Review

Plant RNA virus fitness predictability: contribution of genetic and environmental factors

S. F. Elena\textsuperscript{a,b,*} and J. Lalić\textsuperscript{a}

\textsuperscript{a} Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, Valencia, 46022, Spain;
\textsuperscript{b} The Santa Fe Institute, Santa Fe, NM 87501, USA

* E-mail: sfelena@ibmcp.upv.es

Short title: The genetic architecture of RNA virus fitness

Keywords: emerging viruses; epistasis; mutational effects; pleiotropy; reaction norms; virus evolution
Forecasting plant virus emergence depends on identifying the factors that determine the distribution of genetic variants within the primary host as well as across potential new hosts. It is crucial (i) to determine the distribution of mutational fitness effects (DMFE) on the primary host, (ii) how it changes on different hosts, (iii) the way in which multiple mutations interact in determining viral fitness in the primary host, and (iv) whether this interaction is host-dependent. To illustrate points (i) and (ii) we review recent reports showing that the DMFE for a potyvirus markedly differs between natural and non-natural hosts. Changes in genetic variance for fitness are the main cause of the observed pattern among related hosts, whereas sign pleiotropy mainly explains differences observed among unrelated hosts. To illustrate point (iii), we comment on experiments showing significant epistasis among random pairs of mutations in potyvirus genome. A large fraction of the interactions correspond to the reciprocal sign epistasis, meaning that the sign of the effects of mutations at two loci are mutually dependent. Finally, to illustrate point (iv) we present evidences that epistatic interactions for an RNA virus varied among hosts, with magnitude epistasis being stronger in the primary host but becoming weaker as host’s taxonomic relatedness decreased. The existence of all these interactions jeopardizes predicting the fitness and evolutionary fate of a given mutation, since it will depend on the genetic background but also on the host wherein the virus replicates.
Introduction

The emergence of plant viruses, understanding it as the generation of a new virus or a new viral genotype able of infecting previously non-susceptible hosts, is a complex problem that results from a combination of ecological and genetic factors (Anderson et al., 2004; Woolhouse et al., 2005; Cleveland et al., 2007; Jones, 2009; Elena et al., 2011). The increasing threats imposed by emerging and re-emerging viruses implies urgency in predicting the conditions under which plant RNA virus populations replicating in their primary hosts would 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 first need to identify the factors determining why some viruses, like Cucumber mosaic virus, Potato virus Y (PVY), Barley yellow dwarf, or Pepino mosaic virus, have caused pandemics, whereas other viruses, such as Cotton leaf curl virus, Maize rough dwarf virus or Cocoa swollen shoot disease virus produce outbreaks limited in time and space. Conducio sine qua non for viral emergence is the existence of standing genetic variation within the primary host that enables successful replication within new hosts after occasional spillovers (Holmes, 2009; Elena et al., 2011). Neglecting the effect of genetic drift, the frequency of host-range mutations within the primary host will depend on the equilibrium between the rate at which they are produced (i.e., mutation and recombination rates) and the fitness advantage (or disadvantage) they may have in the primary host. For instance, if host-range mutations are deleterious in the primary host, their frequency will be low and thus the likelihood of emergence will be low as well. By contrast, if they are neutral or beneficial, their frequency will increase, rising up the chances of emergence.

It is generally assumed that RNA viruses have high evolutionary potential as a consequence of their fast and error-prone replication (Sanjuán et al., 2010) along with
incredibly large population sizes (Holmes, 2009; Elena et al., 2011). Regarding fitness
effects, extensive data have shown that host-range mutants confer high fitness in the
new host but usually pay fitness penalties in their primary host (Jenner et al., 2002;
Agudelo-Romero et al., 2008; Bedhomme et al., 2012). Interestingly, fitness trade-offs
should preclude the evolution of generalist multi-host viruses (Gandon, 2004; Agudelo-
Romero et al., 2008; Bedhomme et al., 2012), since specialist will always outcompete
generalists in their corresponding hosts. Sign pleiotropy, i.e. when the sign of an
allele’s effect on fitness depends on the environment (Remold, 2012), has been recently
referred to explain for the existence of such fitness trade-offs (Whitlock, 1996;
Agudelo-Romero et al., 2008; Bedhomme et al., 2012), although the accumulation of
neutral mutations in genes that are not necessary in a given host but essential in
alternative ones maybe a plausible explanation for specialization (Kawecki, 1994).

