Effect of Host Species on the Distribution of Mutational Fitness Effects for an RNA Virus

Jasna Lalić1, José M. Cuevas1, Santiago F. Elena1,2*

1 Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas–Universidad Politécnica de Valencia, Valencia, Spain, 2 The Santa Fe Institute, Santa Fe, New Mexico, United States of America

Abstract

Knowledge about the distribution of mutational fitness effects (DMFE) is essential for many evolutionary models. In recent years, the properties of the DMFE have been carefully described for some microorganisms. In most cases, however, this information has been obtained only for a single environment, and very few studies have explored the effect that environmental variation may have on the DMFE. Environmental effects are particularly relevant for the evolution of multi-host parasites and thus for the emergence of new pathogens. Here we characterize the DMFE for a collection of twenty single-nucleotide substitution mutants of Tobacco etch potyvirus (TEV) across a set of eight host environments. Five of these host species were naturally infected by TEV, all belonging to family Solanaceae, whereas the other three were partially susceptible hosts belonging to three other plant families. First, we found a significant virus genotype-by-host species interaction, which was sustained by differences in genetic variance for fitness and the pleiotropic effect of mutations among hosts. Second, we found that the DMFEs were markedly different between Solanaceae and non-Solanaceae hosts. Exposure of TEV genotypes to non-Solanaceae hosts led to a large reduction of mean viral fitness, while the variance remained constant and skewness increased towards the right tail. Within the Solanaceae hosts, the distribution contained an excess of deleterious mutations, whereas for the non-Solanaceae the fraction of beneficial mutations was significantly larger. All together, this result suggests that TEV may easily broaden its host range and improve fitness in new hosts, and that knowledge about the DMFE in the natural host does not allow for making predictions about its properties in an alternative host.

Citation: 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(11): e1002378. doi:10.1371/journal.pgen.1002378

Introduction

The emergence of new epidemic viruses is a critical issue for public health and economic welfare [1–7]. Virus emergence is a complex, multilevel problem that results from a combination of ecological and genetic factors [3–8]. The increasing threats imposed by emerging and re-emerging viruses make it even more urgent to predict whether and when virus populations replicating in their reservoir hosts will acquire the ability to successfully infect individuals of a new host species, adapt to it and, eventually, turn into an epidemic. To make such predictions we must first identify the factors determining why some viruses, like Hepatitis C virus, Human immunodeficiency virus type 1 (HIV-1), Influenza A virus or Cucumber mosaic virus have been so successful in causing pandemics whereas other viruses such as SARS coronavirus, Ebola virus, Hantavirus, or Cocoa swollen shoot disease virus produced outbreaks limited in time and space. A pre-requisite for viral emergence is the existence of standing genetic variation within the reservoir host that enables successful virus replication within naive hosts after spillover by chance [2,3,8]. As a first approximation, and neglecting the effect of genetic drift, the frequency of these host-range mutants in the reservoir population will directly depend on the equilibrium between (i) the rate at which they are produced and (ii) the fitness effects they may have in the reservoir host. If host-range mutations are deleterious in the reservoir host, their frequency will be low and thus the likelihood of emergence will be low as well, whereas if they are neutral or perhaps even beneficial, their frequency will increase, which will in turn increase the chances of emergence. RNA viruses are characterized not only by extremely high mutation rates [9], but also by short generation times and large population sizes [3,8]. For these reasons RNA viruses have a high evolutionary potential and are over-represented among emerging viruses. Regarding fitness effects, extensive data have shown that host-range mutants have high fitness in the new host but pay fitness penalties in the reservoir host [10–13]. This fitness trade-offs should also preclude the evolution of generalist, multi-host viruses [11,13–15]. Antagonistic pleiotropy is often called to explain the existence of such fitness trade-offs [11,13]. However, an alternative, although not mutually exclusive, mechanism promoting host specialization is the accumulation of neutral mutations in the genes that are not necessary in a given host but are essential in alternative hosts, making these mutations deleterious in the alternative host environment [14,15]. Therefore, to predict the probability of a virus to infect new hosts, it is necessary to characterize the distribution of mutational fitness effects (DMFE) on its primary hosts as well as on potential new hosts. DMFE across hosts show the fraction of all possible mutations that may be beneficial in new hosts and reveal their...
Not much is known about the extent and underlying form of $G \times E$ interactions. Previous attempts to answer these questions suffer from one or another weakness (e.g., non-random samples of mutations taken from standing variation formerly filtered by selection, unknown number of mutations, traits of unclear relationship with fitness, etc.) [26]. To overcome these problems, Remold and Lenski [26] proposed using a collection of mutant genotypes that differ from the wildtype in a single and well defined mutation. Mutational fitness effects should further be evaluated in environments not previously experienced by the organism. By applying this simple experimental design to the bacterium *Escherichia coli*, these authors found that $G \times E$ interactions were quite common even for genotypes that differed by only one mutation and across environments that differed in a single component.

In this study, we sought to study how different host species affect the parameters describing the DMFE for a plant RNA virus, *Tobacco etch potyvirus* (TEV). Furthermore, we were interested in testing whether single point mutations are sufficient to give rise to $G \times E$ interactions in simple and compacted RNA genomes. To do so, we randomly selected 20 single-nucleotide substitution mutants from the collection previously described in Carrasco et al. [20].

Then, we quantified the absolute fitness (i.e., Malthusian growth rate) of all these mutants in eight different host species and characterized the parameters describing the DMFE and how they varied across hosts. Furthermore, we evaluated the amount of observed variability that was explained by genetic differences among viral genotypes, by differences among host species and, more interestingly, by the non-linear interaction between these two factors (e.g., the genotype-by-environment variance). In nature, TEV infects five of these hosts (*Nicotiana tabacum*, *Nicotiana benthamiana*, *Solanum lysipicum*, *Capsicum annuum*, and *Datura stramonium*), all belonging to the same plant family, the Solanaceae. The other three species are not TEV natural hosts, although they are experimentally susceptible to systemic infection. They belong to two plant families, the Asteraceae (*Helianthus annuus*) and the Amaranthaceae (*Gomphrena globosa* and *Spinacea oleracea*). Both the Solanaceae and the Asteraceae are within the Asterids, while the Amaranthaceae are not [27].

## Results

### Characterization of the DMFE on different hosts

For this study, we have used a collection of 21 TEV genotypes (20 mutants plus the wildtype) drawn from a larger collection of mutants obtained by Carrasco et al. [20]. Each mutant contained a single nucleotide change whose position and substitution were chosen at random. In 14 cases, the mutation resulted in an amino acid substitution whose position and substitution were chosen at random. This resulted in an amino acid substitution (Table 1). Our set of mutants consisted in changes among viral genotypes, by differences among host species and, and mutations that were not randomly dispersed throughout the viral genome but concentrated in a regulatory non-coding region.

