VNIVERSITAT DE VALÈNCIA Facultat de Ciències Biològiques

Nonlinearities in plant RNA virus fitness

Tesis presentada por

Jasna Lalić

para optar al grado de doctor en Ciencias Biológicas por la Universitat de València
tras haber cursado el Programa Oficial de Postgrado en Biodiversidad

Fdo: Jasna Lalic en Valencia, 20 de Septiembre de 2012

MINISTERIO
DE ECONOMÍA
Y COMPETITIVIDAD

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

INSTITUTO DE BIOLOGIA MOLECULAR Y CELULAR DE PLANTAS

Santiago F. Elena Fito, Doctor en Ciencias Biológicas y Profesor de Investigación del Consejo Superior de Investigaciones Científicas (CSIC) en el Instituto de Biología Molecular y Celular de Plantas (IBMCP), centro mixto del CSIC y de la Universidad Politécnica de Valencia.

CERTIFICA

Que Dña. Jasna Lalić, Licenciada en Ciencias Biológicas, ha realizado bajo mi supervisión la tesis doctoral titulada "*Nonlinearities in plant RNA virus fitness*".

Y para que así conste, firmo la presente en Valencia, a 30 de Julio de 2012

Prof. Santiago F. Elena, PhD
santiago.elena@csic.es
http://bioxeon.ibmcp.upv.es/EvolSysVir

Campus UPV, CPI 8E
Ingeniero Fausto Elío, s/n
46022 València
Tel: 963 877 895
Fax: 963 877 859

Acknowledgements

Pasar cuatro años en València ha sido un placer en todos los sentidos y una experiencia que me ha marcado profundamente; y no hubiera sido así sin las siguientes personas. La mayor contribución en este trabajo la tiene Santi y por lo tanto le doy muchísimas gracias: por aceptarme en su laboratorio, por enseñarme y guiarme y, sobre todo, por darme libertad y aliento, apoyo y compresión inmensa. He tenido suerte de tener un jefe extraordinario como científico y, más importante, como persona.

Gracias a Paqui por estar siempre de buen humor y ayudarme mucho con *la selva* y otras plantas, y a Àngels por su ayuda para *morterizar* las plantas y hacer las *minis*. Gracias a José Antonio Daròs por estar siempre dispuesto a darme consejos y recetas de biología molecular. Gracias a Manuel Serra por ayudarme con toda la burocracia del doctorado.

Gracias a Susana por enseñarme castellano, qPCR y otras cosas de biología molecular, pero más que nada, por hacerme relajar y reír, por aguantarme, apoyarme, y aconsejarme. Es decir, por ser mi mejor amiga.

Gracias a Stéphanie por sus consejos y varias ayudas con la estadística y evolución. Y más importante, por su amistad y su amable comprensión.

Gracias a Ricardo por su alegría y por compartir conmigo tiempos de descanso y peculiaridades españolas y valencianas.

Gràcies a l'Arcadi i el Lluso per ser els meus amics; per ajudar-me, per preocupar-se per mi, per regalar-me coses i per enriquir la meva vida amb els viatges, vols i sonriures.

Gracias a la gente anónima por subir gratuitamente libros a Internet. Google is my friend.

Najboljim prijateljima: Marinu, Goranu i Tanji, neizmjerno hvala što su uvijek uz mene.

Puno ljubavi i zahvale mojim roditeljima i sestri na apsolutno svemu.

Table of Contents

PROLOGUE
PART I
CHAPTER I – GENERAL INTRODUCTION 1
1.1. Emerging viruses 1
1.2. RNA viruses 2
1.2.1. The genome 2
1.2.1.1. Overlapping reading frames 4
1.2.1.2. RNA secondary structures 5
1.2.1.3. Multifunctional proteins 5
1.3. Plant RNA virus life cycle 6
1.3.1. Replication of positive single stranded RNA viruses 7
1.3.2. Host defences and counter defences 8
1.3.3. Transmission 8
1.4. General features of *Tobacco etch virus* 8
1.5. The concept of viral fitness 13
1.6. Key evolutionary concepts 15
1.6.1. Generalism *vs.* specialism 15
1.6.2. Genotype-by-environment ($G\times E$) interactions 18
1.6.2.1. The nature and causes of $G\times E$ 19
1.6.3. Source of variation: mutation 22
1.6.3.1. Mutation in RNA viruses 23
1.6.3.2. Mutational fitness effects 25
1.6.4. Epistasis 28
1.6.5. Robustness 31
1.6.6. Adaptive fitness landscapes 36
CHAPTER II – OBJECTIVES 42
PART II: $G\times E$, $G\times G$ and $G\times G\times E$ interactions
CHAPTER III – $G\times E$ INTERACTIONS 43
Effect of host species on the distribution of mutational fitness effects for an RNA virus
CHAPTER IV – $G\times G$ INTERACTIONS 56
Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus
CHAPTER V – $G\times G\times E$ INTERACTIONS 67
Epistasis between mutations is host-dependent for an RNA virus
PART III: Adaptive fitness landscape
CHAPTER VI – EMPIRICAL FITNESS LANDSCAPES 75
Empirical fitness landscape for the adaptation of an RNA virus to *Arabidopsis thaliana*
GENERAL SUMMARY 87
APPENDIX:
SUMMARY 94
RESUMEN EN CASTELLANO 97
BIBLIOGRAPHY 101