Probability that a viral genotype infects new hosts depends on the change in the
distribution of mutational fitness effects (DMFE) between the primary and the new
hosts, that is, whether the fraction of lethal, deleterious, neutral, and beneficial
mutations remains constant or varies across hosts. In addition, it is also essential to
know whether the effect of a given host-range mutation depends on the genetic
background where it appears or its effect is background-independent. These questions
are particular cases of two more general biological problems: (i) the extent to which a
phenotype (here viral fitness, $W$) is determined by the interaction between different loci
in the genome, also known as epistasis, and (ii) to which extent viral fitness results from
the genotype-by-environment interaction ($G \times E$ or reaction norm), host species or
genotypes being the environment for viruses (Hodgins-Davies & Townsend, 2010).

Epistasis (the genotype-by-genotype component or $G \times G$) is particularly relevant for
understanding adaptive evolution, as it determines the ruggedness of the adaptive
landscape (Whitlock et al., 1995; Poelwijk et al., 2011) as well as the accessibility of adaptive pathways throughout the landscape (Weinreich, 2005; Welch & Waxman, 2005; Franke et al., 2011). Evolutionary trajectories may end up at suboptimal fitness peaks due to the ruggedness of the landscape; thus epistasis can therefore hamper the efficiency of natural selection and thus slow down the rate of adaptation (Whitlock et al., 1995). Moreover, epistasis can make certain evolutionary pathways selectively inaccessible because of the valleys in the fitness landscape: intermediate genotypes have reduced fitness compared with surrounding genotypes.

The extent, origin and consequences of $G \times E$ interactions in determining phenotypes 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. The fate of genetic variation in viral population depends on the form of the $G \times E$ interactions (Futuyma & Moreno, 1988) and, for instance, a change in the rank order of fitness of virus genotypes in different hosts may support a balanced polymorphism in the viral population (Gillespie & Turelli, 1989).

In more quantitative terms, the fitness $W$ of a viral genotype $G$ infecting a host $E$ would be given by the relationship

\[ W \sim G + E. \]  

Does Eq. 1 provide a good approximation to viral fitness? How many additional terms need to be added to achieve a good prediction of viral fitness? In an effort to tackle these issues for a plant RNA virus, we have been conducting a series of experiments with Tobacco etch virus (TEV; genus Potyvirus, family Potiviridae). In a first stage, we created a collection of single-nucleotide substitution mutants and evaluated the DMFE on the primary host Nicotiana tabacum (Carrasco et al., 2007) and in a set of new hosts that differed in degree of taxonomic relatedness to tobacco (Lalić et al., 2011). These
experiments allowed us to demonstrate the existence and the causes of $G \times E$. In a second set of experiments, we characterized the amount and type of epistasis among random pairs of point mutations in the primary host (Lalić & Elena, 2012a). Finally, in a third set of experiments we tested whether epistasis itself varied across hosts (Lalić & Elena, 2012b). Here, we provide an overview of these experiments and provide an integration of the different results into a unified conceptual framework that tries to shed light onto the problem of emerging viruses. Those readers interested in methodological details are kindly directed to the original articles.

Definition of viral fitness and properties of the DMFE in the primary host

Fitness is a macroscopic property that measures the reproductive success of a viral genotype on a given host. As such, it includes many different components, for instance, genome unpacking, translation, replication, coating into new particles, and cell-to-cell and systemic movement. In all these steps, fitness depends on the quality of the interactions with many different cellular components that the virus uses on its own benefit. Furthermore, viral fitness would also depend on the successful interaction between the virus and the defense mechanisms of the plant, by dismounting or evading them. Finally, viral fitness also depends on the stability of virion particles and, obviously, on the efficiency of the processes within the vector that would ensure a successful transmission to the next host. In most plant virus evolution experiments, and in those regarding this review, vectors do not play any role, since transmission is always mechanical. In our studies we have used real-time quantitative PCR to determine virus concentration systemically infected leaves. From these determinations, we estimated a Malthusian growth rate per day, $m$, for each TEV genotype on each particular host.
Absolute fitness was then defined as $W = e^m$ (Crow & Kimura, 1970). In Lalić et al. (2011) we directly reported $m$ as a measure of fitness, whereas in all other studies we reported $W$. Here, we homogenize fitness definitions and use $W$ in all cases.

