Review

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

S. F. Elena[a,b,*] and J. Lalić[a]

[a] *Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, Valencia, 46022, Spain;*

[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. *Condicio 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

63 incredibly large population sizes (Holmes, 2009; Elena *et al*., 2011). Regarding fitness
64 effects, extensive data have shown that host-range mutants confer high fitness in the
65 new host but usually pay fitness penalties in their primary host (Jenner *et al*., 2002;
66 Agudelo-Romero *et al*., 2008; Bedhomme *et al*., 2012). Interestingly, fitness trade-offs
67 should preclude the evolution of generalist multi-host viruses (Gandon, 2004; Agudelo-
68 Romero *et al*., 2008; Bedhomme *et al*., 2012), since specialist will always outcompete
69 generalists in their corresponding hosts. Sign pleiotropy, i.e. when the sign of an
70 allele's effect on fitness depends on the environment (Remold, 2012), has been recently
71 referred to explain for the existence of such fitness trade-offs (Whitlock, 1996;
72 Agudelo-Romero *et al*., 2008; Bedhomme *et al*., 2012), although the accumulation of
73 neutral mutations in genes that are not necessary in a given host but essential in
74 alternative ones maybe a plausible explanation for specialization (Kawecki, 1994).

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

86 Epistasis (the genotype-by-genotype component or $G \times G$) is particularly relevant for
87 understanding adaptive evolution, as it determines the ruggedness of the adaptive

88 landscape (Whitlock *et al*., 1995; Poelwijk *et al*., 2011) as well as the accessibility of
89 adaptive pathways throughout the landscape (Weinreich, 2005; Welch & Waxman,
90 2005; Franke *et al*., 2011). Evolutionary trajectories may end up at suboptimal fitness
91 peaks due to the ruggedness of the landscape; thus epistasis can therefore hamper the
92 efficiency of natural selection and thus slow down the rate of adaptation (Whitlock *et al*.,
93 1995). Moreover, epistasis can make certain evolutionary pathways selectively
94 inaccessible because of the valleys in the fitness landscape: intermediate genotypes have
95 reduced fitness compared with surrounding genotypes.

96 The extent, origin and consequences of $G \times E$ interactions in determining phenotypes
97 and fitness has been a central aim of ecology, genetics and evolution. Therefore, it
98 should also be central for the epidemiology and evolution of infectious diseases. The
99 fate of genetic variation in viral population depends on the form of the $G \times E$ interactions
100 (Futuyma & Moreno, 1988) and, for instance, a change in the rank order of fitness of
101 virus genotypes in different hosts may support a balanced polymorphism in the viral
102 population (Gillespie & Turelli, 1989).

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

105 $$W \sim G + E. \qquad (1)$$

106 Does Eq. 1 provide a good approximation to viral fitness? How many additional terms
107 need to be added to achieve a good prediction of viral fitness? In an effort to tackle
108 these issues for a plant RNA virus, we have been conducting a series of experiments
109 with *Tobacco etch virus* (TEV; genus *Potyvirus*, family Potiviridae). In a first stage, we
110 created a collection of single-nucleotide substitution mutants and evaluated the DMFE
111 on the primary host *Nicotiana tabacum* (Carrasco *et al*., 2007) and in a set of new hosts
112 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 leafs. 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

163 positive/negative depending on whether the double mutant is more/less fit than expected
164 under the multiplicative null model (Fig. 1). Sign epistasis refers to cases where the
165 sign of the mutational effect changes depending on the genetic background (i.e., a
166 mutation may be beneficial in one background but deleterious in another; Fig. 1)
167 (Weinreich, 2005; Poelwijk *et al*., 2011). A particular case of sign epistasis is
168 reciprocal sign epistasis, when the sign of the fitness effect of a mutation is conditional
169 upon the state of another locus and *vice versa* (Fig. 1). Reciprocal sign epistasis is a
170 necessary condition for an adaptive landscape to be rugged (Poelwijk *et al*., 2011).
171 Positive magnitude epistasis has been shown to be the norm in animal and
172 bacteriophage RNA viruses (reviewed in Elena *et al*., 2010). Would this be the case for
173 a plant RNA virus? To answer this question Lalić & Elena (2012a) sought to
174 characterize the patterns of multidimensional epistasis in TEV. To do so, pairs of
175 mutations from the Carrasco *et al.* (2007) collection were drawn at random and the
176 corresponding double mutants were generated by site-directed mutagenesis. The
177 absolute fitness of the wild-type (W_{00}), the corresponding single (W_{x0} and W_{0y}) and the
178 double mutants (W_{xy}) were evaluated as described above. Magnitude epistasis among
179 mutations x and y, ε_{xy}, was calculated as $\varepsilon_{xy} = W_{00}W_{xy} - W_{x0}W_{0y}$ (Kouyos *et al*., 2007).
180 Several interesting results were found by Lalić & Elena (2012a). First, magnitude
181 epistasis was widespread, with some pairs showing negative epistasis and others
182 positive epistasis. Cases of negative epistasis were associated to the generation of
183 synthetic lethals, i.e., two mutations that were independently viable resulted in lethality
184 when combined. Otherwise, the average epistasis was positive, in agreement with
185 former observations for other RNA viruses. Fig. 1 shows the number of cases of
186 magnitude, sign and reciprocal sign epistasis within our dataset of 53 TEV double
187 mutants. Another very interesting observation is the pervasiveness of reciprocal sign