PROLOGUE

This thesis is written in a deductive form; starting from the introduction to the worldwide problem of emergent infectious diseases; describing in continuation the properties of RNA viruses and evolutionary biology terms and parameters relevant to addressing the aforementioned problem of virus emergence. All topics elaborated in the introduction part are mutually highly interconnected, some even being synonymous.

Viruses are the most diverse and the most abundant organisms on the planet (Edwards & Rohwer 2005), infecting organisms from all three domains of life. Viruses can be defined as *obligate intracellular parasites or symbionts that possess their own genomes encoding information required for virus reproduction and, hence, a degree of autonomy from the host genetic system, but do not encode a complete translation system or a complete membrane apparatus* (Koonin 2003). Due to their parasitic life cycle, viruses have evolved numerous, complex and fascinating ways of interacting with their hosts, which in turn enable them to switch between hosts and become emergent. Emergent viruses are those that had "crossed species barriers", that is, expanded their host range to other species. Emergent virus diseases of viral aethiology nowadays represent a major threat to public health and to the agronomy. Among emerging infectious diseases of plants, viruses cause almost a half (Anderson *et al.* 2004).

The majority of both past and ongoing pandemics are caused by RNA viruses (Domingo & Holland 1997). Due to their great evolvability; a consequence of combining highly error-prone replication, large population sizes and rapid replication rates (Elena & Sanjuán 2008), RNA viruses have great capacity of adaptation to environmental challenges such as new hosts, resistance genes and antiviral treatments (Holmes 2009). Understanding the evolutionary mechanisms by which a virus may become an emergent one is pivotal for the rational design of control strategies and antiviral therapies (McDonald & Linde 2002). Herewith, the aim of this thesis is to explore the evolutionary genetics underlying the empirical phenomena of RNA viruses switching their hosts.

Thus, can virus emergence and jump to a new host species be predicted by knowing its phenotype, *i.e.*, fitness, in its natural host? If so, then the architecture of virus fitness would be determined only by its genotype (the *G* component; *i.e.*, mutations in the genome) and environment (the *E* component; *i.e.*, host). Still, interactions between these components may exist and compromise the predictability of

virus phenotype in an alternative host. How mutations affect the fitness of viral populations is essential to understanding viral emergence and adaptation to new hosts. The widespread observation that the majority of mutations are deleterious coincides with the theoretical prediction that an organism will be well adapted to its particular environment, so that any genomic change would represent a move-away from the optimal phenotype. In reality, this view is overly simplified; mutational fitness effects constitute a continuum and are conditional upon the environment and genetic background; effects commonly referred to as genotype-by-environment ($G \times E$) genetic ($G \times G$ or epistasis) and $G \times G \times E$ (*i.e.*, epistasis-by-host) interactions. In this Thesis the contributions of genetic and environmental components to the architecture of viral fitness and, in the final part, the description of an empirical fitness landscape, was addressed using a plant positive sense RNA virus: *Tobacco etch potyvirus* (TEV). TEV pertains to the genus *Potyvirus* within the family *Potyviridae*. Potyviruses are the most abundant and economically the most significant plant viruses. TEV genome is composed of single-stranded RNA of about 9.5 kb that directly serves as a template for translation into a polyprotein precursor, which is further being proteolyticaly processed by its own three proteases into at least 10 mature proteins, plus an additional peptide encoded in the +2 reading frame. TEV has relatively wide host range although most of its natural hosts belong to the family *Solanaceae*.

CHAPTER I - GENERAL INTRODUCTION

1.1. Emerging viruses

Emergence of novel infectious diseases is a critical issue of public health and economic welfare (Holmes 2009). An emerging virus is "*the causal agent of an infectious disease of viral aetiology whose incidence is increasing following its first introduction into a new host population or whose incidence is increasing in an existing host population as a result of long-term changes in its underlying epidemiology*" (Woolhouse & Dye 2001). The majority of emerging and reemerging pathogens are viruses, among which numerically predominate RNA viruses, comprising 37% of all emerging and reemerging pathogens (Woolhouse & Gowtage-Sequeria 2005; Holmes 2009). Even though most of public attention has been dedicated to emerging viruses of humans and animals, emerging plant viruses are equally common and known to cause significant economic losses in crops (Roossinck 2008; Navas-Castillo *et al.* 2011). Some rough estimates put total worldwide damage due to plant viruses as high as US$ 6×10^{10} per year (Cann 2005).

