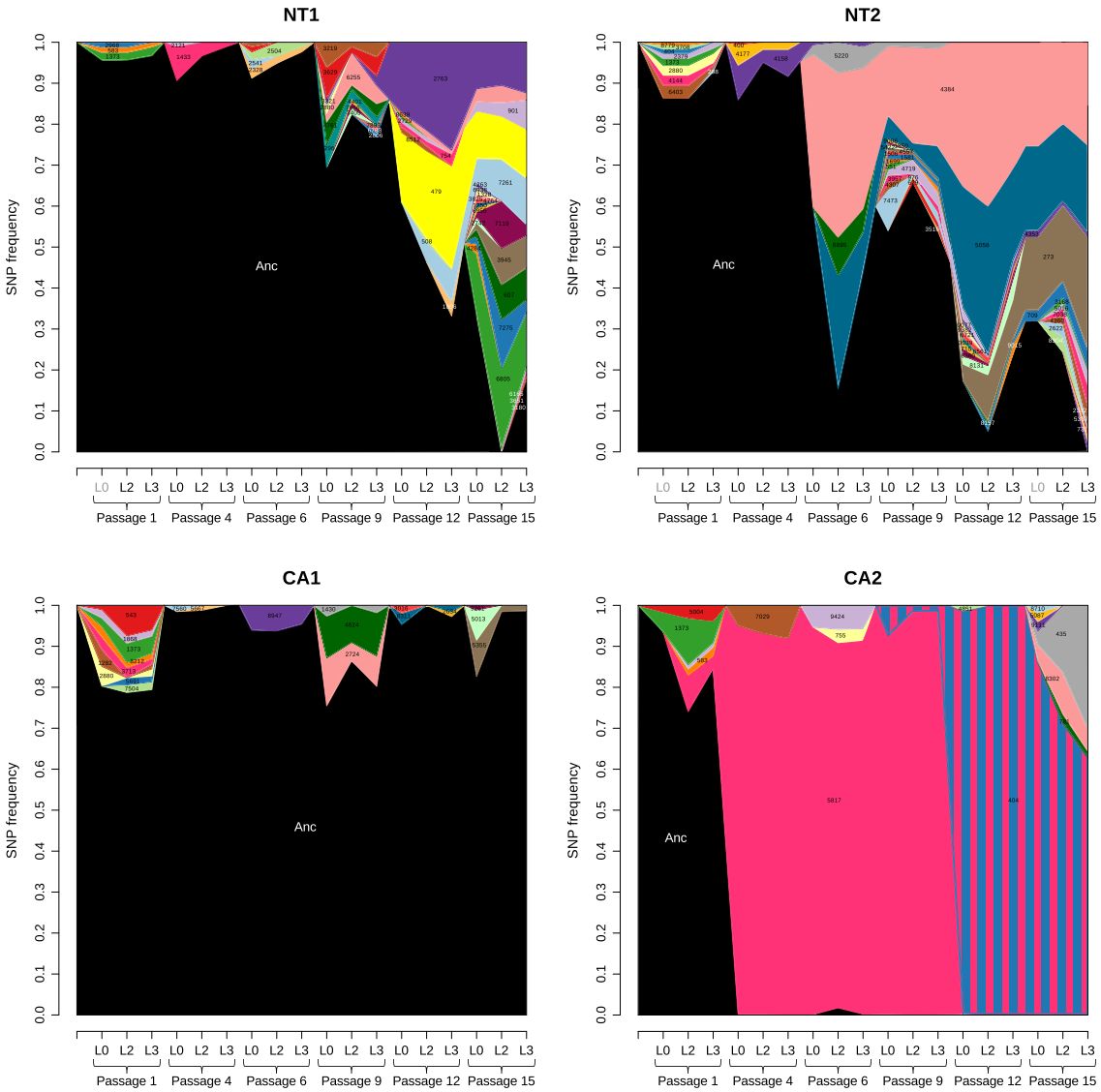

Fig. 4. Dynamics of genetic diversity in experimentally evolving TEV populations. The evolution of the frequency of different SNPs is illustrated using Muller plots (Barrick and Lenski 2013), which show the frequency of mutations over time as colored segments. Note that there is no data for the leaves indicated in light grey; NT1 passage 1 L0, NT2 passage 1 L0 and NT2 passage 15 L0.