DMFE have been characterized in recent years for a handful of single-stranded DNA and RNA viruses in their primary hosts (reviewed by Sanjuán, 2010). In all cases, site-directed mutagenesis was performed on infectious clones, generating collections of random single-nucleotide substitution mutants. The fitness of each mutant was then determined. Carrasco et al. (2007) characterized the DMFE for the first plant virus, TEV on its primary host *N. tabacum*. Notice that this study reported relative fitness, rather than absolute fitness, evaluated by means of competition experiments between the mutant genotypes and an engineered surrogated wild-type. Three major conclusions could be drawn from this study. First, TEV shows very little tolerance to mutations, with a large fraction (ca. 41%) being lethal. Second, for non-lethal mutations, the mean fitness loss associated to a single nucleotide substitution is about 50%. Third, the DMFE is left-skewed (i.e., containing more negative values than the Gaussian) and leptokurtic (i.e., comprising less central values than the Gaussian and having heavier tails). Accordingly, the probability density function (PDF) that better fits the data belongs from the heavy-tailed family (e.g., Weibull) or a highly skewed one (Beta).

**Epistasis: mutational fitness effects depend on the genetic background**

Multi-dimensional epistasis refers to all possible individual interactions among a set of mutations, providing a precise description of the fitness landscape (Kondrashov & Kondrashov, 2001) (Fig. 1). Magnitude epistasis occurs when the fitness value of a mutation depends on the genetic background, while its sign remains constant (Weinreich, 2005; Poelwijk et al., 2011). Magnitude epistasis can be either
positive/negative depending on whether the double mutant is more/less fit than expected under the multiplicative null model (Fig. 1). Sign epistasis refers to cases where the sign of the mutational effect changes depending on the genetic background (i.e., a mutation may be beneficial in one background but deleterious in another; Fig. 1) (Weinreich, 2005; Poelwijk et al., 2011). A particular case of sign epistasis is reciprocal sign epistasis, when the sign of the fitness effect of a mutation is conditional upon the state of another locus and vice versa (Fig. 1). Reciprocal sign epistasis is a necessary condition for an adaptive landscape to be rugged (Poelwijk et al., 2011).

Positive magnitude epistasis has been shown to be the norm in animal and bacteriophage RNA viruses (reviewed in Elena et al., 2010). Would this be the case for a plant RNA virus? To answer this question Lalić & Elena (2012a) sought to characterize the patterns of multidimensional epistasis in TEV. To do so, pairs of mutations from the Carrasco et al. (2007) collection were drawn at random and the corresponding double mutants were generated by site-directed mutagenesis. The absolute fitness of the wild-type \( W_{00} \), the corresponding single \( W_{x0} \) and \( W_{0y} \) and the double mutants \( W_{xy} \) were evaluated as described above. Magnitude epistasis among mutations \( x \) and \( y \), \( \varepsilon_{xy} \), was calculated as \( \varepsilon_{xy} = W_{00}W_{xy} - W_{x0}W_{0y} \) (Kouyos et al., 2007).

Several interesting results were found by Lalić & Elena (2012a). First, magnitude epistasis was widespread, with some pairs showing negative epistasis and others positive epistasis. Cases of negative epistasis were associated to the generation of synthetic lethals, i.e., two mutations that were independently viable resulted in lethality when combined. Otherwise, the average epistasis was positive, in agreement with former observations for other RNA viruses. Fig. 1 shows the number of cases of magnitude, sign and reciprocal sign epistasis within our dataset of 53 TEV double mutants. Another very interesting observation is the pervasiveness of reciprocal sign
epistasis; 12 out of the 20 TEV double-mutant genotypes for which significant epistasis were detected fulfilled the mathematical condition of sign epistasis (Poelwijk et al., 2011), and among these, 11 further met the condition for reciprocal sign epistasis.

The dominance of positive epistasis among deleterious mutations and the high frequency of synthetic lethality in TEV genome are side-effects of the low genetic robustness of RNA genomes that lack of redundancy and, by contrast, often code for overlapping reading frames, contain functional RNA secondary structures and encode multi-functional proteins. The abundance of reciprocal sign epistasis suggests that TEV fitness landscape must be highly rugged. This high ruggedness has implications for the evolutionary dynamics of TEV, since it imposes harsh constraints to the evolution. Ruggedness also means that historical contingency should be important: the first mutation to appear in a genome conditions what evolutionary mutational pathways maybe reachable. In other words, the result of evolutionary optimization may not necessarily be the global optima but TEV populations may be trapped into suboptimal fitness peaks.