The situation that we have just described in the context of emerging viruses is a particular case of a more general biological problem: the extent to which a phenotype (here viral fitness) is determined by the interaction between the genotype and the environment (here the host species), or the $G \times E$ interaction [23]. Understanding how genotype and environment interact to determine the phenotype and fitness has been a central aim of ecology, genetics, and evolution. Therefore, it should also be central for the epidemiology and evolution of infectious diseases; even more so in light of the reasons given above. The fate of genetic variation in populations depends on the form of the $G \times E$ interactions [24, 25] and, for instance, a change in the rank order of genotypic fitness in different environments may support a balanced polymorphism [25]. Despite this centrality, not much is
This finding is particularly interesting when observed that the fitness higher than the mean is significantly heavier than the negative tail. This means that the tail of the distribution containing fitness effects only for non-Solanaceae values; a comparison of absolute fitness values between Solanoideae and non-Solanaceae species among hosts ($\chi^2 = 163.262$, 21 d.f., $P<0.001$). However, this heterogeneity is entirely driven by the differences among TEV absolute fitness in Solanaceae hosts ($\chi^2 = 96.161$, 12 d.f., $P<0.001$), but not among non-Solanaceae hosts ($\chi^2 = 0.891$, 6 d.f., $P=0.989$). Indeed, if a new contingency table is constructed by grouping hosts into Solanaceae and non-Solanaceae, a significant heterogeneity is observed among the two host classes ($\chi^2 = 37.884$, 3 d.f., $P<0.001$). These results are explained by the shift from more neutral mutations in the two Nicotianae towards more beneficial and lethal in the three Solanaceae, while the three non-Solanaceae species had similar counts of neutral and beneficial mutations.

To further expand the analyses of the data shown in Figure 1, we compared the absolute fitness of each mutant to that of the wildtype TEV on each host using the bootstrap method described in [18]. Based on the bootstrap results, mutations were classified into lethal, deleterious (i.e., significantly smaller absolute fitness than wildtype), neutral, and beneficial (i.e., significantly larger absolute fitness than wildtype) on each alternative host (Table 2). The analysis of this contingency table shows that there is a significant heterogeneity in the distributions of discrete mutational classes among hosts ($\chi^2 = 163.262$, 21 d.f., $P<0.001$). However, this heterogeneity is entirely driven by the differences among TEV absolute fitness in Solanaceae hosts ($\chi^2 = 96.161$, 12 d.f., $P<0.001$), but not among non-Solanaceae hosts ($\chi^2 = 0.891$, 6 d.f., $P=0.989$). Indeed, if a new contingency table is constructed by grouping hosts into Solanaceae and non-Solanaceae, a significant heterogeneity is observed among the two host classes ($\chi^2 = 37.884$, 3 d.f., $P<0.001$). These results are explained by the shift from more neutral mutations in the two Nicotianae towards more beneficial and lethal in the three Solanaceae, while the three non-Solanaceae species had similar counts of neutral and beneficial mutations.

Interestingly, neutral and non-neutral cases were evenly distributed among synonymous and nonsynonymous mutations for all hosts (Fisher’s exact test, $P=0.131$ in all hosts). In recent years, increasing evidence supports the notion that, for compacted RNA genomes, synonymous mutations are not necessarily neutral mutations [20,28]. This observation is most likely due to the overlapping nature of many viral genes, the existence of secondary RNA structures essential for regulating gene expression, the adaptation to the host’s codon usage bias, and the pressure for evading RNAi-based host defenses.

The above classification of viable mutants into deleterious, neutral or beneficial depends on whether their fitness values

| Genotype     | Protein | Location | Nucleotide substitution | Amino acid change | Polarity change |
|--------------|---------|----------|-------------------------|------------------|----------------|
| PC2          | P1      | 158      | U→G                     | F→C              | apolar→polar   |
| PC6          | P1      | 375      | A→G                     | L→M              |                |
| PC7          | P1      | 475      | A→C                     | K→Q              | basic→polar    |
| PC12         | P1      | 872      | A→C                     | M→L              |                |
| PC19         | HC-Pro  | 1503     | A→G                     | synonymous       |                |
| PC22         | HC-Pro  | 1655     | A→G                     | N→S              |                |
| PC26         | HC-Pro  | 2119     | A→U                     | synonymous       |                |
| PC40         | P3      | 3238     | T→C                     | synonymous       |                |
| PC41         | P3      | 3406     | C→A                     | Q→K              | polar→basic    |
| PC44         | P3      | 3468     | U→G                     | synonymous       |                |
| PC49         | CI      | 4418     | G→C                     | S→T              |                |
| PC60         | CI      | 5349     | U→C                     | synonymous       |                |
| PC63         | 6K2     | 5582     | A→G                     | K→R              |                |
| PC67         | Nla-Vpg | 6012     | U→G                     | I→M              |                |
| PC69         | Nla-Vpg | 6044     | C→A                     | T→N              |                |
| PC70         | Nla-Vpg | 6197     | U→G                     | M→R              | apolar→basic   |
| PC72         | Nla-Vpg | 6251     | U→C                     | F→S              | apolar→polar   |
| PC76         | Nla-Pro | 6519     | U→C                     | synonymous       |                |
| PC83         | Nib     | 7315     | A→G                     | I→V              |                |
| PC95         | Nib     | 8501     | A→C                     | E→A              | acid→polar     |

Table 1. TEV genotypes used in this study and some of their properties.
Figure 1. DMFEs across different host species. Host species belong to the taxonomic families Solanaceae, Asteraceae and Amaranthaceae. The first two families belong to the Asterids class. In nature, TEV is found infecting members of the Solanaceae family. The ancestral isolate used in this study was obtained from and subsequently passed in N. tabacum plants. Lethal mutations (which have a Malthusian fitness of $\approx 0$) are indicated in the histograms with $< -0.1$ fitness values. The vertical dashed lines represent the fitness value of the wildtype genotype in each host.

doi:10.1371/journal.pgen.1002378.g001

Table 2. Parameters describing the DMFE shown in Figure 1 and number of mutations classified as lethal, deleterious, neutral, and beneficial on each host.

|                  | Mean | Median | Std. deviation | Skewness | Kurtosis | Lethal | Deleterious | Neutral | Beneficial |
|------------------|------|--------|----------------|----------|----------|--------|-------------|---------|------------|
| N. tabacum       | 0.280| 0.283  | 0.016          | -1.974** | 4.608**  | 0      | 6           | 14      | 0          |
| N. benthamiana   | 0.267| 0.277  | 0.050          | -3.949***| 16.879***| 0      | 10          | 10      | 0          |
| D. stramonium    | 0.307| 0.322  | 0.040          | -1.566** | 1.364    | 2      | 15          | 3       | 0          |
| C. annuum        | 0.200| 0.260  | 0.116          | -1.037** | -0.389   | 0      | 9           | 11      | 0          |
| S. lycopersicum  | 0.338| 0.349  | 0.029          | -0.768   | 0.062    | 8      | 0           | 2       | 10         |
| H. annuus        | 0.026| 0.020  | 0.043          | 0.527    | 0.579    | 0      | 0           | 15      | 5          |
| G. globosa       | 0.019| 0.010  | 0.041          | 0.997    | 0.561    | 0      | 0           | 17      | 3          |
| S. oleracea      | -0.018| -0.039 | 0.053          | 1.479**  | 1.915    | 0      | 0           | 17      | 3          |

*Test significance levels for skewness and kurtosis:
*0.05>P>0.01,
**0.01>P>0.001;
***P<0.001.
doi:10.1371/journal.pgen.1002378.t002
deviates significantly from that of the wildtype TEV in the bootstrap test. However, given the statistical uncertainties inherent to our measurements, it is difficult to distinguish between small-effect mutations and lack of fitness effects. For the *Solanaceae*, relative fitness values $< -0.03$ were generally significantly deleterious, whereas mutations were assigned to the beneficial class if they had relative fitness $> 0.05$ as in *S. lycopersicum*, although the threshold for *C. annuum* rose up to $> 0.2$. For the non-*Solanaceae*, in general, mutations were considered as beneficial if they had relative fitness values $> 0.05$. However, since the concept of neutrality depends on the effective population size [29], modeling the continuous DMFE rather than their discretization, at length, is to be more informative. In the next section we will address this problem.