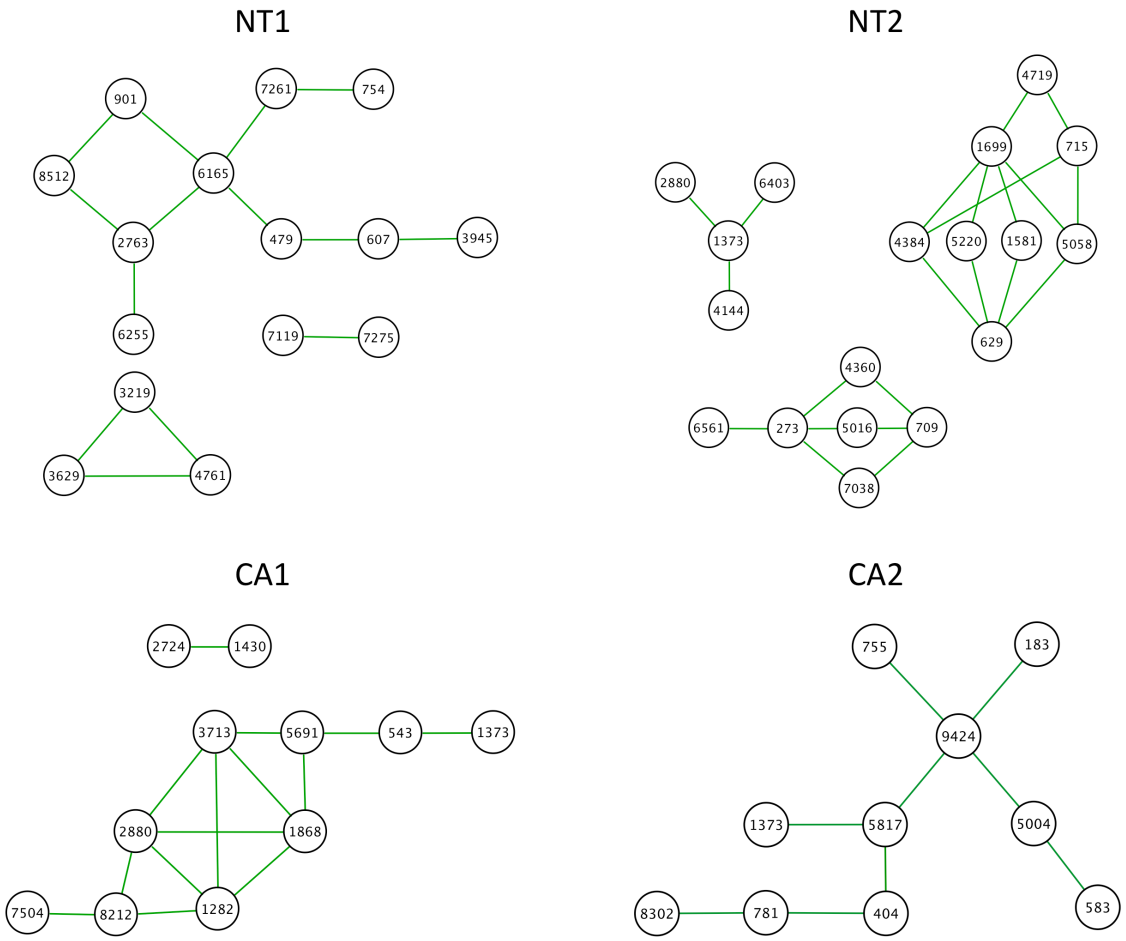

Fig. 5. Networks of co-occurring mutations on the different evolutionary lineages. Edges indicate that the temporal trajectories of two mutations across samples are significantly and positively correlated.