A virus becomes emergent after jumping from its reservoir species to a new host and successfully establishes infection within the latter. Virus emergence results from complex dynamics of pathogen, host and environmental factors; such as transmission chains (Parrish *et al*. 2008; Elena *et al.* 2011). Hence, it is of great importance to pinpoint the contribution of these factors to disease emergence. The appearance of new diseases, and resurgence of old ones, implies the necessity for interdisciplinary involvement in the case. Within this context, experimental evolution provides an invaluable tool for exploring the processes underlying origin, emergence and spread of emerging viruses (Elena *et al.* 2011). This sort of knowledge may be of help to design intervention and prevention measures to control disease propagation and identifying and managing the host reservoirs.

One of the most intuitive and most frequently cited ideas in the prism of viral emergence is that the more closely related the host species in question, the greater the chance of virus jump between them and successful infection (DeFilippis & Villareal 2000; Longdon *et al*. 2011). Since phylogenetically close hosts share common cell-surface receptors needed for establishing a successful infection, it is thought that this mechanistic basis should underlie the increased likelihood of viral emergence to a

phylogenetically proximate host, especially in animal viruses (Baranowski *et al*. 2001; Woolhouse 2002). Still, spillovers or epidemic infections have occurred between hosts that are closely or distantly related, and there is no rule for predicting the susceptibility of a new host (Parrish *et al.* 2008). It has been suggested that the probability of successful cross-species transmission tightly relates to the probability of exposure (Holmes 2009a), which depends on the ecology of the two hosts and of the transmission biology of the virus, including any relevant vectors. However, upon initial infection of the new host, the infectivity, viral load and transmissibility is usually very low (Elena *et al.* 2011), so in order to expand its host range, a virus needs to adapt to its new host. Genetic mechanisms underlying posterior adaptation are described in Section 3.

A prerequisite for viral emergence is the existence of host-range mutants within the standing genetic variation in the reservoir host (Elena *et al.* 2011). The amount of standing genetic variation would depend (*i*) on the rates of mutation and recombination, (*ii*) on the distribution of mutational effects on viral fitness and (*iii*) on the strength of genetic drift and gene flow among subpopulations.

1.2. RNA viruses

Out of 2000 known species of viruses, approximately a half use RNA as their genomic material (the other half are the DNA viruses) (Holmes 2009). Moreover, positive-strand RNA viruses encompass over one-third of all virus genera and most of plant viruses have such genome. RNA viruses are the only organisms that store their genetic information in RNA. Other organisms use RNA only temporary: for regulatory and various other tasks.

Here I summarize some of the distinguishing features of RNA viruses with special interest in their genome characteristics and replication. These are the mechanistic reasons that underlie and/or constrain the great adaptability hallmark of RNA viruses reflected in the following population-genetics parameters: *i*) high mutation rates, *ii*) large population sizes and *iii*) short generation times that enable them to generate great genetic variability and adapt to many different environments (*i.e.*, hosts).

1.2.1. The genome

One of the most prominent features of RNA viruses is the restricted size of their genome. All RNA viruses have small genomes, spanning the range from 2.3 kb in *Ophiostoma novo-ulmi* mitovirus 6-Ld up to *ca*. 30 kb in coronaviruses (Figure 1);

typically having around 10 kb (Holmes 2009a). In comparison, genome sizes of DNA viruses can differ by as much as ~3 orders of magnitude. The mechanistic reasons that underlie the genome size limitation observed in RNA (and ssDNA) viruses are unknown. It has been suggested that the ultimate size of single-stranded RNA viruses is limited by the fragility of RNA and the tendency of long strands to break (Cann 2005). In the case of plant RNA viruses, which genome size does not exceed 15 kb, it has been suggested that plasmodesmata (through which the virus moves cell-to-cell) constrains the genome size (Roossinck 2008a). Since a small genome is replicated faster than a long one, this property has important repercussions on the RNA virus population size as will be discussed further on. Moreover, restricted genome size in RNA viruses is selected by intrinsically error-prone replication (Holmes 2003). In this sense, longer genome presents bigger mutational target, so great number of mutations (majority of which are deleterious) will in final have a negative impact on virus evolution. In comparison, dsDNA viruses that replicate with high fidelity, bear far lower mutation rates and are able to achieve much larger genome sizes. In RNA virus world, the exception are animal *Nidovirales* which possess the largest known RNA genomes (from 26 to 32 kb) and a set of enzymes involved in proofreading and repair linked to their replicase (Gorbalenya *et al.* 2006; Minskaia *et al.* 2006). This hints that *Nidovirales* are able to reduce mutational load and thus evolve larger genome sizes. From the experimentalist perspective, small genomes have an advantage of being easily sequenced, thus allowing for experimental observations to be explained mechanistically by attributing certain phenotypic response (such as, for example, change in viral accumulation, infectivity, transmissibility, etc.) to particular mutation(s). In other words, it is relatively easy to map genotypes into phenotypic space.