188 epistasis; 12 out of the 20 TEV double-mutant genotypes for which significant epistasis
189 were detected fulfilled the mathematical condition of sign epistasis (Poelwijk *et al*.,
190 2011), and among these, 11 further met the condition for reciprocal sign epistasis.
191 The dominance of positive epistasis among deleterious mutations and the high
192 frequency of synthetic lethality in TEV genome are side-effects of the low genetic
193 robustness of RNA genomes that lack of redundancy and, by contrast, often code for
194 overlapping reading frames, contain functional RNA secondary structures and encode
195 multi-functional proteins. The abundance of reciprocal sign epistasis suggests that TEV
196 fitness landscape must be highly rugged. This high ruggedness has implications for the
197 evolutionary dynamics of TEV, since it imposes harsh constraints to the evolution.
198 Ruggedness also means that historical contingency should be important: the first
199 mutation to appear in a genome conditions what evolutionary mutational pathways
200 maybe reachable. In other words, the result of evolutionary optimization may not
201 necessarily be the global optima but TEV populations may be trapped into suboptimal
202 fitness peaks.
203 A particularly illustrative study of the effect of epistasis among viral loci on the
204 emergence of resistance-breaking viruses was recently provided by Monterry *et al*.
205 (2011). These authors found that certain alleles of the VPg protein conferred PVY the
206 ability to infect and accumulate in *Capsicum annuum* plants that carried the *pvr2*
207 resistance allele (a particular genetic variant of the eukaryotic translation initiation
208 factor 4E, eIF4E). However, the beneficial effect of the escape mutations at VPg was
209 conditional upon the alleles present at the CI viral protein.
210 Therefore, Eq. 1 has to be modified by decomposing the *G* term into two factors, one
211 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:

$$W \sim G + G \times G + E \quad (2)$$

$G \times E$: 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:

$$W \sim G + E + G\times E \qquad (3)$$

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*×*E*: differences in genetic variance for fitness and antagonistic pleiotropy

A significant *G*×*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*×*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*×*E* can be evaluated using Robertson (1959) decomposition of $\sigma^2_{G\times E}$. The amount of *G*×*E* expressed by a collection of viral genotypes across two heterogeneous hosts could be written as:

$$\sigma^2_{G\times E} = \tfrac{1}{2}\left(\sigma_{G_H} - \sigma_{G_{N.tabacum}}\right)^2 + \sigma_{G_H}\sigma_{G_{N.tabacum}}\left(1 - \rho_{G_H G_{N.tabacum}}\right), \quad (4)$$

where σ_{G_H} and $\sigma_{G_{N.tabacum}}$ are the genetic standard deviations for fitness in novel host *H* and the primary host *N. tabacum*, respectively, and $\rho_{G_H G_{N.tabacum}}$ 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 $\rho_{G_H G_{N.tabacum}} < 0$, 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 (σ_G^2 and $\sigma_{G\times E}^2$) and the genetic correlation, $\rho_{G_H G_{N.tabacum}}$, 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 $\sigma_{G\times E}^2$ 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