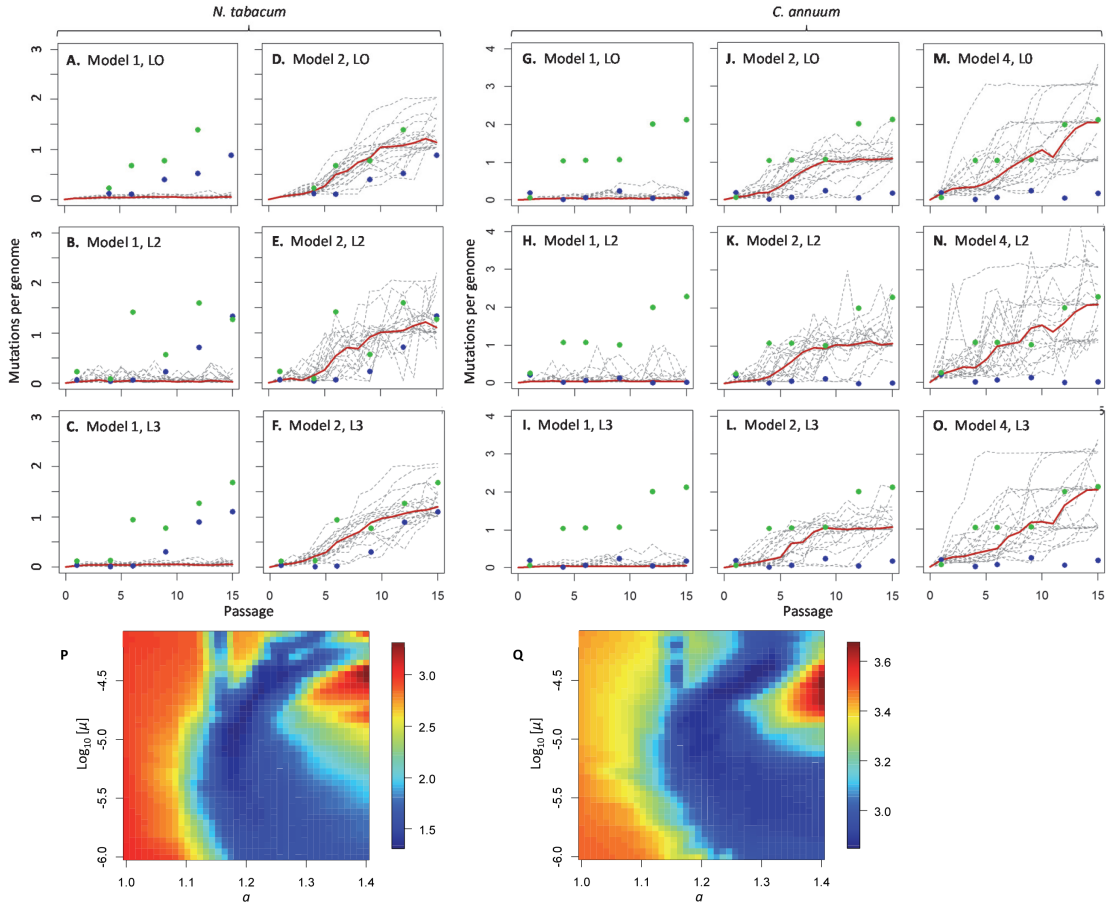

Fig. 6. Results obtained with simulation model of TEV molecular evolution are shown for the *N. tabacum* data (Panels A-F, P) and *C. annuum* (Panels G-O, Q). In panels A-O, simulation results for virus populations in individual leaves are shown (panels A, D, G, J, and M: L0; panels B, E, H, K, and N: L2; panels C, F, I, L, and O: L3). On the abscissae is the number of serial passages, whereas on the ordinates the mean number of mutations accumulated in the virus population are given. Blue and green circles are the data of replicates 1 and 2, respectively. Twenty individual simulations were run (dotted grey lines), and the mean of these simulations (solid red lines) is also given. For the *N. tabacum* data, simulations were first run for Model 1 ($\alpha = 1$, panels A-C), and then for Model 2 ($\alpha = 1.18$, panels D-F), where α is the maximum value of the DMFE. For these data, Model 2 is better supported than Model 1. Model 1 predicts maintenance of the ancestral sequence in all lineages, whereas Model 2 can predict the observed rate of mutation accumulation. In this instance Model 4 (not shown) is not better supported (table 3). For *C. annuum*, Model 1 (panels G-I) again poorly predicts the observed rate of mutation accumulation. Model 2 (panels J-L) predicts the rate better, whereas Model

4 – the best-supported model (table 3) – also predicts the tempo of mutation accumulation (panels M-O). For all models and both host species, the greatest amount of genetic drift is visible in L2, in which there is a strong bottleneck. In panels P and Q, the fit of the model for different combinations of α and μ is given for the *N. tabacum* and *C. annuum* data, respectively. The heat of the maps corresponds to values of the ln-transformed residual sum of squares (RSS), as indicated by the scale to the right of each map. What is surprising about model fit is the large range of μ values (3.16×10^{-6} - 2.00×10^{-5}) for which the model fits for *N. tabacum*, as long as an appropriate α value is chosen ($\alpha \approx 1.2$). For *C. annuum*, the results are similar, although the best model fit (dark blue regions) is more constrained to a region with high mutation rates.

Table 1. GLMM Analyses of the Viral Load Data.

Term	LRT	d.f	P	η_P^2
Intercept	1264.664	1	< 0.001	0.924
HOST	933.341	1	< 0.001	0.978
PASSAGE	642.955	6	< 0.001	0.417
HOST×*PASSAGE*	362.054	5	< 0.001	0.300
LINEAGE(*HOST*)	128.234	2	< 0.001	0.087
PASSAGE×*LINEAGE*(*HOST*)	555.992	10	< 0.001	0.057
LEAF(*PASSAGE*×*LINEAGE*(*HOST*))	1324.104	72	< 0.001	0.990

NOTE.— *LRT* is the value of the likelihood ratio test, *P* is its corresponding significance level and η_P^2 statistic that represents the proportion of the total variability attributable to a each factor in the model. See Material and Methods section for details.