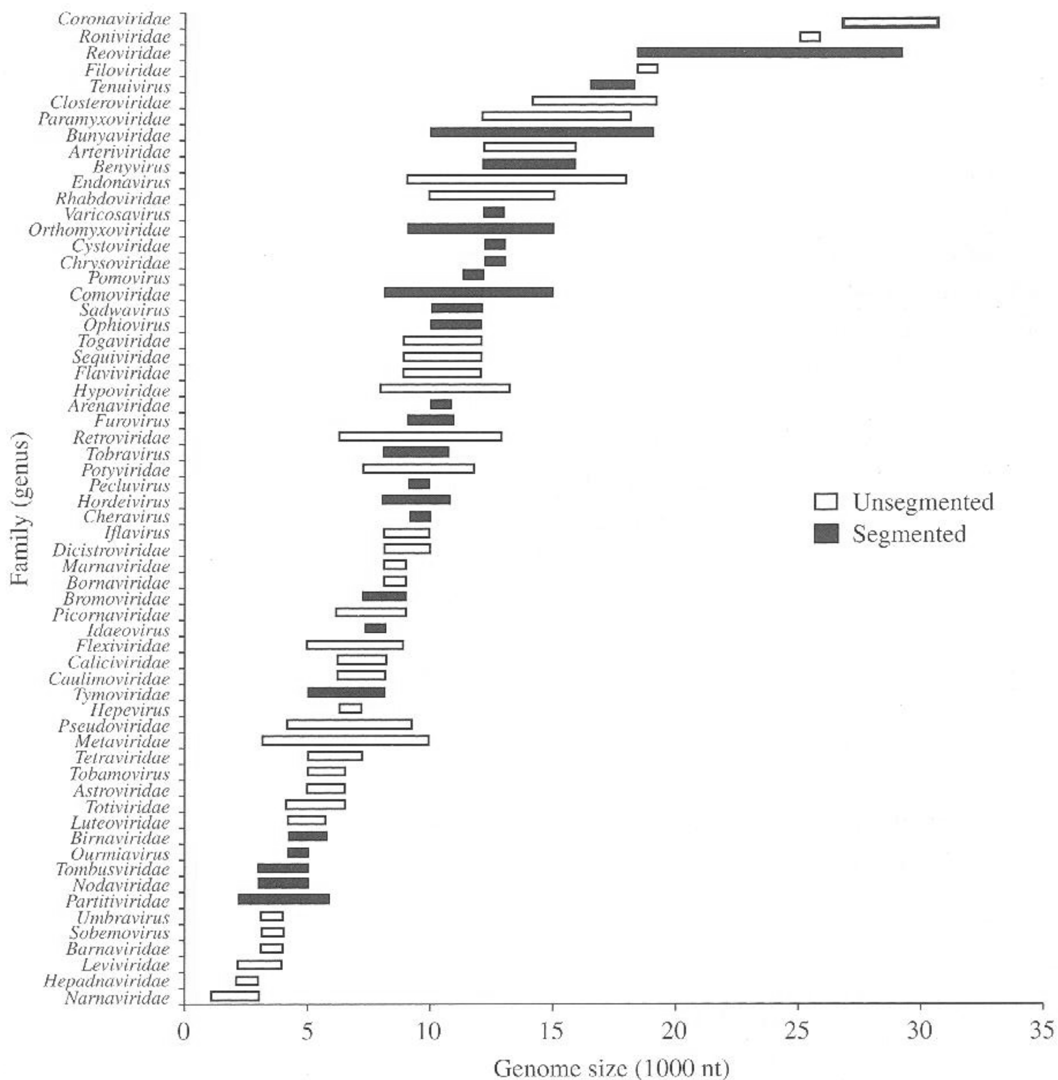


Figure 1. The distribution of genome sizes among different families (and some genera) of RNA viruses. Segmented (closed bars) and unsegmented (open bars) genomes are indicated (Holmes 2009a).

RNA viruses resemble low complexity in terms of their genome structure and number of proteins it codes for; according to their parasitic life history. They have no genetic redundancy and the compactness of their genetic information is reflected in different levels: *i*) trough common existence of overlapping reading frames, *ii*) via secondary RNA and higher-order structures and *iii*) multifunctional proteins.

1.2.1.1. Overlapping reading frames

Overlapping reading frames are created *de novo* by mutations within a coding sequence that leads to the expression of a novel protein in another reading frame. Mechanisms

that lead to gene overlap include ribosomal frameshifting, the use of non-AUG start codons and RNA splicing (Belshaw *et al.* 2007). It has been suggested that the proteins created by gene overlaps are typically accessory proteins that play a role in viral pathogenicity or spread (Rancurel *et al.* 2009). These overlaps are typically assumed to be a form of genome compression, allowing the virus to increase its repertoire of proteins without increasing its genome length (Barrell *et al.* 1976; Scherbakov & Garber 2000; Chung *et al.* 2008). In addition, Belshaw *et al.* (2007) showed that viruses with longer genomes tend to have less gene overlap compared to shorter RNA viruses, presumably because their genome already comprises enough protein diversity. However, other authors argue that the gene overlap evolved because the size of the capsid physically limits the genome length (Chirico *et al.* 2010).