A particularly illustrative study of the effect of epistasis among viral loci on the emergence of resistance-breaking viruses was recently provided by Monterry et al. (2011). These authors found that certain alleles of the VPg protein conferred PVY the ability to infect and accumulate in Capsicum annuum plants that carried the pvr2 resistance allele (a particular genetic variant of the eukaryotic translation initiation factor 4E, eIF4E). However, the beneficial effect of the escape mutations at VPg was conditional upon the alleles present at the CI viral protein.

Therefore, Eq. 1 has to be modified by decomposing the $G$ term into two factors, one accounting for the net fitness effect of point mutations and an additional one that
accounts for the epistatic interactions between mutations at different loci in TEV genome:

\[ W \sim G + G \times G + E \]  

(2)

**G×E: mutational fitness effects are dependent on the host species**

Lalić et al. (2011) undertook the task of exploring how different host species would affect the parameters describing the DMFE for TEV, as well as specifically testing whether point mutations would be sufficient to give rise to a significant \( G \times E \) in a viral genome. To do so, they randomly selected 20 single mutants from Carrasco et al. (2007) collection and quantified their fitness across a panel of eight host species. Five hosts belonged to the natural host range of TEV (the Solanaceae species \( N. \) tabacum, *Nicotiana benthamiana*, *Solanum lycopersicum*, *C. annuum*, and *Datura stramonium*). The other three species were not TEV natural hosts, although they were experimentally susceptible to systemic infection (the Asteraceae *Helianthus annuus*, and the Amaranthaceae *Gomphrena globosa* and *Spinacea oleracea*). Table 1 shows the parameters describing the DMFE and the classification of mutations on each host. Overall, mutations are either neutral or deleterious in hosts that are close relatives to the primary one (*N. tabacum*), with the expected value of the DMFEs being close to the one estimated for the primary host and the distributions being left-skewed (i.e., most mutations being deleterious or even lethal; Table 1). As hosts taxonomic relatedness to the primary one decreases, the DMFEs suffer a change in their location and shape: the expected deleterious fitness effect became larger but the distributions also become right-skewed (i.e., a certain fraction of mutations become beneficial; Table 1). This suggests that the number of mutations that may potentially expand TEV host range is large and increasing as the taxonomic relatedness to the primary host decreases. In all cases,
regardless the host, the PDF that better fits the data belong to the heavy-tailed family (e.g., Weibull).

The analyses of the DMFE already suggest the existence of a significant $G \times E$ component. Proper analysis of the fitness data (GLM using host species and TEV genotypes as random factors) confirms that most of the observed variation (66.82%) was attributable to the $G \times E$ interaction, whereas 26.13% was due to pure differences among host species and 4.29% to pure genetic differences among TEV mutants. This large significant interaction means that we cannot accurately predict a particular genotype’s absolute fitness in a given host from the main effects. Henceforth, this result confirms that Eq. 1 needs to be modified to account for the dependence of mutational fitness effects on the host wherein effects are being evaluated:

$$W \sim G + E + G \times E$$  \hspace{1cm} (3)

Lalić et al. (2011) data demonstrate that single random nucleotide substitutions are sufficient to produce a significant $G \times E$. Mutations involved in significant $G \times E$ were scattered along the genome and they were randomly chosen irrespective of their fitness effects. Thus, it is possible to conclude that phenotypic plasticity in TEV was not associated to the expression of any particular cistron but results form the contribution of different ones. In the context of emerging plant virus infections, the existence of a significant $G \times E$ means that knowing the absolute fitness of a viral genotype in the primary host informs us little about what it may be in alternative ones, thus minimizing our ability to predict which genetic variants may be relevant for expanding TEV host-range.

There is a compelling idea that taxonomic relatedness among primary and novel hosts may constrain the chances for a virus to jump the host species barrier, and that the more closely related the primary and the new host are, the greater are the chances for a
successful spillover (DeFilippis & Villareal, 2000). 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 genetically 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 the new host (Holmes & Drummond, 2007). Viral host switches between closely related species (e.g., species within the same genera) may also be limited by cross-immunity to related pathogens.