Table 2. GLMM Analyses of the Virulence Data.

Term	***LRT***	**d.f**	***P***	η_P^2
Intercept	5.595	1	0.018	0.354
HOST	528.990	1	< 0.001	0.990
PASSAGE	26.786	14	0.021	0.434
HOST×*PASSAGE*	57.955	13	< 0.001	0.636
LINEAGE(*HOST*)	10.198	2	0.006	0.224
PASSAGE×*LINEAGE*(*HOST*)	34.169	27	0.161	0.076

NOTE.— *LRT* is the value of the likelihood ratio test, *P* is its corresponding significance level and η_P^2 statistic that represents the proportion of the total variability attributable to a each factor in the model. See Material and Methods section for details.

Table 3. Model Fitting Results.

Host	Model	Parameter estimates	*NLL*	*AIC*	Δ*AIC*	*AW*
N. tabacum	1	-	64.194	128.388	95.396	0
	2	$\alpha = 1.18$	15.496	32.992	-	0.616
	3	$\mu = 2.51\times10^{-5}$	60.500	122.999	90.007	0
	4	$\alpha = 1.20$; $\mu = 2.00\times10^{-5}$	14.967	33.934	0.942	0.384
C. annuum	1	-	810.896	1621.792	1516.656	0
	2	$\alpha = 1.22$	54.160	110.320	5.184	0.070
	3	$\mu = 2.51\times10^{-5}$	422.575	847.150	742.014	0
	4	$\alpha = 1.28$; $\mu = 3.16\times10^{-5}$	50.568	105.136	-	0.930

NOTES.— *NLL* is the negative log likelihood, a measure of model fit; *AIC* is the Akaike information criterion; Δ*AIC* is the difference in *AIC* between a given model and the best supported model; *AW* is the Akaike weight, a measure of the relative support for the model. Parameters α and μ are the maximum value at which the Weibull DMFE was truncated and the mutation rate per base and per generation, respectively.

Table 4. Fixed Model Parameters for the Evolution Model

Parameter	Description	Value	Source
λ_3	Population bottleneck size in the mechanically inoculated tobacco third leaf	417.0	This study
λ_3	Population bottleneck size in the mechanically inoculated pepper third leaf	27.2	This study
λ_5	Bottleneck for systemic viral movement to the fifth leaf	5.83	Tromas et al. (2014)
λ_6	Bottleneck for systemic viral movement to the sixth leaf	107.0	Tromas et al. (2014)
σ^a	Shape parameter for negative binomial distribution of founders	11.71	This study
σ^b	Shape parameter for negative binomial distribution of founders	8.70	This study
κ	Carrying capacity for a logistic growth of the number of viral genomes	3.99×10^8	Zwart et al. (2012)
r_0	Initial growth rate for a logistic growth of the number of viral genomes	2.303	Zwart et al. (2012)

β	Constant for cell-to-cell transmission of infection with a leaf for SI model	3.16×10	Tromas et al. (2014)
χ_5	Constant for between-leaf transmission to the fifth leaf for SI model	8.32×10	Tromas et al. (2014)
χ_6	Constant for between-leaf transmission to the sixth leaf for SI model	3.72×10	Tromas et al. (2014)
I_0	Fraction of infected cells in the inoculated leaf at the time of inoculation	1.23×10	Tromas et al. (2014)
ψ_3	Spatial aggregation of infected cells in third leaf.	0.096	Tromas et al. (2014)
ψ_5	Spatial aggregation of infected cells in fifth leaf.	0.019	Tromas et al. (2014)
ψ_6	Spatial aggregation of infected cells in sixth leaf.	0.221	Tromas et al. (2014)
L	Virus genome length	9539	Carrasco et al. (2007)
ξ	Fraction of non-lethal single-nucleotide mutations	0.951	Carrasco et al. (2007)
γ	Shape parameter of the Weibull distribution describing the DMFE	1.487	Carrasco et al. (2007)[c]

τ	Scale parameter of the Weibull distribution describing the DMFE	0.971	Carrasco et al. (2007)[c]
μ[d]	Mutation rate per base per generation	4.75×10^{-9}	Tromas and Elena (2010)

NOTES.— [a] Estimates for *N. tabacum*. [b] Estimates for *C. annuum*. [c] Estimated here but based on the data from this previous work. [d] Estimates of μ were used only in Models 1 and 2.