1.2.1.2. RNA secondary structures

Defined RNA secondary and higher-order structures are shown to play fundamental roles in many different cell processes; including transcription, translation, RNA localization, splicing, transport, stability and catalytic activity. RNA viruses, as obligate intracellular parasites, need to interact with the cellular synthetic machinery for completing these processes. So, in that sense, functional secondary structures within the genome of RNA viruses must be crucial for the various stages of the viral life cycle and in evasion of host defences. Important viral RNA structures include internal ribosome entry sites (IRES) (Witwer *et al*. 2001), packaging and splicing signals, pseudoknots, transfer RNA mimics, ribosomal frameshift motifs, and cis-regulatory elements (Cann 2005).

The observation of widespread RNA secondary structures argues against selective neutrality of synonymous sites (Simmonds & Smith 1999; Plotkin & Kudla 2011). Synonymous changes that do not affect the translated product may still result in non viable virus progeny, since RNA structure has been increasingly found to influence virus pathogenesis and survival (Simmonds *et al.* 2004, Coleman *et al*. 2008). In addition, it has been shown that structured regions of virus genome bearded more mutations than nonstructured indicating that the presence of RNA secondary structure in the virus genome affects viral mutation rate (Pathak & Temin 1992; Pita *et al.* 2007).

1.2.1.3. Multifunctional proteins

Finally, plant RNA viruses often comprise multifunctional proteins. One of such is HC-Pro from viruses within the genus *Potyvirus* (Maia *et al.* 1996) that in addition to being self-cleaving proteinase (Carrington & Herndon 1992), is also involved in a number of infectious processes such as aphid transmission (Govier *et al.* 1977), cell-to-cell (Rojas *et al.* 1997) and long-distance movements (Saenz *et al.* 2002), suppression of gene silencing (Llave *et al.* 2000), synergism between co-infecting viruses (Pruss *et al.* 1997) and symptom development (Redondo *et al.* 2001). *Potyvirus* coat protein (CP) is also involved in multitude of interactions with factors encoded by the virus, host plant, or viral vector (biological transmission agent) that influence the infection and epidemiological facets of plant disease (Callaway *et al.* 2001).

1.3. Plant RNA virus life cycle

Life history of a plant RNA viruse consists of several sequential intra-host steps: cell entry, replication, cell-to-cell movement and systemic (long-distance) movement (Hull 2002); in addition to inter-host transmission mediated by vectors (Figure 2). Among these, the mechanism of systemic propagation through the phloem is the most poorly understood (Ruiz-Medrano *et al.* 2012). To move from cell-to-cell, plant viruses encode special proteins; the movement proteins, that modify the plasmodesmata because the viruses are too big to pass trough (Boevink & Oparka 2005). All these steps (Melcher 2012) require counteracting host defenses as well as intimate interactions between the viral genomes and/or virus-encoded proteins and numerous host's transcriptional, translational, and macromolecular trafficking factors (Harries & Ding 2011).

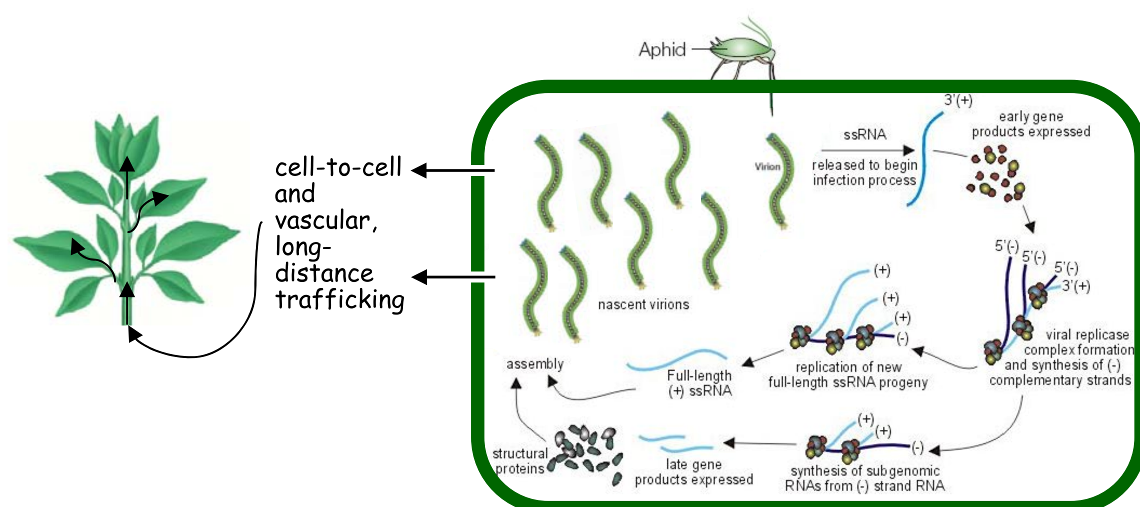


Figure 2. Schematic representation of a plant ssRNA(+) virus life cycle.