The causes of $G \times E$: differences in genetic variance for fitness and antagonistic pleiotropy

A significant $G \times E$ can be produced by two non-mutually exclusive mechanisms (Remold & Lenski, 2001). First, pleiotropic effects may change the rank order of mutations from the primary to alternative hosts (e.g., a mutation beneficial in the new host may not be so in the primary one). Second, whilst retaining the rank order of fitness effects, $G \times E$ can also be generated by altering the genetic component of phenotypic variance across hosts ($\sigma^2_{G \times E}$). The relative contribution of these two mechanisms to the observed $G \times E$ can be evaluated using Robertson (1959) decomposition of $\sigma^2_{G \times E}$. The amount of $G \times E$ expressed by a collection of viral genotypes across two heterogeneous hosts could be written as:

$$\sigma^2_{G \times E} = \frac{1}{2} (\sigma_{G_H}^2 - \sigma_{G_{N.tobacum}}^2)^2 + \sigma_{G_H}^2 \sigma_{G_{N.tobacum}}^2 (1 - \rho_{G_HG_{N.tobacum}}),$$  \hspace{1cm} (4)

where $\sigma_{G_H}$ and $\sigma_{G_{N.tobacum}}$ are the genetic standard deviations for fitness in novel host $H$ and the primary host $N. tabacum$, respectively, and $\rho_{G_HG_{N.tobacum}}$ is the genetic correlation for fitness across both hosts. The first right-hand term in Eq. 4 corresponds
to the variance resulting from the differences between genetic variation expressed in the
two hosts. $G \times E$ will be generated if there is more genetic variance in one host than in
the other because the differences between viral genotypes will depend on the host that
they are infecting. The second right-hand term in Eq. 4 involves the genetic correlation
between hosts. In this case $G \times E$ will be generated if the collection of genotypes
responds inconsistently to different hosts, that is, if the rank order of fitness effects is
altered from the primary host to each alternative one. If $\rho_{G_{H}G_{N\text{.tabacum}}} < 0$, then
selection would generate sign pleiotropy (sensu Remold, 2012) thus favoring different
viral genotypes in different hosts.

Table 2 shows the estimated components of genetic variance ($\sigma^2_{G}$ and $\sigma^2_{G \times E}$) and the
genetic correlation, $\rho_{G_{H}G_{N\text{.tabacum}}}$, that are necessary to evaluate the relative
contribution of pleiotropy and change in genetic variances. Two interesting
observations can be drawn from Table 2. First, on average, the genetic variances were
larger for the Solanaceae than for the non-Solanaceae. Second, genetic correlations
were positive for all the Solanaceae, suggesting weak magnitude pleiotropy (sensu
Remold, 2012): on average, mutations beneficial in N. tabacum tend to remain
beneficial, although to a different extent, in phylogenetically related hosts. However,
correlations become negative for the non-Solanaceae, indicating sign pleiotropy: on
average, mutations being beneficial in the new hosts tend to be deleterious in the
primary one. Fig. 2 shows the fraction of $\sigma^2_{G \times E}$ attributable to each mechanism.
Whereas changes in genetic variances between primary and alternative hosts explain
most of the observed differences in $G \times E$ for alternative hosts that are phylogenetically
related to the primary one, sign pleiotropy largely explains the observed differences in
$G \times E$ for hosts that are unrelated to the primary one. This has profound evolutionary
implications. Changes in genetic variance imply that the relative influence of selection
and drift on the fate of mutations depends on the host. Exposure to the hosts within which the genetic variance for fitness is low minimizes the efficiency by which natural selection operates either removing deleterious alleles or fixing beneficial ones and thus enhances the role of drift. This seems to be the situation for the Solanaceae hosts, suggesting that different TEV alleles may dominate in one host or another as a consequence of a balance between drift and selection. By contrast, sign pleiotropy implies that selection favors different mutations in different hosts thus driving to a balanced polymorphism across hosts and leads to specialization. The sign pleiotropy observed between N. tabacum and 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 interactions 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 primary host (reviewed in Elena et al., 2011). A particularly illustrating example is the interaction between the VPg of potyviruses and the host’s eIF4E (Robaglia & Caranta, 2006).

Translation of the viral genomic RNA into the polyprotein depends upon the correct interaction between VPg and eIF4E. Mutations in eIF4E have been identified as the cause of PVY resistant phenotype of pepper cultivars. Not surprisingly, PVY overcomes the resistance by fixin amino acid changes in the central domain of VPg that reconstitute the correct binding. These mutations pay a fitness cost in the non-resistant pepper cultivars (Ayme et al., 2007; Montarry et al., 2011).