Failed inoculation experiments and lethal mutations produce the same apparent result: a lack of viral accumulation in the inoculated plants. To rule out the possibility that the putative lethal mutations observed in *D. stramonium* and *S. lycopersicum* are just a succession of failed inoculation experiments, we applied the following statistical argument. First, we evaluated our rate of failure to produce an infection when starting the experiment with viruses that are viable in each host species. In the case of *D. stramonium*, two mutants were assigned to the class of lethals. Of 171 *D. stramonium* plants inoculated with viable viruses, 72 plants were infected and thus the failure rate was $1 - 72/171 = 0.579$ per inoculation event. After nine trials (corresponding to the number of replicates per mutant and per host species), the probability of failing all cases should be $0.579^9 = 0.007$. Therefore, in a sample of 21 genotypes, we expect less than one case ($21 \times 0.007 = 0.13$) to be erroneously assigned to the category of lethal mutations. Similarly, in the case of *S. lycopersicum*, where eight mutants were putatively lethal, 72 out of 117 plants inoculated with viable viruses were infected, which represents a failure rate of 0.385 per inoculation experiment. From this, we expect $21 \times 0.385^9 = 0.004$ much less than one case to be classified as lethal but resulting from multiple inoculation failures. Therefore, on these grounds, we are confident that the mutations classified as lethal on these two hosts were really so.

**Fit of empirical DMFE to theoretical probability density functions**

Next, we sought to determine which of several competing statistical models better describes the observed DMFEs. Following previous analyses of the DMFE for RNA viruses [16,18,19,20], we evaluated the goodness-of-fit of distributions sharing the property of asymmetry and with heavy tails to the empirical DMFEs observed in each host. Lethal mutations were excluded from the analyses. The probability density functions (pdf) tested were: Exponential, Gaussian, Gamma, Beta, Log-normal, Laplace, Pareto, and Weibull. Nonlinear regression techniques were used to fit models to the data. Table 3 shows the best-fitting model for each host and the relevant parameters describing each distribution, as well as the statistics measuring the goodness of fit (Akaike’s weight and $R^2$). The Weibull pdf was the model that better described the DMFEs measured in *N. tabacum, N. benthamiana, D. stramonium, S. lycopersicum*, and *G. globosa*. A Weibull pdf is described by two parameters, the scale $\lambda$ and the shape $\kappa$, related to the expected value of the distribution as $E(m) = \lambda \Gamma(1 + 1/\kappa)$, where $\Gamma(\cdot)$ is the gamma function evaluated at the given argument. However, the Aikake’s weight for this pdf is $< 0.95$ in all cases, suggesting that alternative models, or combinations of models, can still contribute to better describe the observed distributions. In the cases of *C. annuum* and *S. oleracea* the pdf that better explained the observed DMFEs were Laplace and Pareto, respectively. These two distributions are from the power-law family. In the case of the Laplace pdf, the expected fitness value is equal to the location parameter $E(m) = \mu$, whereas in the case of the Pareto, the expected value is $E(m) = \mu / (\alpha - 1)$, where $\alpha$ is the shape parameter.

**Table 3. Probability distribution models that best describe the observed DMFEs on each host (excluding lethal mutations).**

| Model          | Parameter estimates$^a$ | Expected fitness | Akaike’s weight$^b$ | $R^2$ | ER (to second best model)$^c$ |
|----------------|-------------------------|------------------|---------------------|-------|-----------------------------|
| *N. tabacum*   | Weibull                 | scale $\lambda = 0.286 \pm 0.000$ | 0.286              | 0.706 | 0.988 | 7.675 (Normal)             |
|                |                         | shape $\kappa = 33.138 \pm 1.433$ |                     |       |     |                             |
| *N. benthamiana* | Weibull                | scale $\lambda = 0.282 \pm 0.000$ | 0.274              | 0.917 | 0.989 | 28.924 (Normal)            |
|                |                         | shape $\kappa = 20.371 \pm 0.840$ |                     |       |     |                             |
| *D. stramonium* | Weibull                | scale $\lambda = 0.323 \pm 0.002$ | 0.311              | 0.643 | 0.849 | 4.990 (Laplace)            |
|                |                         | shape $\kappa = 12.992 \pm 2.317$ |                     |       |     |                             |
| *C. annuum*    | Laplace                 | location $\mu = 0.253 \pm 0.010$ | 0.223              | 0.521 | 0.842 | 5.495 (Weibull)            |
|                |                         | scale $\beta = 0.104 \pm 0.019$ |                     |       |     |                             |
| *S. lycopersicum* | Weibull               | scale $\lambda = 0.324 \pm 0.004$ | 0.300              | 0.479 | 0.873 | 2.514 (Normal)            |
|                |                         | shape $\kappa = 5.774 \pm 0.785$ |                     |       |     |                             |
| *H. annuus*    | Laplace                 | location $\mu = 0.067 \pm 0.001$ | 0.020              | 1.000 | 0.992 | 3721.827 (Normal)         |
|                |                         | scale $\beta = 0.032 \pm 0.014$ |                     |       |     |                             |
| *G. globosa*   | Weibull                 | scale $\lambda = 0.058 \pm 0.001$ | $-0.322$           | 0.400 | 0.992 | 1.159 (Beta)              |
|                |                         | shape $\kappa = 1.358 \pm 0.046$ |                     |       |     |                             |
| *S. oleracea*  | Pareto                  | threshold $c = 0.829 \pm 0.001$ | $-0.024$           | 0.997 | 0.930 | 553.409 (Laplace)         |
|                |                         | shape $\kappa = 22.189 \pm 1.493$ |                     |       |     |                             |

$^a \pm$ 1 SE of the estimated value.

$^b$The set of pdf models fitted and compared was: Exponential, Normal, Gamma, Beta, Log-normal, Laplace, Pareto, and Weibull.