1.3.1. Replication of positive single-stranded RNA viruses

Unlike other RNA viruses, positive-strand RNA viruses do not encapsidate their RNA-dependent RNA polymerase (RdRp). Instead, upon entering the host cell, the virus decapsidates and the positive single-stranded (+) genomic RNA serves directly as a template (*i.e.*, analogous to mRNA) for translation by host ribosomes. The polyprotein product of translation further self-processes giving rise to structural and nonstructural (*i.e.*, replication-associated) proteins. Once the viral RdRp and other essential proteins are synthesized, the genomic RNA is used as a template for transcription (*i.e.*, replication). First, a minus-stranded (-) RNA intermediates are produced, followed by the synthesis of (+) RNA. This results in asymmetric production of an excess of genomic (+) over (-) strand RNA, characteristic of all (+) strand viruses. This may be accomplished through transition of the replicase from competence for (-) to (+) strand synthesis by the recruitment of additional host factors. For (+) strand transcription, similarities in *cis*-acting sequence motifs and RNA secondary structures within 5' termini of genomic (+) strands have been shown to participate in binding of host factors (Pogue *et al.* 1994). Even though it has not been clearly demonstrated, it is usually thought that viral transcription and translation are spatiotemporally coupled events. Virus genomes are further packaged along with structural proteins into virions.

Virus replication mode is important issue in virology because it has direct repercussions on the number and distribution of mutations in virus populations (Elena *et al.* 2008). Viruses can replicate either linearly, via Luria's stamping machine model (Luria, 1951), using solely a few (-) strand intermediates as templates for the production of (+) strand progeny; or exponentially (geometrically) (French & Stenger 2003), whereby both (+) and (-) strands are used as templates for the production of (+) strand progeny. Both strategies are mutually non-exclusive. Neglecting the purifying selection, under the stamping machine model, the frequency of mutant progeny depends only on the genomic mutation rate, whereas if the replication is solely exponential, the frequency of mutants produced additionally depends on the number of replication events required to produce N viral genomes per cell. As a result, the replication would be faster but more mutants would be produced if the replication would be exclusively exponential because the transcription errors would be geometrically amplified (Malpica *et al*. 2002). Empirically, it has been found that the stamping machine is the predominating mode of replication for dsRNA (Chao *et al.* 2002), ssRNA(+) (Martínez *et al.* 2011) and DNA viruses (Denhardt & Silver 1966), with exponential mode

contributing only to the minor fraction (*i.e.*, less than 10%) of the viral progeny. These studies are evidences that RNA viruses shift the balance on replication accuracy depending only on the RdRP error rate rather than on replication speed.

1.3.2. Host defences and counter defences

Plants actively resist viral infection trough the expression of resistance (*R*) genes and via RNA silencing. For many plant viruses, *R* genes has been described (Martin *et al.* 2003). On the other hand, RNA silencing in plants and its animal counterpart RNA interference (RNAi) is a mechanism for sequence-specific gene silencing triggered by dsRNA (Ratcliff *et al.* 1997). During virus replication, dsRNAs are produced that trigger virus-induced RNA silencing (VIGS) pathway (Ding & Voinnet 2007). As a counter-defense, viruses have commonly evolved proteins that suppress the silencing pathway (Kasschau & Carrington 1998) such as the HC-Pro of potyviruses or the 2b from *Cucumber mosaic cucumovirus* (CMV) (Voinnet 2005).

1.3.3. Transmission

Animal viruses transmit passively, trough air, or via vectors, and enter their host by fusing with the cell membrane. By contrast, due to robust and thick plant cell wall, plant viruses developed the strategy of using vectors that feed on plants in order to propagate between hosts. Most common vectors are arthropods; although nematodes, fungi or even bacteria can also transmit viruses (Walkey 1991). Only relatively few viruses, such as *Tobacco mosaic tobamovirus*, rely on passive mechanical transmission from plant to plant. Besides, some viruses are transmitted trough seed and pollen or by vegetative propagation, which represents a good strategy for surviving the winter. Transmission is a specific process: a particular virus can be transmitted by only one vector species or genera and not by anothers. Moreover, the interaction between a virus and its specific vector is variable: some viruses just attach on their vector mouthparts, while some other multiply both in the cells of their insect vectors and in plants.

1.4. General features of *Tobacco etch virus*

The model virus used in this thesis is *Tobacco etch virus* (TEV). TEV is a member of the genus *Potyvirus* within the family *Potyviridae* belonging to picornavirus supergroup of positive-strand RNA viruses (Figure 3). *Potyvirus* is the largest genus in the family.

Potyviruses form the largest family (30%) of plant viral pathogens and produce severe crop losses worldwide (Shukla *et al.* 1994; Rajamäki *et al.* 2004). One of distinguishing features of potyviruses is that gene order and protein sequences are conserved throughout the family (King *et al.* 2012).