312 and drift on the fate of mutations depends on the host. Exposure to the hosts within
313 which the genetic variance for fitness is low minimizes the efficiency by which natural
314 selection operates either removing deleterious alleles or fixing beneficial ones and thus
315 enhances the role of drift. This seems to be the situation for the *Solanaceae* hosts,
316 suggesting that different TEV alleles may dominate in one host or another as a
317 consequence of a balance between drift and selection. By contrast, sign pleiotropy
318 implies that selection favors different mutations in different hosts thus driving to a
319 balanced polymorphism across hosts and leads to specialization. The sign pleiotropy
320 observed between *N. tabacum* and the non-*Solanaceae* hosts suggests that TEV may be
321 interacting with different host factors and that the improved interaction with tobacco
322 may led to less efficient interactions with an orthologous factor, if available, in the
323 alternative hosts. In this regard, many examples exist in the plant virology literature
324 showing that host-range mutations have negative pleiotropic effects in the primary host
325 (reviewed in Elena *et al*., 2011). A particularly illustrating example is the interaction
326 between the VPg of potyviruses and the host's eIF4E (Robaglia & Caranta, 2006).
327 Translation of the viral genomic RNA into the polyprotein depends upon the correct
328 interaction between VPg and eIF4E. Mutations in eIF4E have been identified as the
329 cause of PVY resistant phenotype of pepper cultivars. Not surprisingly, PVY
330 overcomes the resistance by fixing amino acid changes in the central domain of VPg
331 that reconstitute the correct binding. These mutations pay a fitness cost in the non-
332 resistant pepper cultivars (Ayme *et al.*, 2007; Montarry *et al*., 2011).
333

334 ## Epistasis among mutations also depends on host: *G*×*G*×*E*

335 The results presented so far suggest that (*i*) epistasis is common in TEV genome and
336 that (*ii*) mutational effects depend on the host. Therefore, it is logical to expect that

337 epistasis may also vary depending on the host, that is, a significant $G \times G \times E$ component
338 may exist to determine TEV absolute fitness. To test this prediction, Lalić & Elena
339 (2012b) evaluated the strength and type of epistasis for a set of TEV double mutants on
340 four experimental hosts (*N. tabacum*, *D. stramonium*, *H. annuus*, and *S. oleracea*). The
341 10 double mutants used were randomly chosen among the larger collection described in
342 Lalić & Elena (2012a). Fig. 3 shows the distribution of epistasis across the four hosts,
343 after removing synthetic lethals from the dataset (which is justified since they are
344 irrelevant in terms of evolutionary dynamics). In short, average epistasis was positive
345 in the primary host, as already shown above, but became negative, although not
346 significant, on all alternative hosts, with a tendency to reduce in magnitude as the
347 taxonomic relatedness to the primary host decreased (Fig. 3). Furthermore, the number
348 of non-epistatic interactions was significantly larger in non-*Solanaceae* hosts.

349 These results indicate that host effects on epistasis, similarly to what happened with
350 the effect of point mutations, are modulated by the degree of genetic divergence
351 between the primary and alternative hosts. This result is in good agreement with the
352 prediction that mutations shall be more severe in poor environments and milder in rich
353 ones (You & Yin, 2002). Furthermore, mild mutations are expected to be involved in
354 negative epistatic interactions in poor environments but in positive interactions in rich
355 ones (You & Yin, 2002). Our results are in good agreement with these predictions:
356 average mutational effects are milder and mutations show positive epistasis in the
357 primary host but switch to larger effects and negative or no epistasis in alternative hosts.
358 Together, these observations suggest that the primary host, and those that are closely
359 related to it, represent rich environments for TEV while the alternative and unrelated
360 hosts represent more stressful environments. This makes sense, considering that TEV
361 has a coevolutionary history with *Solanaceae* hosts and thus its interaction with cellular

362 resources and defenses is optimal. By contrast, alternative hosts may not provide the
363 necessary resources at the right time, amount or location.

364 $G \times G \times E$ is equivalent to the concept of epistatic pleiotropy (Remold, 2012). Under
365 epistatic pleiotropy, virus populations may achieve either specialization for a single host
366 or, alternatively, become generalist with no cost, depending on the host in which they
367 evolve. More importantly, no-cost generalists can evolve despite the existence of true
368 genetic trade-offs. We will discuss this possibility in large in the next section.