$^c$ER: evidence ratio. In this case, ER measures how many times the best fitting model is more likely than the model ranked in second place.

doi:10.1371/journal.pgen.1002378.t003
The phylogenetic distance between natural and naive hosts influences the location and shape of DMFE

Next, we sought to evaluate whether the location and shape characteristics of the DMFE were affected by the genetic relationship between the hosts. Figure 2a shows that a statistically significant negative correlation (Spearman’s \(r_S = -0.798\), 6 d.f., \(P = 0.013\)) exists between the expected centrality parameter of the DMFE, \(E(m)\) (taken from Table 3), and the ranked phylogenetic distance of each host to the natural one; \(N.\ tabacum\). This negative correlation indicates that the average absolute fitness decreases as the host becomes more and more distant from the one to which the virus was originally adapted. By contrast, a significant positive correlation has been observed between the skewness of the DMFE and host’s phylogenetic distance from the natural one (Figure 2b; Spearman’s \(r_S = 0.877\), 6 d.f., \(P = 0.004\)). This result is congruent with the above observation that the skewness of the DMFE shifts from negative to positive as hosts become more phylogenetically distant from the natural one. The phylogenetic distance did not significantly affect the variance and kurtosis of the distributions (in both cases Spearman’s \(r_S \leq 0.569\), 6 d.f., \(P \geq 0.153\)).

Contribution of G×E interactions to TEV absolute fitness

Model I in Table 4 shows the GLM analysis of the absolute fitness data using host species and TEV genotype as random factors. First, there is a highly significant difference among TEV genotypes in their absolute fitness. This is in agreement with previous analyses of the larger collection of genotypes from which these 20 were drawn [20]. However, only \(~4\%\) of total observed variability is explained by genetic differences among TEV genotypes. There is also a highly significant effect of the host species on viral fitness, which explains \(26\%\) of the observed variability in absolute fitness. Finally, and more interestingly from the perspective of predicting emerging viral infections by using information about fitness effects in natural hosts, the G×E interaction term is also highly significant, and explains \(67\%\) of the observed variability in absolute fitness. This significant interaction means that we cannot accurately predict a particular genotype’s absolute fitness in a given host from the main effects, thus adding an unpredictability component to viral emergence. Finally, it is worth noting that only \(2.76\%\) of the observed variance remained unexplained by the model and was used as error variance in the computation of the different variance components.

To account for the fact that hosts are not independent but phylogenetically related, we fitted a more complicated model to the data (Model II in Table 4). This alternative model treated the host species as a binary factor; belonging to one of two classes (Solanaceae vs. non-Solanaceae). Then, host species were nested within these two classes and the G×E component was evaluated by looking the significance of the interaction between hosts within classes and TEV genotype. This model has an appreciably lower \(AIC\) value than the Model I and thus should be taken as a better one, although the conclusions do not qualitatively depart from those reached from the simpler model (Model I): the genetic component only explains a minor fraction of observed fitness variance whereas most of it is explained by the G×E interaction term.

The Weibull pdf is the best descriptor in five out of eight host species. Hence, one may ask how good a descriptor it is for the three remaining hosts. In the case of \(C.\ annuum\), the Weibull was ranked as the second best fitting, performing only \(~5.5\) times worse than the Laplace pdf. For \(H.\ annuum\), the Weibull pdf ranked in third position, with an \(ER = 38609.153\), thus providing a much worse fit than the Laplace pdf. Finally, in the case of \(S.\ oleracea\) the Weibull pdf ranked in seventh position, with an \(ER = 190955.254\), indicative of a very poor fit compared to the best fitting Pareto pdf.

To make this analysis, we used an evidence ratio (\(ER\)) computed as the likelihood of the best model divided by the likelihood of the alternative model of interest [30]. The last column in Table 3 shows the \(ER\) values computed for models ranked in second place. The Weibull pdf is the best descriptor in five out of eight host species. Hence, one may ask how good a descriptor it is for the three remaining hosts. In the case of \(C.\ annuum\), the Weibull pdf ranked as the second best fitting, performing only \(~5.5\) times worse than the Laplace pdf. For \(H.\ annuum\), the Weibull pdf ranked in third position, with an \(ER = 38609.153\), thus providing a much worse fit than the Laplace pdf. Finally, in the case of \(S.\ oleracea\) the Weibull pdf ranked in seventh position, with an \(ER = 190955.254\), indicative of a very poor fit compared to the best fitting Pareto pdf.

The Weibull pdf is the best descriptor in five out of eight host species. Hence, one may ask how good a descriptor it is for the three remaining hosts. In the case of \(C.\ annuum\), the Weibull was ranked as the second best fitting, performing only \(~5.5\) times worse than the Laplace pdf. For \(H.\ annuum\), the Weibull pdf ranked in third position, with an \(ER = 38609.153\), thus providing a much worse fit than the Laplace pdf. Finally, in the case of \(S.\ oleracea\) the Weibull pdf ranked in seventh position, with an \(ER = 190955.254\), indicative of a very poor fit compared to the best fitting Pareto pdf.

Next, we sought to evaluate whether the location and shape characteristics of the DMFE were affected by the genetic relationship between the hosts. Figure 2a shows that a statistically significant negative correlation (Spearman’s \(r_S = -0.798\), 6 d.f., \(P = 0.013\)) exists between the expected centrality parameter of the DMFE, \(E(m)\) (taken from Table 3), and the ranked phylogenetic distance of each host to the natural one; \(N.\ tabacum\). This negative correlation indicates that the average absolute fitness decreases as the host becomes more and more distant from the one to which the virus was originally adapted. By contrast, a significant positive correlation has been observed between the skewness of the DMFE and host’s phylogenetic distance from the natural one (Figure 2b; Spearman’s \(r_S = 0.877\), 6 d.f., \(P = 0.004\)). This result is congruent with the above observation that the skewness of the DMFE shifts from negative to positive as hosts become more phylogenetically distant from the natural one. The phylogenetic distance did not significantly affect the variance and kurtosis of the distributions (in both cases Spearman’s \(r_S \leq 0.569\), 6 d.f., \(P \geq 0.153\)).

Contribution of G×E interactions to TEV absolute fitness

Model I in Table 4 shows the GLM analysis of the absolute fitness data using host species and TEV genotype as random factors. First, there is a highly significant difference among TEV genotypes in their absolute fitness. This is in agreement with previous analyses of the larger collection of genotypes from which these 20 were drawn [20]. However, only \(~4\%\) of total observed variability is explained by genetic differences among TEV genotypes. There is also a highly significant effect of the host species on viral fitness, which explains \(26\%\) of the observed variability in absolute fitness. Finally, and more interestingly from the perspective of predicting emerging viral infections by using information about fitness effects in natural hosts, the G×E interaction term is also highly significant, and explains \(67\%\) of the observed variability in absolute fitness. This significant interaction means that we cannot accurately predict a particular genotype’s absolute fitness in a given host from the main effects, thus adding an unpredictability component to viral emergence. Finally, it is worth noting that only \(2.76\%\) of the observed variance remained unexplained by the model and was used as error variance in the computation of the different variance components.