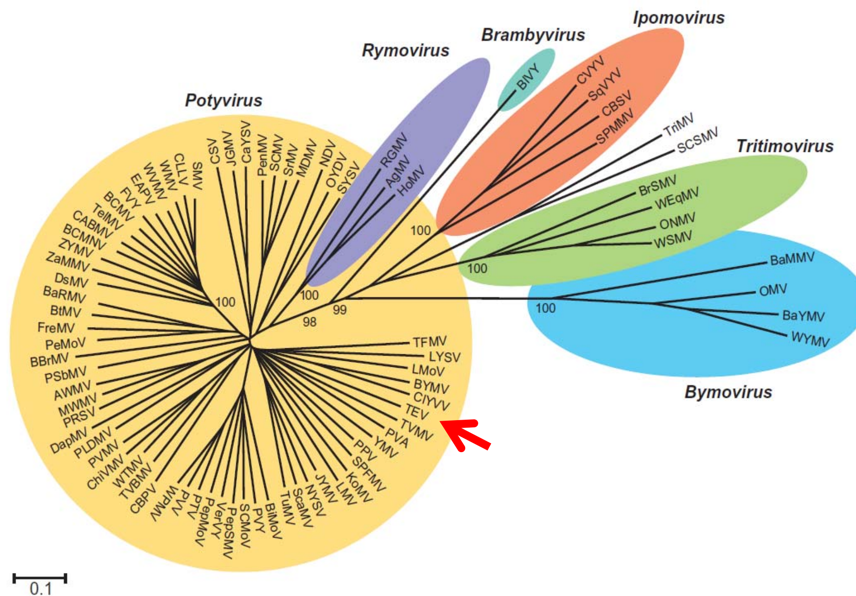


Figure 3. Unrooted phylogenetic tree based on the codon-aligned nucleotide sequences of the polyproteins of fully-sequenced members of the family *Potyviridae* (King *et al*. 2012). The position of TEV in the tree is denoted with an arrow.

TEV is distributed worldwide, but occurs with more prevalence in North and South America (http://www.dpvweb.net/dpv/showdpv.php?dpvno=258).

TEV posses relatively broad host range: it infects around 150 species from 19 families (Shukla *et al.* 1994), among which *Solanaceae* members (such as tobacco, pepper, tomato, etc.) are its primary hosts.

Typical symptoms induced by TEV on its primary hosts are stunting and mottling, necrotic etching and leaf malformations (Shukla *et al.* 1994; Figure 4). The nature and extent of symptoms depend upon the virus and particular virus strain as well as the specific host species.

Figure 4. Symptoms in a *Nicotiana tabacum* plant infected with TEV.

Potyviruses are transmitted in a non-persistent manner by aphid insects that feed on plants (Figure 5). The virus produces a helper component protein, HC-Pro, that "glues" the virions to aphid stylets. The adsorbed virus is released from the bucal apparatus of the aphid when the insect punches another plant in order to feed. It does not replicate nor circulate within the vector body (Pirone & Blanc 1996). *In vitro*, transmission to most hosts is readily accomplished by mechanical inoculation.

Figure 5. An aphid (*Myzus persicae*) responsible for *Potyvirus* transmission.

TEV is an unsegmented (monopartite) virus. It replicates in the cytoplasm and packs into a filamentous, flexuous nucleocapsid with a length of about 800 nm and a width of approximately 13 nm (Figure 6). Virions consist of genomic RNA with a 3' poly(A) tail and 5' covalently attached genome-linked protein (the VPg) of about 24 kDa encapsidated by approx. 2000 molecules of coat protein (CP). In addition to multifunctional VPg, the CP, the cyclindrical inclusion protein (CI) also forms parts of viral particle (Oruetxebarria *et al*. 2001; Puustinen *et al*. 2002; Gabrenaite-Verkhovskaya *et al*. 2008). During infection, all potyviruses form cytoplasmic cylindrical inclusion (CI) bodies.