Epistasis among mutations also depends on host: $G \times G \times E$

The results presented so far suggest that (i) epistasis is common in TEV genome and that (ii) mutational effects depend on the host. Therefore, it is logical to expect that...
Epistasis may also vary depending on the host, that is, a significant $G \times G \times E$ component may exist to determine TEV absolute fitness. To test this prediction, Lalić & Elena (2012b) evaluated the strength and type of epistasis for a set of TEV double mutants on four experimental hosts (*N. tabacum*, *D. stramonium*, *H. annuus*, and *S. oleracea*). The 10 double mutants used were randomly chosen among the larger collection described in Lalić & Elena (2012a). Fig. 3 shows the distribution of epistasis across the four hosts, after removing synthetic lethals from the dataset (which is justified since they are irrelevant in terms of evolutionary dynamics). In short, average epistasis was positive in the primary host, as already shown above, but became negative, although not significant, on all alternative hosts, with a tendency to reduce in magnitude as the taxonomic relatedness to the primary host decreased (Fig. 3). Furthermore, the number of non-epistatic interactions was significantly larger in non-*Solanaceae* hosts.

These results indicate that host effects on epistasis, similarly to what happened with the effect of point mutations, are modulated by the degree of genetic divergence between the primary and alternative hosts. This result is in good agreement with the prediction that mutations shall be more severe in poor environments and milder in rich ones (You & Yin, 2002). Furthermore, mild mutations are expected to be involved in negative epistatic interactions in poor environments but in positive interactions in rich ones (You & Yin, 2002). Our results are in good agreement with these predictions: average mutational effects are milder and mutations show positive epistasis in the primary host but switch to larger effects and negative or no epistasis in alternative hosts. Together, these observations suggest that the primary host, and those that are closely related to it, represent rich environments for TEV while the alternative and unrelated hosts represent more stressful environments. This makes sense, considering that TEV has a coevolutionary history with *Solanaceae* hosts and thus its interaction with cellular
resources and defenses is optimal. By contrast, alternative hosts may not provide the
necessary resources at the right time, amount or location.

\[ G \times G \times E \] is equivalent to the concept of epistatic pleiotropy (Remold, 2012). Under
epistatic pleiotropy, virus populations may achieve either specialization for a single host
or, alternatively, become generalist with no cost, depending on the host in which they
evolve. More importantly, no-cost generalists can evolve despite the existence of true
genetic trade-offs. We will discuss this possibility in large in the next section.

Finally, here we have provided evidences that Eqs. 2 and 3 are still insufficient to
describe the variability in TEV fitness and that a more realistic description would be
provided by the following equation, which incorporates all levels of genetic and
genetic-by-environmental interactions:

\[ W \sim G + G \times G + E + G \times E + G \times G \times E \] (5)

**Pleiotropy and epistasis**

Pleiotropy and epistasis have strong parallelism because for both interactions, the effect
of an allele depends on its context: the host species for pleiotropy and the virus’ genetic
background for epistasis. Indeed, it has been postulated that pleiotropy is a prerequisite
for epistasis (Martin et al., 2007; De Visser et al., 2011). This dependence is easy to
understand for the case of sign pleiotropy, where mutations with a positive effect in the
new host have a negative effect in the primary one (Remold, 2012). In the context of
compensatory evolution, sign pleiotropy is a precondition for sign epistasis (Fig. 1),
because it allows for the negative pleiotropic effects of previously selected mutations to
be compensated by additional ones (De Visser et al., 2011). Therefore, the question to
be answered is whether a positive association exists between the tendency of mutations
to be involved in significant epistasis and how often they are pleiotropic. To evaluate
whether such positive association holds for TEV, we have proceeded as follows. First, 
the tendency of a given mutation $x$ to be involved into epistatic interactions, namely 
epistasis or $E_x$, was evaluated as the average of the squared epistasis coefficients $\varepsilon_{xy}$ 
for all pairs in which mutation $x$ has been tested: $E_x = \langle \varepsilon_{xy}^2 \rangle$, where $\langle \cdot \rangle$ represents the 
average value. The square was taken to remove signs as we are interested in whether a 
mutation is involved in epistasis, regardless its sign. Second, the average pleiotropic 
effect of a mutation $x$ across the seven alternative hosts was calculated as $P_x = 
\left( \left( \frac{W_{x,H}}{W_{x,N.tabacum}} - 1 \right)^2 \right)$, which measures the average quadratic difference in fitness 
between host $H$ and the primary host $N. tabacum$. For a mutation with no pleiotropic 
effect $W_{x,H} = W_{x,N.tabacum}$ and thus $P_x = 0$; for a pleiotropic mutation $W_{x,H} < W_{x,N.tabacum}$ 
and $P_x > 0$. In the extreme case of sign pleiotropy, i.e. mutation $x$ being lethal in all 
alternative hosts ($W_{x,H} = 0$), then $P_x = 1$.