To account for the fact that hosts are not independent but phylogenetically related, we fitted a more complicated model to the data (Model II in Table 4). This alternative model treated the host species as a binary factor; belonging to one of two classes (Solanaceae vs. non-Solanaceae). Then, host species were nested within these two classes and the G×E component was evaluated by looking the significance of the interaction between hosts within classes and TEV genotype. This model has an appreciably lower \(AIC\) value than the Model I and thus should be taken as a better one, although the conclusions do not qualitatively depart from those reached from the simpler model (Model I): the genetic component only explains a minor fraction of observed fitness variance whereas most of it is explained by the G×E interaction term.
The causes of $G \times E$

A significant $G \times E$ interaction can be produced by two non-mutually exclusive mechanisms [26]. First, pleiotropic effects may change the rank order of mutations across environments (e.g., a mutation beneficial in one environment may not be so in an alternative one). Second, while still retaining the rank order of fitness effects, $G \times E$ can also be generated by altering the genetic component of phenotypic variance ($\sigma^2_G$) across hosts. To evaluate the contribution of these two mechanisms to the observed $G \times E$, we run two different analyses.

As a first statistical test, we computed Spearman’s rank correlation coefficients between absolute fitness values in the primary host $N. \ tabacum$ and the values estimated on each alternative host (Figure 3). Lethal mutations were assigned to the lowest rank. A negative correlation would indicate negative or antagonistic pleiotropy (e.g., mutations change the strength and sign of their effects on different hosts), whereas a positive correlation would indicate positive pleiotropy. Interestingly, the correlations were positive for all the Solanaceae hosts (although only reached significance in two cases, $N. \ benthamiana$ and $D. \ stramonium$). By contrast, for the three non-Solanaceae hosts the correlation coefficients had negative non-significant values. We used the frequency of discrete mutational signs on each host class to construct a contingency table, and applied a Fisher’s exact test to confirm that the difference in correlation signs among host classes was significant ($P=0.029$) despite the small sample size. Furthermore, the shift from negatively skewed DMFE (excess of deleterious effects) in the Solanaceae to positively skewed distributions (excess of beneficial effects) in the non-Solanaceae described above is also consistent with antagonistic pleiotropy. Therefore, from these analyses we concluded that antagonistic pleiotropy contributed to generate $G \times E$ when the new host species are phylogenetically distant from the natural host (i.e., outside the plant family), but not when host species belong to the same family. Nevertheless, this conclusion needs to be qualified because the most extreme cases of antagonistic pleiotropy are mutations that indicate the percentage of total phenotypic variance explained by genetic differences among TEV genotypes. For the five Solanaceae hosts, $\sigma^2_G$ ranged from 0.051 to 0.115, with an average value of 0.083, and $\sigma^2_G$ explaining >95% of the observed phenotypic variance. By contrast, $\sigma^2_G$ within the non-Solanaceae hosts was
Figure 3. Relationship between fitness in N. tabacum and in the seven alternative hosts. Spearman’s non-parametric correlation coefficients and their statistical significance are shown above each plot. The non-parametric test was chosen given its robustness against extreme data points. Dashed lines represent the fitness of the wildtype TEV in the corresponding hosts. The solid lines are only inserted to illustrate the overall trend. 
doi:10.1371/journal.pgen.1002378.g003

significantly smaller (Mann-Whitney test, \( P = 0.036 \)), with an average value of \( \sim 0.002 \). Besides, for these hosts only ca. 25% of phenotypic variance for absolute fitness was explained by genetic differences among TEV mutants. Henceforth, from these analyses we conclude that changes in genetic variance for absolute fitness contributed to the observed \( G \times E \) only when comparing phylogenetically distant hosts.

All together, these results suggest that \( G \times E \) arises from the combined effect of antagonistic pleiotropy and reductions in genetic variance associated to the shift from hosts that belong to the same family as the natural host to hosts that do not belong to this family.

Discussion

Changes in DMFE and the likelihood of crossing the species barrier

New emerging epidemic viruses represent one of the most serious threats to human, animal and crops health [1–8]. The problem of viral emergence is complex and depends on the interaction between host’s genetics, vectors’ abundance, ecology, and virus evolvability. Predicting the potential of a virus to spillover from its natural host reservoir to few individuals of a new host species and successfully establish a productive infection that will trigger a new epidemic seems an insurmountable problem. However, from the perspective of evolutionary genetics, the problem can be simplified by considering that the fate of the viral population entering into the new host depends, in first instance, on whether it contains genetic variants with a positive fitness value.

In other words, a pre-requisite for predicting the ability of a virus to expand its host range is to have information about the distribution of fitness effects associated to mutations (DMFE) across all possible hosts. In this study, we have characterized DMFE across a set of hosts for the plant virus TEV. The host species selected widely ranged in their degree of genetic relatedness with the natural host, N. tabacum: from very close relatives (members of the same genus) to members of other genera within the same family, and finally, to species belonging to different families within the same class or even to different classes.

We found that the central parameter of the DMFE shifted towards smaller values as the phylogenetic distance of each host from tobacco increased (Figure 2a). The distributions did not just displace; they also changed in shape, moving most of the probability mass from the negative to the positive tails. This means that, on average, the absolute fitness of TEV decreased as hosts became more different from the natural one. However, if the fitness of individual mutant genotypes is expressed relative to wildtype virus, the change in shape means that the number of (conditional) beneficial mutations increases as hosts become more phylogenetically distant from tobacco. This suggests that the number of mutations that may potentially expand TEV host range is large. A similar abundance of host-range mutants was also observed for phage \( \phi 6 \) [12]. In this case, the mutations were concentrated in the P3 gene that encodes for the protein responsible for attaching the virion to the bacterial pili. However, in our case, host-range mutations do not concentrate in any particular gene but were scattered along the genome. Notably, Gaussian fitness landscape models [31] predict an increase in the proportion of beneficial mutations under stressful conditions (here represented by those hosts in which absolute fitness was dramatically reduced).

The shape of DMFE is a critical component of many mathematical models of evolutionary dynamics, including the molecular clock, the rate of genomic contamination by Muller’s ratchet, the maintenance of genetic variation at the molecular level, and the evolution of sex and recombination [32]. In more practical terms, characterizing the shape of DMFE is essential for understanding the nature of quantitative genetic variation, here including complex human diseases as well as pathogens virulence [32]. Therefore, it is not surprising that much effort has been recently invested in characterizing the DMFE for many organisms (reviewed in [32]), including several RNA and DNA viruses. Despite differences in the genetic material of these viruses, their sizes and gene contents, the methodology applied has been similar in all cases, namely, generating collections of single-nucleotide substitutions mutants and then characterizing the fitness of each of these mutants relative to the non-mutated parental. In RNA viruses such as bacteriophage \( Qb \) [16], Vesicular stomatitis virus (VSV) [19] and TEV [20], over one third of mutations generated unviable viruses, whereas viable mutations reduced fitness, on average, by \( \sim 10\% \) [21]. Regarding the theoretical pdf that better explained these datasets, VSV fitness data conformed to a complex distribution combining a Log-normal and an Uniform pdfs, the original TEV larger dataset was best fitted by a Beta pdf (notice that in [20] fitness was measured as a relative value, which may justify the difference to the Weibull pdf conclusion reached here), and the \( Qb \) DMFE was well described by a Gamma pdf. In the case of DNA phages \( \phi X 174 \) [16] and \( f1 \) [17] the fraction of unviable mutations was lower (one fifth) but the average effect of viable mutations was almost identical to the one reported for RNA viruses [21]. \( \phi X 174 \) best fitting was to the Exponential pdf whereas for \( f1 \) the Log-Normal and the Weibull fitted equally well. Taken together, all these results suggested the existence of certain common rules: a large fraction of mutations are lethal or have a large negative fitness effects (displaying the fragility of viral genomes). In addition, DMFE for viruses are highly asymmetric.