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

373 $$W \sim G + G \times G + E + G \times E + G \times G \times E \qquad (5)$$

374

375 ## Pleiotropy and epistasis

376 Pleiotropy and epistasis have strong parallelism because for both interactions, the effect
377 of an allele depends on its context: the host species for pleiotropy and the virus' genetic
378 background for epistasis. Indeed, it has been postulated that pleiotropy is a prerequisite
379 for epistasis (Martin et al., 2007; De Visser *et al*., 2011). This dependence is easy to
380 understand for the case of sign pleiotropy, where mutations with a positive effect in the
381 new host have a negative effect in the primary one (Remold, 2012). In the context of
382 compensatory evolution, sign pleiotropy is a precondition for sign epistasis (Fig. 1),
383 because it allows for the negative pleiotropic effects of previously selected mutations to
384 be compensated by additional ones (De Visser *et al*., 2011). Therefore, the question to
385 be answered is whether a positive association exists between the tendency of mutations
386 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 epistasisness or E_x, was evaluated as the average of the squared epistasis coefficients ε_{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 $P_x = \left\langle \left(\frac{W_{x,H}}{W_{x,N.tabacum}} - 1 \right)^2 \right\rangle$, 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 (ρ_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

412 to a situation of no-cost generalism. A long history of evolution in the primary host
413 could have resulted in an adaptive walk towards a host-specific fitness peak involving
414 most, if not all, viral loci. Such population could find itself many mutational steps,
415 through an adaptive valley, away from reaching a generalist peak.
416

417 ## Conclusions

418 Here we have reviewed recent data showing that the expected effect on viral fitness of
419 point mutations depends on the genetic background where they appear in as much as on
420 the host species being infected by the virus. In other words, the reviewed data show
421 that the virus genotype and the host species interact in a non-linear manner to determine
422 the fitness of a potyvirus. The implications of these observations for our understanding
423 of emerging plant viral infections are multiple, but basically all hint on the
424 unpredictability at the level of effect of individual mutations: in the light of information
425 collected on the primary host, one can not anticipate which particular viral genotypes
426 will be more likely to emerge in related hosts. However, the observation of sign
427 pleiotropy in unrelated hosts leaves some room for predictability at least at the level of
428 classes of mutations: beneficial mutations, as a class, in the primary host may become
429 deleterious in new ones.

430 Finally, the existence of epistatic pleiotropy on host usage together with the
431 dominance of reciprocal sign epistasis in the viral genome create rugged adaptive
432 landscapes that may trap viral populations in local peaks and impede their escape
433 towards no-cost generalists. In some sense, these are good news, since the difficulty to
434 generate no-cost generalists reduces the likelihood of successful spillovers, as most
435 genotypes will necessarily pay a large fitness cost after infecting a new host.
436

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

Agudelo-Romero P, de la Iglesia F, Elena SF, 2008. The pleiotropic cost of host-specialization in *Tobacco etch potyvirus*. *Infection, Genetics and Evolution* **8**, 806-814.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P, 2004. Emerging infectious diseases of plant pathogens pollution, climate change and agrotechonological drivers. *Trends in Ecology and Evolution* 13, 535-544.

Ayme V, Petit-Pierre J, Souche S, Palloix A, Moury B, 2007. Molecular dissection of the *Potato virus* Y VPg virulence factor reveals complex adaptations to the *pvr2* resistance allelic series in pepper. *Journal of General Virology* **88**, 1594-1601.

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

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

Cleveland S, Haydon DT, Taylor L, 2007. Overviews of pathogen emergence: which pathogens emerge, when and why? *Current Topics in Microbiology and Immunology* **315**, 85-111.