Fig. 4 shows the relationship between $E_x$ and $P_x$ obtained for the TEV data described 
in the previous sections. A weak, yet significant, positive correlation exits between 
both traits ($\rho_S = 0.400$, 18 df, 1-tailed $P = 0.040$), thus supporting the positive 
association between epistasis and pleiotropy.

So far, the TEV data reviewed here picture that sign pleiotropy in host usage and 
epistasis at genomic level go hand in hand, thus corresponding to a situation that 
Remold (2012) defined as epistatic pleiotropy. Epistatic pleiotropy has two important 
implications. First, unlike either sign or magnitude pleiotropy in the absence of 
epistasis, epistatic pleiotropy allows for the evolution of either specialist or no-cost 
generalist viruses, depending on the virus population’s host. Second, and very 
important to limit the emergence of new viruses, when epistasis is in the form of 
reciprocal sign epistasis, as it is the norm in TEV genome, the ruggedness of the 
adaptive landscape diminishes the ability of viral populations to escape from specialism
to a situation of no-cost generalism. A long history of evolution in the primary host could have resulted in an adaptive walk towards a host-specific fitness peak involving most, if not all, viral loci. Such population could find itself many mutational steps, through an adaptive valley, away from reaching a generalist peak.

**Conclusions**

Here we have reviewed recent data showing that the expected effect on viral fitness of point mutations depends on the genetic background where they appear in as much as on the host species being infected by the virus. In other words, the reviewed data show that the virus genotype and the host species interact in a non-linear manner to determine the fitness of a potyvirus. The implications of these observations for our understanding of emerging plant viral infections are multiple, but basically all hint on the unpredictability at the level of effect 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 in related hosts. However, the observation of sign pleiotropy in unrelated hosts leaves some room for predictability at least at the level of classes of mutations: beneficial mutations, as a class, in the primary host may become deleterious in new ones.

Finally, the existence of epistatic pleiotropy on host usage together with the dominance of reciprocal sign epistasis in the viral genome create rugged adaptive landscapes that may trap viral populations in local peaks and impede their escape towards no-cost generalists. In some sense, these are good news, since the difficulty to generate no-cost generalists reduces the likelihood of successful spillovers, as most genotypes will necessarily pay a large fitness cost after infecting a new host.
Acknowledgements

We thank Stéphanie Bedhomme, José M. Cuevas and Susanna K. Remold for insightful discussions and suggestions. This work was supported by grants BFU2009-06993 and BFU2012-30805 from Spanish Dirección General de Investigación Científica y Técnica to S.F.E. J.L. was supported by a JAE-pre contract from CSIC.

References


Current Topics in Microbiology and Immunology 315, 51-66.

overcome host resistance. Virus Research 86, 1-6.

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 Research 141, 113-130.

Kawecki TJ, 1994. Accumulation of deleterious mutations and the evolutionary cost of 

Kondrashov FA, Kondrashov AS, 2001. Multidimensional epistasis and the 
disadvantage of sex. Proceedings of the National Academy of the USA 98, 12089- 
12092.


Lalić J, Cuevas JM, Elena SF, 2011. Effect of host species on the distribution of 

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

Lalić J, Elena SF, 2012b. Epistasis between mutations is host-dependent for an RNA 
virus. Biology Letters 9, 20120396.

Martin G, Elena SF, Lenormand T. 2007. Distribution of epistasis in microbes fit 
predictions from a fitness landscape model. Nature Genetics 39, 555-560.

plant-virus gene-for-gene interaction is strongly influenced by genetic contexts. 
Molecular Plant Pathology 12, 911-920.


**Figure 1** Abundance of the different types of genetic interactions between 53 pairs of mutant alleles observed for TEV. The effect of each type of epistasis (or lack of it) on the landscape ruggedness is illustrated (modified from Dawid et al., 2010). The red numbers next to each panel correspond to the abundance of each type of epistasis within TEV dataset. Data taken from Lalić & Elena (2012a).
**Figure 2** Contribution of sign pleiotropy and changes in variance for fitness to the observed variance in $G \times E$ when comparing the primary host (*N. tabacum*) and the alternative ones. For *Solanaceae* hosts, $G \times E$ is mostly generated by changes in genetic variance for fitness across hosts; by contrast, sign pleiotropy is the main cause of $G \times E$ for non-*Solanaceae* hosts. Data taken from Lalić *et al.* (2011).
Figure 3  Distribution of epistasis among pairs of non-lethal mutations in TEV genome evaluated on four different host species. Error bars represent ±1 SEM. Data taken from Lalić & Elena (2012b).
Figure 4  Relationship between the tendency of a mutation to be involved in epistatic interactions (epistasis) and its pleiotropic effect across hosts. Data taken from Lalić et al. (2011) and Lalić & Elena (2012a, 2012b).
Table 1 Parameters describing the DMFE and the number of mutations classified as lethal, deleterious, neutral and beneficial on each host. Data taken from Lalić et al. (2011).