Table 5. Maximum likelihood estimators for the variance components of absolute fitness estimated on each host (± variance of the estimator).

| Host species      | \( \sigma^2_x \times 10^{-4} \) | \( \sigma^2_e \times 10^{-4} \) | \( \rho^2 \) |
|-------------------|-------------------------------|-------------------------------|---------------|
| N. benthamiana    | 7.323                         | 16.052                        | 0.097         |
| D. stramonium     | 9.462 ± 0.097                 | 40.160 ± 0.006                | 0.959         |
| C. annuum         | 5.162 ± 0.028                 | 61.520 ± 0.015                | 0.894         |
| S. lycopersicum   | 11.475 ± 0.203                | 6.204 ± 0.000                 | 0.995         |
| H. annua          | 0.148 ± 0.000                 | 48.061 ± 0.006                | 0.236         |
| G. globoasa       | 0.190 ± 0.000                 | 47.062 ± 0.006                | 0.188         |
| S. oleracea       | 0.195 ± 0.000                 | 46.762 ± 0.005                | 0.294         |

*Lethal alleles were removed from the computations because they have absolute fitness \( \sim \infty \).
and can be reasonably well described by theoretical pdfs with heavy tails. In a recent study [33], the reason for this generality was grounded into the thermodynamic properties of protein folding, suggesting that the effect of mutations on protein folding and stability was a good explanation for the observed DMFEs. Despite being important for understanding the evolution of a virus in its natural host, these results were, even so, insufficient to understand the likelihood of a virus expanding its host range. Here, we have contributed to cover this lack of knowledge by describing the effect of changing hosts on the properties of DMFE. One of the most striking conclusions from our study is that the fraction of lethal, deleterious, neutral and beneficial mutations and hence the shape and location of the distributions, radically depends on the host in which the fitness effects of mutations is evaluated, and that this dependence is, itself, conditioned by the phylogenetic distance among hosts. Furthermore for host species belonging to the same family as the primary host, the Weibull pdf fitted best (or second to best for C. annuum) model to describe DMFE, although for hosts outside the family this model is the best only in one out of three cases (Table 3).

Martin and Lenormand [31] proposed three possible outcomes for the DMFEs measured in permissive vs. stressful environments: (i) conditional expression means that some mutations have a detectable fitness effect in some environments but are neutral in others, (ii) conditional average means that the average mutational effect differs between the two types of environments and (iii) conditional variance, meaning that variance in mutational effects changes between the two types of environments. In a survey of DMFE across benign and stressful environments for organisms as diverse as the fungi S. cerevisiae and C. neoformans, the nematode Caenorhabditis elegans, and the fruitfly Drosophila melanogaster, Martin and Lenormand [31] found that stressful conditions tend to inflate the variance of the DMFE while leaving the central value of the distributions almost unaffected. These results contrast with those reported here: for TEV, DMFE evaluated in stressful hosts (the non-Solanaceae) had lower average (Figure 2a) and more positive skewness (Figure 2b) than in permissive hosts (the Solanaceae), while no significant effects on variance were observed. Furthermore, we found that some mutations that were neutral in the natural host had reduced absolute fitness in alternative ones. Therefore, our data contain all three possible outcomes proposed by Martin and Lenormand [31], thus suggesting that their expectations were somewhat simplistic.

A compelling idea of the phylogenetic constraints for a virus jumping the host species barrier resides in the argument that the more closely related the primary host and the new host are, the greater are the chances for a successful spillover [34]. There are good mechanistic reasons that argue for it; if the ability to recognize and infect a host cell is important for cross-species transmission, then phylogenetically related species are more likely to share related cell receptors and defense pathways. However, others support the opposed view based on the observation that spillovers have occurred between hosts that can be either closely or distantly related, and no rule appears to predict the susceptibility of a new host [35]. Whether or not phylogenetic relatedness between reservoir and new hosts may be a factor for host switching, the rate and intensity of contact may be even more critical. Viral host switches between closely related species (e.g., species within the same genera) may also be limited by cross-immunity to related pathogens [2]; paraphrasing Holmes and Drummond [35] ‘although a species might be exposed to a novel pathogen, they might, through a combination of shared common ancestry and good fortune, already posses a sufficient immune response to prevent the infection from being established’. Our results shed some light into this debate: certainly the absolute fitness of a virus may be reduced when colonizing a new host, especially those distantly related ones, but the fraction of mutations that may be beneficial in this new host also increases with phylogenetic distance between the new host and the reservoir.

### Pleiotropy and changes in genetic variance as sources of G×E interactions

The existence of G×E interactions in determining fitness has been well established for many organisms, however, many of these studies used genotypes that differed in a large and unknown number of mutations [23,36–39], making unclear whether G×E depended on single plasticity genes or on the quantitative contribution of multiple genes. Furthermore, in many examples, these studies used genotypes sampled from natural populations and thus have been filtered out by natural selection. Interestingly, our data demonstrate that single random nucleotide substitutions are sufficient to produce a significant G×E interaction. Mutations involved in significant G×E were scattered along the genome and they were randomly chosen irrespective of their fitness effects, provided they were viable in the primary host N. tabacum. Thus, we can conclude that phenotypic plasticity of TEV is not associated to the expression of any particular gene but results from the contribution of different genes. The concordance of these results with those previously reported by Remold and Lenski [26] for the bacterium E. coli and using knockout mutations suggests that the contribution of individual mutations to G×E is a general norm. In the context of emerging viral infections, the existence of a significant G×E interaction means that by knowing the absolute viral fitness in the natural host informs us little about what it may be in an alternative one, thus minimizing our ability to predict which genetic variants may be relevant for expanding TEV host-range.

Two non-mutually exclusive explanations can be brought forward to explain the existence of G×E. A change in the rank order of mutational effects across hosts (i.e., pleiotropy) and a change in the magnitude of the genetic variance but without changing the rank order. The evolutionary implications for these two mechanisms are different. Changes in genetic variance imply that the relative influence of selection and drift on the fate of mutations depends on the host. Exposure to hosts where the genetic variance in absolute fitness effects is low minimizes the efficiency by which selection operates either removing deleterious alleles or fixing beneficial ones and thus enhances the role of drift. By contrast, changes in rank order imply that selection favor different mutations in different hosts thus driving to a balanced polymorphism and specialization. We have assessed the extent to which these two possibilities may contribute to the observed G×E and found that both indeed coexist. Antagonistic pleiotropy does not contribute significantly to G×E when the novel host is closely related to the natural one, however, it becomes an important factor when hosts are distantly related (Figure 3). Similarly, genetic variance for absolute fitness was similar within Solanaceae hosts, but approximately one order of magnitude smaller for hosts outside the Solanaceae. Therefore, we conclude that the observed G×E interaction can be explained both by antagonistic pleiotropy and by changes in the genetic component of variance. Previous studies with E. coli showed that G×E was mainly explained by changes in genetic variance but not by changes in the rank order of fitness effects across environments [26]. However, other authors found that the contribution of new mutations to G×E for fitness traits in D. melanogaster was mostly via antagonistic pleiotropy [40].