462 Crow JF, Kimura M, 1970. An Introduction to Population Genetics Theory. New York,
463 USA: Harper and Row.
464 Dawid A, Kiviet DJ, Kogeranu M, de Vos M, Tans SJ, 2010. Multiple peaks and
465 reciprocal sign epistasis in an empirically determined genotype-phenotype landscape.
466 *Chaos* **20**, 026105.
467 DeFillipis EC, Villareal LP, 2000. An introduction to the evolutionary ecology of
468 viruses. In: Hurst CJ, ed. *Viral Ecology*. New York, USA: Academic Press, 126-208.
469 De Visser JAGM, Cooper TF, Elena SF, 2011. The causes of epistasis. *Proceedings of*
470 *the Royal Society B* **278**, 3617-3624.
471 Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, Lafforge G, Lalić J,
472 Pròsper A, Tromas N, Zwart MP, 2011. The evolutionary genetics of emerging plant
473 RNA viruses. *Molecular Plant-Microbe Interactions* **24**, 287-293.
474 Elena SF, Solé RV, Sardanyés J, 2010. Simple genomes, complex interactions:
475 epistasis in RNA virus. *Chaos* **20**, 026106.
476 Franke J, Klözer A, de Visser JAGM, Krug J, 2011. Evolutionary accessibility of
477 mutational pathways. *PLoS Computational Biology* **7**, e1002134.
478 Futuyma DJ, Moreno G, 1988. The evolution of ecological specialization. *Annual*
479 *Review of Ecology and Systematics* **19**, 207-234.
480 Gandon S, 2004. Evolution of multihost parasites. *Evolution* **58**, 455-469.
481 Gillespie JH, Turelli M, 1989. Genotype-environment interactions and the maintenance
482 of polygenic variation. *Genetics* **121**, 129-138.
483 Hodgins-Davies A, Townsend JP, 2010. Evolving gene expression: from *G* to *E* to *G*×*E*.
484 *Trends in Ecology and Evolution* **24**, 649-658.
485 Holmes EC, 2009. The evolutionary genetics of emerging viruses. *Annual Review of*
486 *Ecology, Evolution and Systematics* **40**, 353-372.

Holmes EC, Drummond AJ, 2007. The evolutionary genetics of viral emergence. *Current Topics in Microbiology and Immunology* **315**, 51-66.

Jenner CE, Wang X, Ponz F, Walsh JA, 2002. A fitness cost for *Turnip mosaic virus* to 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 being generalist. *American Naturalist* **144**, 833-838.

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

Kouyos RD, Silander OK, Bonhoeffer S, 2007. Epistasis between deleterious mutations and the evolution of recombination. *Trends in Ecology and Evolution* **6**, 308-315.

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

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.

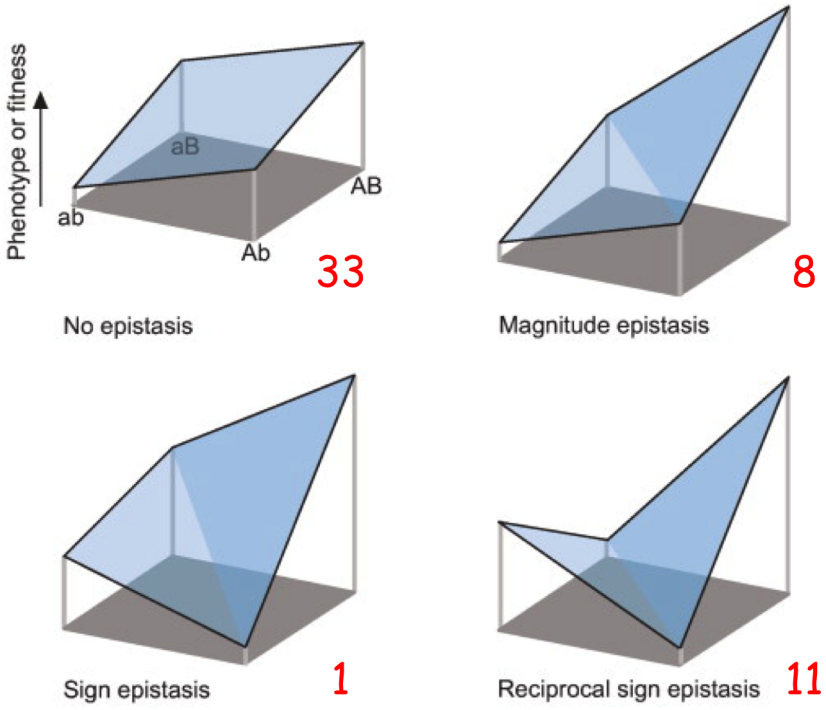
Martin G, Elena SF, Lenormand T. 2007. Distribution of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genetics* **39**, 555-560.

Montarry J, Doumayrou J, Simon V, Moury B, 2011. Genetic background matters: a plant-virus gene-for-gene interaction is strongly influenced by genetic contexts. *Molecular Plant Pathology* **12**, 911-920.