<table>
<thead>
<tr>
<th>Host species</th>
<th>Expected $W$</th>
<th>Median</th>
<th>Standard deviation</th>
<th>Skweness</th>
<th>Kurtosis</th>
<th>Lethal</th>
<th>Deleterious</th>
<th>Neutral</th>
<th>Beneficial</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. tabacum</td>
<td>1.331</td>
<td>1.327</td>
<td>0.021</td>
<td>-1.974</td>
<td>4.608</td>
<td>0</td>
<td>6</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>N. benthamiana</td>
<td>1.315</td>
<td>1.319</td>
<td>0.065</td>
<td>-3.949</td>
<td>16.879</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>D. stramonium</td>
<td>1.365</td>
<td>1.380</td>
<td>0.054</td>
<td>-1.566</td>
<td>1.364</td>
<td>2</td>
<td>15</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>C. annuum</td>
<td>1.246</td>
<td>1.297</td>
<td>0.142</td>
<td>-1.037</td>
<td>-0.389</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>S. lycopersicum</td>
<td>1.350</td>
<td>1.418</td>
<td>0.041</td>
<td>-0.768</td>
<td>0.062</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>H. annuus</td>
<td>1.020</td>
<td>1.020</td>
<td>0.044</td>
<td>0.527</td>
<td>0.579</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>G. globosa</td>
<td>0.725</td>
<td>1.010</td>
<td>0.042</td>
<td>0.997</td>
<td>0.561</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>S. oleracea</td>
<td>0.976</td>
<td>0.962</td>
<td>0.052</td>
<td>1.479</td>
<td>1.915</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 2  Different components of the variance for fitness evaluated in eight susceptible hosts. Data taken from Lalić et al. (2011).

<table>
<thead>
<tr>
<th>Host species</th>
<th>$\sigma^2_G$</th>
<th>$\rho_{GH_{N.tabacum}}$</th>
<th>$\sigma^2_{G\times E}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. tabacum</td>
<td>3.210±1.245 × 10^{-4}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. benthamiana</td>
<td>2.683±0.922 × 10^{-3}</td>
<td>0.244±0.222</td>
<td>1.275±1.311 × 10^{-3}</td>
</tr>
<tr>
<td>D. stramonium</td>
<td>7.916±2.510 × 10^{-2}</td>
<td>0.220±0.224</td>
<td>3.864±4.528 × 10^{-2}</td>
</tr>
<tr>
<td>C. annuum</td>
<td>1.202±0.466 × 10^{-2}</td>
<td>0.010±0.229</td>
<td>6.148±6.587 × 10^{-3}</td>
</tr>
<tr>
<td>S. lycopersicum</td>
<td>4.639±0.143 × 10^{-1}</td>
<td>0.468±0.203</td>
<td>2.264±3.216 × 10^{-1}</td>
</tr>
<tr>
<td>H. annuus</td>
<td>9.250±6.548 × 10^{-4}</td>
<td>-0.592±0.185</td>
<td>9.458±9.555 × 10^{-4}</td>
</tr>
<tr>
<td>G. globosa</td>
<td>7.360±6.135 × 10^{-4}</td>
<td>-0.336±0.216</td>
<td>6.920±6.988 × 10^{-4}</td>
</tr>
<tr>
<td>S. oleracea</td>
<td>1.788±0.902 × 10^{-3}</td>
<td>-0.619±0.180</td>
<td>1.523±1.548 × 10^{-3}</td>
</tr>
</tbody>
</table>

$\sigma^2_G = $ genetic variance for fitness on each host.

$\rho_{GH_{N.tabacum}} = $ genetic correlation for fitness across host $H$ and the primary host $N. tabacum$.

$\sigma^2_{G\times E} = $ variance for the interaction between viral genotype and host ($G\times E$) computed using Eq. 4.