The significant positive pleiotropy observed between absolute fitness in the natural host N. tabacum and in two closely related
alternative ones (N. benthamiana and D. stramonium) suggests that mutations ameliorate aspects of the virus interaction with host factors that may be common to all three hosts but not to the other hosts. By contrast, the antagonistic pleiotropy observed between absolute fitness in N. tabaci and in the non-Solanaceae hosts suggests that TEV may be interacting with different host factors and that the improved interaction with tobacco may led to less efficient interaction with an orthologous factor, if available, in the alternative hosts. In this regard, many examples exist in the plant virology literature showing that host-range mutations have negative pleiotropic effects in the natural host (reviewed in [8,41]). An illustrative example is the interaction between the VPG protein of other potyviruses and the host translation initiation factor eIF4E [42,43]. Translation of the viral genomic RNA into the polyprotein depends upon the correct attachment between VPG and eIF4E. Mutations in eIF4E have been identified as the cause of the Potato virus Y (PVY) resistant phenotype of pepper cultivars. However, PVY overcomes the resistance by fixing amino acid changes in the central domain of VPG that reconstitutes the correct binding. These mutants pay a fitness cost in the non-resistant pepper.

Concluding remarks

Here we have shown for the first time how DMFE for an RNA virus vary across hosts. Our results suggest that the location of the DMFE moves towards smaller values as the phylogenetic distance to the natural host increases. In parallel, the distribution switches from negative to positive skewness, indicating that the probability of potential beneficial mutations increases along with host genetic distance. Similarly, we have found that the virus genotype and the host species interact in a non-linear manner to determine viral fitness. Both pleiotropic effects and reductions in genetic variance contribute to generate this genotype-by-host interaction. The implications of these observations for our understanding of emerging viral infections are multiple, but basically all hint on the unpredictability at the level of individual mutations: in the light of information collected on the primary host one can not anticipate which particular viral genotypes will be more likely to emerge. However, antagonistic pleiotropy still leaves some room for predictability at the level of classes of mutations: beneficial mutations, as a class, in the natural host may become deleterious in an alternative one, or vice versa.

Materials and Methods

Virus genotypes

For this study, a subset of 20 mutants non-lethal in N. tabaci (Table 1) was randomly chosen from a larger collection used in a previous study [20]. A plasmid containing the TEV genome, pMTEV [44], generously gifted by Dr. J.A. Darós, was used to generate both the wildtype virus and the mutant genotypes. Single-nucleotide substitution mutants were generated by site-directed mutagenesis using QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene) as described in [20] and following the manufacturer’s instructions. The kit incorporates PhUltra high fidelity DNA polymerase that minimizes the introduction of undesired mutations. The uniqueness of each mutation was confirmed by sequencing an 800 bp fragment encompassing the mutagenized nucleotide.

Infectious RNA of each genotype was obtained by in vitro transcription after BglII linearization of the corresponding plasmid as described in [45]. The infectivity of each RNA genotype was tested by inoculating five N. tabacum plants. All TEV genotypes were confirmed to be infectious on N. tabacum.

Host species

Eight host species previously described as susceptible to TEV systemic infection (VIDE database; pvo.bio-mirror.cn/refs.htm) were chosen for these experiments. Five hosts belong to the Solanaceae family: N. tabacum, N. benthamiana, D. stramonium, C. annuum, and S. lycopersicum. The first two belong to the same genus of the Nectarinoideae subfamily whereas the other three belong to the Solanoidoeae subfamily [27]. One host, H. annuus, pertains to the Asteraceae family. Both Solanaceae and Asteraceae are classified as Asterids [27]. The remaining two hosts, G. globoza and S. oleracea belong to the family Amaranthaceae. The three plant families are Eudicots [27].

Inoculation experiments

All hosts were at similar growth stages when inoculated in order to minimize infectivity error due to possible variation in defense response to infection with developmental stage. All inoculations were done in a single experimental block. Nine plants per host per TEV genotype (9 x 0 x 21 = 1512) were inoculated by rubbing the first true leaf with 5 μL containing 5 μg RNA in vitro transcript of the virus and 10% carborundum [100 mg/mL]. Solanaceae hosts show clear symptoms when infected and thus visual inspection was enough for determining infection. Nonetheless, some randomly chosen asymptomatic Solanaceae plants were subjected to RT-PCR for detection of infection as described in [46]. None was positive in this test. In the case of the non-Solanaceae hosts, symptoms were not recognizable and thus, infection was confirmed by RT-PCR.

Ten days post-inoculation (dpi), the whole infected plant, except the inoculated leaf, was collected. The whole tissue was frozen in liquid nitrogen and ground with mortar and pestle.

RNA purification and virus quantification

An aliquot of approximately 100 mg of grounded tissue was taken and mixed with 200 μL of extraction buffer (0.2 M Tris, 0.2 M NaCl, 50 mM EDTA, 2% SDS; pH 8). An equal volume of phenol:chloroform:isoamyl alcohol (25:25:1) was added, thoroughly vortexed and centrifuged at 14000 g for 5 min at 25°C. Ca. 160 μL of the upper aqueous phase were mixed with 80 μL of a solution containing 7.5 M LiCl and 50 mM EDTA and incubated overnight on ice at 4°C for RNA precipitation. The precipitated RNAs were centrifuged at 14000 g for 15 min at 4°C, washed once with 70% ice-cold ethanol, dried in a SpeedVac (Thermo) and resuspended in 30 μL of DEPC-treated ultrapure water. RNA concentration was measured spectrophotometrically and the samples were diluted to a final concentration of 50 ng/μL.

Within-plant virus accumulation was measured by absolute RT-qPCR using external standard [47]. Standard curves were constructed using five serial dilutions of TEV RNA produced by in vitro transcription and diluted in RNA obtained from the corresponding healthy host plant species. Samples were grouped by host species and quantity of viral RNA was calculated using the corresponding standard curve.

RT-qPCR reactions were performed in 20 μL volume using One Step SYBR PrimeScript RT-PCR Kit (TaKaRa) following the instructions provided by the manufacturer. The primers forward TEV-CP 5′-TTTGGTCTTGTATGGCAAGCTG and reverse TEV-CP 5′-TTTGCGCGTTCAAGTTTCTCCGTTTGTT amplify a 71 nt fragment within the TEV CP cistron. CP was chosen because it locates in the 3′ end of TEV genome and hence would only quantify complete genomes but not partial incomplete amplicons. Each RNA sample was quantified three times in independent experiments. Amplifications were done using the ABI PRISM Sequence Analyzer 7000 (Applied Biosystems). The thermal profile was as follows: RT phase consisted of 5 min. at
was used. No qualitative differences were observed between the results obtained with these alternative distributions. Results reported will be those obtained using the Gaussian model.

Acknowledgments
We thank Francisca de la Iglesia and Angels Prósper for excellent technical assistance and Stéphanie Bedhomme, José A. Darós, Guillame Martin, and Susanna K. Remold for help, comments, or discussion.