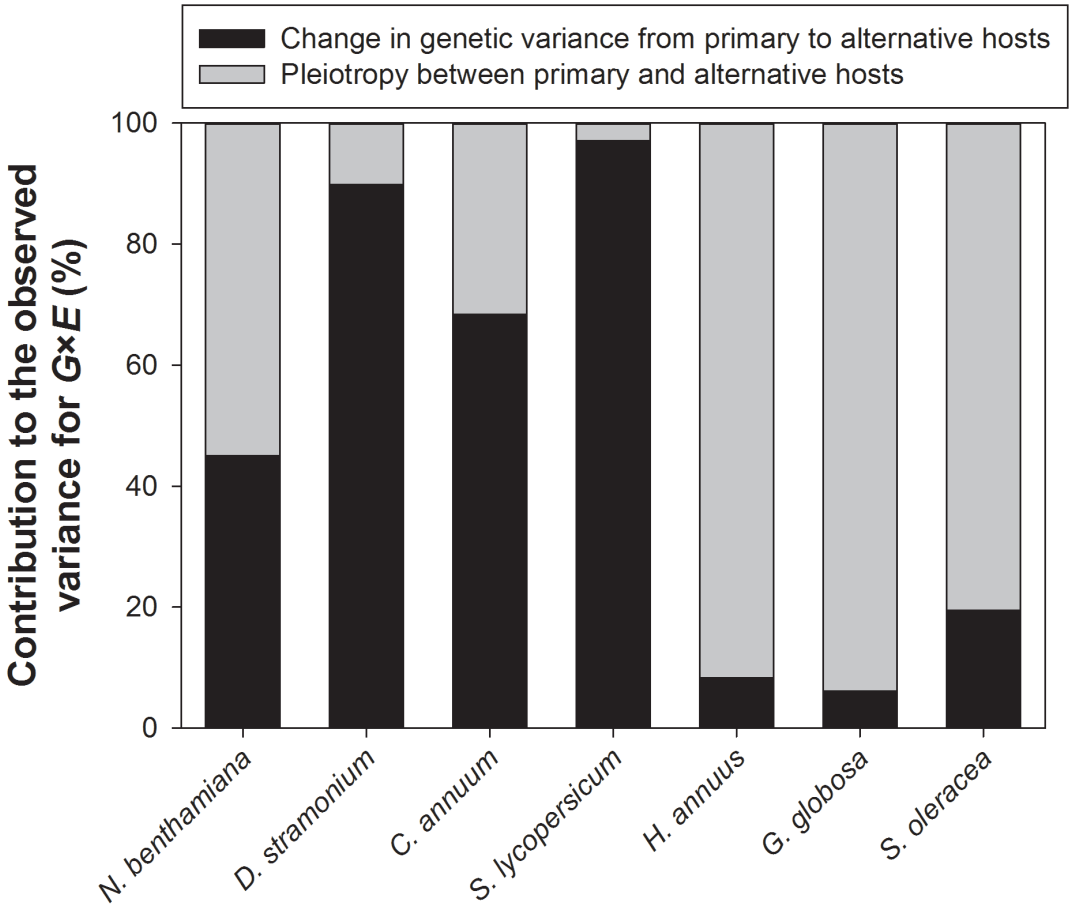
512 Poelwijk FJ, Tanase-Nicola S, Kiviet DJ, Tans SJ, 2011. Reciprocal sign epistasis is a
513 necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical*
514 *Biology* **272**, 141-144.
515 Remold SK, 2012. Understanding specialism when the jack of all trades can be the
516 master of all. *Proceeding of the Royal Society B* **279**, 4861-4869.
517 Remold SK, Lenski RE, 2001. Contribution of individual random mutations to
518 genotype-to-environment interactions in *Escherichia coli*. *Proceeding of the*
519 *National Academy of the USA* **98**, 11388-11393.
520 Robaglia C, Caranta C, 2006. Translation initiation factors: a week link in plant RNA
521 virus infection. *Trends in Plant Sciences* **11**, 40-45.
522 Robertson A, 1959. The sampling variance of the genetic correlation coefficient.
523 *Biometrics* **15**, 469-485.
524 Sanjuán R, 2010. Mutational fitness effects in RNA and single-stranded DNA viruses:
525 common patterns revealed by site-directed mutagenesis studies. *Philosophical*
526 *Transactions of the Royal Society B* **365**, 1975-1982.
527 Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R, 2010. Viral mutation rates.
528 *Journal of Virology* **84**, 9733-9748.
529 Weinreich DM, 2005. The rank ordering of genotypic fitness values predicts genetic
530 constraints on natural selection on landscapes lacking sign epistasis. *Genetics* **171**,
531 1397-1405.
532 Welch JJ, Waxman D, 2005. The *nk* model and population genetics. *Journal of*
533 *Theoretical Biology* **234**, 329-340.
534 Whitlock MC, 1996. The Red Queen beats the jack-of-all-trades: the limitations on the
535 evolution of phenotypic plasticity and niche breadth. *American Naturalist* **148**, S65-
536 S77.