Author Contributions
Conceived and designed the experiments: SFE. Performed the experiments: JL. Analyzed the data: JL SFE. Wrote the paper: JL. SFE. Contributed in performing experiments: JMC.

References
1. Woodhouse MJ, Haydon DT, Antia R (2005) Emerging pathogens: the epidemiology and evolution of species jumps. Trends Ecol Evol 20: 238–244.
2. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, et al. (2008) Cross-species transmission and the emergence of new infectious diseases. Microbiol Mol Biol Rev 72: 457–470.
3. Holmes EC (2009) The evolutionary genetics of emerging viruses. Annu Rev Environ Resour 34: 535–572.
4. Elena SF, Frosaert R (2010) New experimental and theoretical approaches towards the understanding of the emergence of viral infections. Phil Trans R Soc B 365: 1679–1689.
5. Anderson PS, Cunningham AA, Patel NG, Morales FJ, Epstein PR, et al. (2008) Emerging infectious diseases of plants: pathogen infection, climate change and agrotechnological drivers. Trends Ecol Evol 19: 535–544.
6. Cleaveland S, Haydon DT, Taylor L (2007) Overviews of pathogen emergence: which pathogens emerge, when and why? Curr Top Microbiol Immunol 315: 85–111.
7. Jones RAC (2009) Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. Virus Res 141: 113–130.
8. Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, et al. (2011) The evolutionary genetics of emerging plant RNA viruses. Mol Plant-Microbe Interact 24: 287–293.
9. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010) Viral mutation rates. J Virol 84: 9733–9748.
10. Turner PE, Elena SF (2000) Cost of host radiation in an RNA virus. Genetics 156: 1465–1470.
11. Duffy S, Turner PE, Burch CL (2006) Pleiotropic costs of niche expansion in the RNA bacteriophage Qβ. Genetics 172: 751–757.
12. Ferris ML, Joyce P, Burch CL (2007) High frequency of mutations that expand the host range of an RNA virus. Genetics 176: 1013–1022.
13. Agudelo-Romero P, de la Iglesia F, Elena SF (2008) The pleiotropic cost of host specialization in Tobacco etch potyivirus. Infect Genet Evol 8: 806–814.
14. Gandon S (2004) Evolution of multihost parasites. Evolution 58: 455–469.
15. Erenold SK, Rambaut A, Turner PE (2003) Evolutionary genomes of host adaptation in Vescicular staminius virus. Mol Biol Evol 20: 1138–1147.
16. Domingo-Calap P, Cuevas JM, Sanjuán R (2009) The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. PLoS Genet 5: e1000742. doi:10.1371/journal.pgen.1000742.
17. Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuán R (2010) Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage φ1. Genetics 155: 603–609.
18. Elena SF, Moya A (1999) Rate of deleterious mutations and the distribution of substitutions in their effects on fitness in Vescicular staminius virus. J Evol Biol 12: 1078–1088.
19. 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: 8566–8569.
20. Carrasco P, de la Iglesia F, Elena SF (2007) Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco etch virus. J Virol 81: 12979–12984.
21. Sanjuán R (2010) Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis. Phil Trans R Soc B 365: 1975–1982.
22. Van Oprijeen T, Boerlijst MC, Berkhoft B (2006) Effects of random mutations in the Human immunodeficiency virus type 1 transcriptional promoter on viral fitness in different host cell types. J Virol 80: 6678–6685.
23. Hodgins-Davies A, Townsend JP (2010) Evolving gene expression: from G to E to GxE. Trends Ecol Evol 24: 649–658.
24. Futuyma DJ, Moreno G (1983) The evolution of ecological specialization. Annu Rev Ecol Syst 19: 207–234.
25. Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121: 129–138.
26. Remold SK, Lenski RE (2001) Contributions of individual random mutations to genotype-to-environment interactions in Escherichia coli. Proc Natl Acad Sci USA 98: 11388–11393.
27. Solis DE, Solis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol Biol 42: 45–75.
28. Novella IS, Zárate S, Metzgar D, Ebenhoch-Consuejo RE (2004) Positive selection of synonymous mutations in Vescicular staminius virus. J Mol Biol 342: 1415–1421.
29. Ohra T (1992) The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst 23: 263–286.
30. Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends Ecol Evol 19: 101–108.
31. Martin G, Lenormand T (2006) The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 60: 2413–2427.
32. Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. Nat Rev Genet 8: 610–618.
33. Wylie CS, Shakhnovich EI (2011) A biophysical protein folding model accounts for most mutational fitness effects in viruses. Proc Natl Acad Sci USA 108: 9916–9921.
34. DeFilippis VR, Villarreal LP (2000) An introduction to the evolutionary ecology of viruses. In Viral Ecology (ed. Hurst CJ). Academic Press: New York, USA. pp 126–208.
35. Holmes EC, Drummond AJ (2007) The evolutionary genetics of viral emergence. Curr Top Microbiol Immunol 315: 51–66.
36. Kondrashov AS, Houle D (1994) Genotype-environment interactions and the estimation of the genomic mutation rate in Drosophila melanogaster. Proc R Soc B 259: 221–227.
37. Via S, Gemmell K, de Jong G, Scheiner SM, Schlichting CD, et al. (1995) Adaptive phenotypic plasticity: consensus and controversy. Trends Ecol Evol 10: 212–217.
38. Korona R (1999) Genetic load of the yeast Saccharomyces cerevisiae under diverse environmental conditions. Evolution 53: 1966–1971.
39. Auhl JR, Agrawal AA, Reyba RA (2010) Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proc R Soc B 277: 503–511.
40. Fry JD, Heinsohn SL, Mackay TFC (1996) The contribution of new mutations to genotype-environment interaction for fitness in Drosophila melanogaster. Evolution 50: 2316–2327.
41. Elena SF, Agudelo-Romero P, Carrasco P, Codoto FM, Martin S, et al. (2008) Experimental evolution of plant RNA viruses. Heredity 100: 478–483.
42. Ayne V, Souché S, Caranta C, Jacquemoud M, Chabotedral J, et al. (2006) Different mutations in the genome-linked VPg of Potato virus X confer virulence on the pot23(5) resistance in pepper. Mol Plant-Microbe Interact 19: 557–563.
43. Charron C, Nicolai M, Gallow JL, Robaglia C, Mouy B, et al. (2008) Natural variation and functional analysis provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J 54: 56–68.
44. Bedoya LC, Darós JA (2010) Stability of Tobacco eth virus infectious clones in plasmid vectors. Virus Res 149: 234–240.
45. Carrasco P, Darós JA, Agudelo-Romero P, Elena SF (2007) A real-time RT-PCR assay for quantifying the fitness of Tobacco eth virus in competition experiments. J Virol Meth 139: 181–188.
46. Lalik J, Agudelo-Romero P, Carrasco P, Elena SF (2010) Adaptation of Tobacco eth potyvirus to a susceptible ecotype of Arabidopsis thaliana capacities it for systemic infection of resistant ecotypes. Phil Trans R Soc B 65: 1997–2008.
47. Pfaffl MW (2004) Quantification strategies in real-time PCR. In A-Z of Quantitative PCR (ed. SA Basman). International University Line: La Jolla, USA. pp 87–112.