537 Whitlock MC, Phillips PC, Moore FGB, 1995. Multiple fitness peaks and epistasis.
538 *Annual Review of Ecology, Evolution and Systematics* **26**, 601-629.
539 Woolhouse MEJ, Haydon DT, Antia R, 2005. Emerging pathogens: the epidemiology
540 and evolution of species jumps. *Trends in Ecology and Evolution* **20**, 238-244.
541

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).

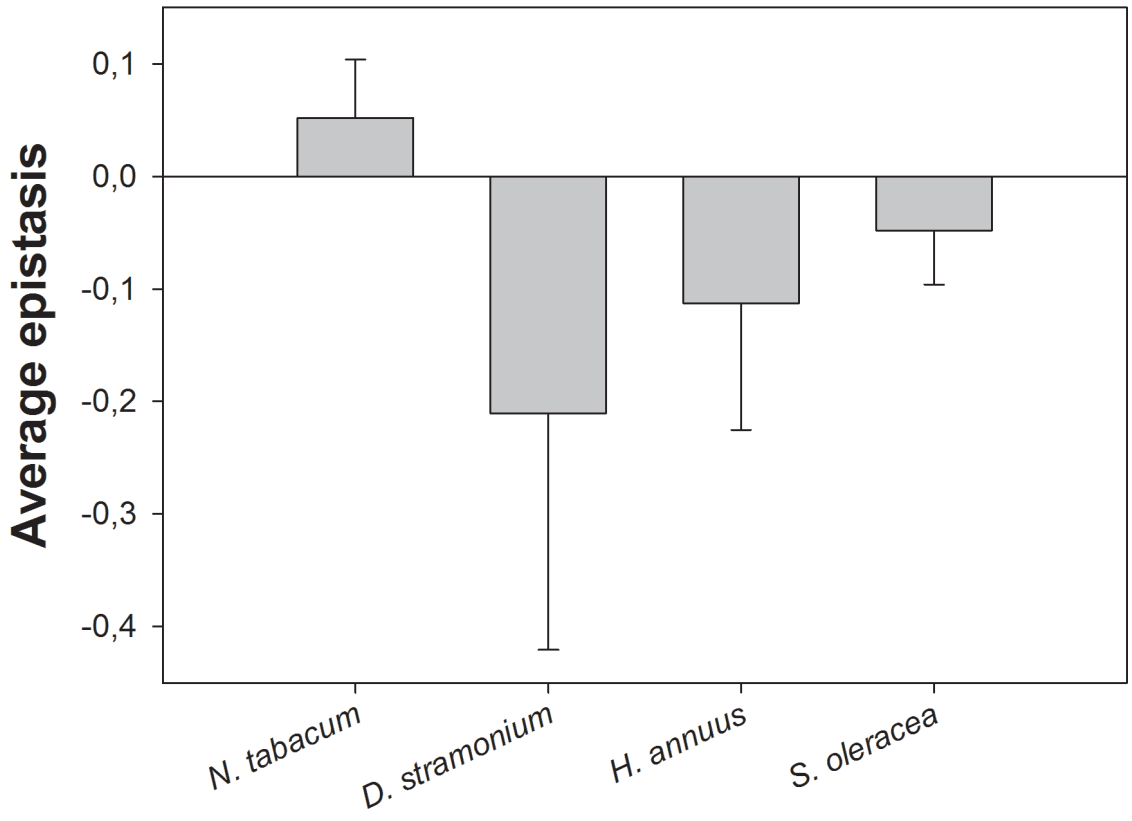


549 **Figure 2** Contribution of sign pleiotropy and changes in variance for fitness to the
550 observed variance in $G \times E$ when comparing the primary host (*N. tabacum*) and the
551 alternative ones. For *Solanaceae* hosts, $G \times E$ is mostly generated by changes in genetic
552 variance for fitness across hosts; by contrast, sign pleiotropy is the main cause of $G \times E$
553 for non-*Solanaceae* hosts. Data taken from Lalić *et al*. (2011).
554

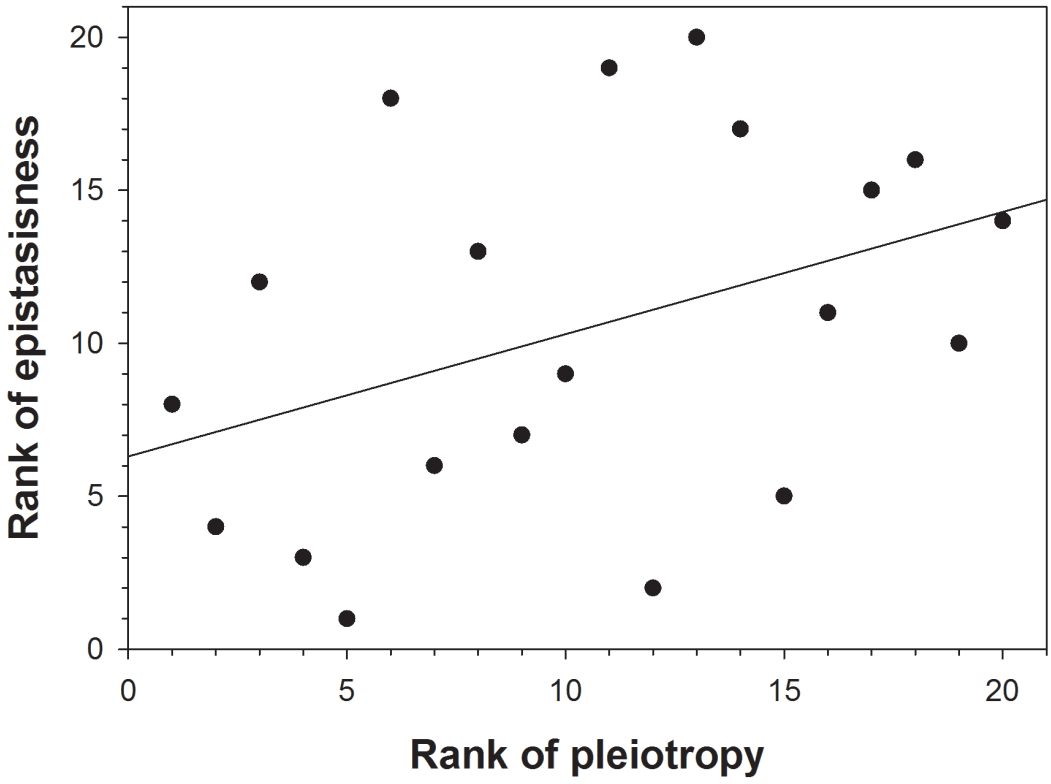


555
556

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).



562 **Figure 4** Relationship between the tendency of a mutation to be involved in epistatic
563 interactions (epistasisness) and its pleiotropic effect across hosts. Data taken from Lalić
564 *et al*. (2011) and Lalić & Elena (2012a, 2012b).
565



566

567

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).

Host species	Expected *W*	Median	Standard deviation	Skweness	Kurtosis	Lethal	Deleterious	Neutral	Beneficial
N. tabacum	1.331	1.327	0.021	-1.974	4.608	0	6	14	0
N. benthamiana	1.315	1.319	0.065	-3.949	16.879	0	10	10	0
D. stramonium	1.365	1.380	0.054	-1.566	1.364	2	15	3	0
C. annuum	1.246	1.297	0.142	-1.037	-0.389	2	0	9	11
S. lycopersicum	1.350	1.418	0.041	-0.768	0.062	8	0	2	10
H. annuus	1.020	1.020	0.044	0.527	0.579	0	0	15	5
G. globosa	0.725	1.010	0.042	0.997	0.561	0	0	17	3
S. oleracea	0.976	0.962	0.052	1.479	1.915	0	0	17	3

568
569

570

Table 2 Different components of the variance for fitness evaluated in eight susceptible hosts. Data taken from Lalić *et al*. (2011).

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

σ_G^2 = genetic variance for fitness on each host.
$\rho_{G_H G_{N.tabacum}}$ = genetic correlation for fitness across host *H* and the primary host *N. tabacum*.
$\sigma_{G\times E}^2$ = variance for the interaction between viral genotype and host (*G*×*E*) computed using Eq. 4.

571