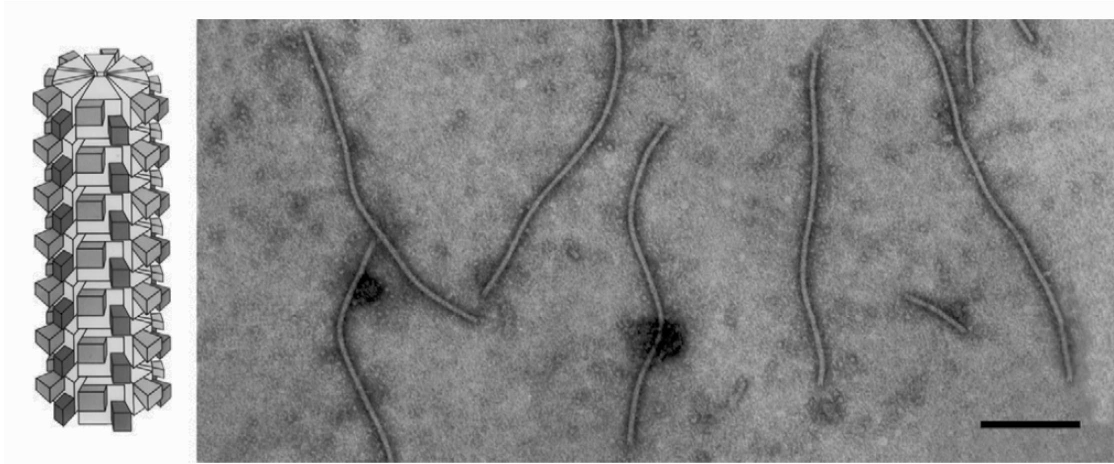


Figure 6. (Left) Schematic diagram of a potyvirus particle. The N-terminal (*ca*. 30 amino acids; large rectangle) and C-terminal (*ca*. 19 amino acids; small rectangle) of the CP molecules are exposed on the surface of the intact virus particle (from Shukla & Ward 1989). (Right) Negative contrast electron micrograph of particles of an isolate of *Plum pox virus* (from King *et al.* 2012). The bar represents 200 nm.

TEV genome is composed of positive sense RNA of 9496 nts (Figure 7) that directly serves as a template for translation into a 350 kDa polyprotein precursor which is co-translationally proteolyticaly cleaved by its own three proteases into at least 10 mature proteins (Riechmann *et al.* 1992; Urcuqui-Inchima *et al.* 2001). Seven of the nine cleavage sites are cut by the viral NIa-Pro acting both *in cis* and *in trans*, while P1 and HC-Pro act only *in cis* separating themselves from the polyprotein during the ongoing traduction (Adams *et al*. 2005). First the parental RNA is copied into a complementary minus strand. The minus strand then serves as the template for the production of progeny plus strands. The virus RNA-dependent-RNA polymerase (RdRp) uses an uridylylated form of the VPg peptide as primer for the synthesis of both (+) and (-) strand RNAs that is cleaved off as elongation of the initial complex occurs to become a 5'-genome-linked protein. Additionally, RdRp requires cellular proteins and *cis*-acting RNA elements to achieve complete replication of the viral RNA genomes (Cameron *et al.* 2009).

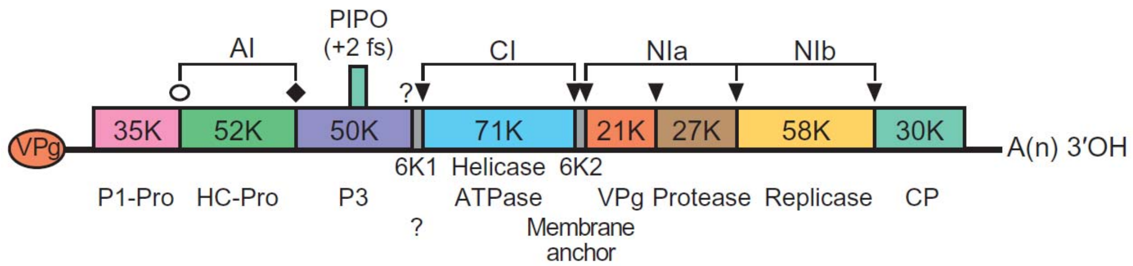


Figure 7. Genomic map of TEV (from King *et al.* 2012). The ssRNA genome is represented by a line and an open box representing the ORF translated into a polyprotein. Functions associated with the mature proteins proteolytically processed from the polyprotein are shown. VPg, genome-linked viral protein covalently attached to the 5'-terminal nt (represented by the oval at the 5' end); P1, a protein with a proteolytic activity responsible for cleavage at typically Tyr/Phe-Ser (○); HC-Pro, a protein with aphid transmission helper-component activity and proteolytic activity responsible for cleavage at typically Gly-Gly (♦); NIa-Pro, serine-like proteolytic activity responsible for cleavage at Gln/Glu-(Ser/Gly/Ala) (▼). Some of these proteins of particular viruses of the family *Potyviridae* aggregate to form inclusion bodies during infection. The protein involved and the particular type of inclusion body is shown above the genetic map; AI, amorphous inclusion; CI, cylindrical-shaped inclusion body found in the cytoplasm; NIa and NIb, small and large nuclear inclusion proteins, respectively, which aggregate in the nucleus to form a nuclear inclusion body. The small ORF PIPO is putatively translated by +2 frameshift of the polyprotein ORF, and its product is expressed as a fusion with the N-terminal part of P3.

The properties of these 10 viral proteins are listed in Table 1. Besides, a novel viral protein, PIPO, resulting from a +2 frameshift in the P3 cistron has recently been reported (Chung *et al*. 2008) to participate in cell-to-cell movement (Vijayapalani *et al.* 2012).

Table 1. Properties of potyviral proteins (from Urcuqui-Inchima *et al.* 2001; Adams *et al.* 2005; King *et al.* 2012).

Protein	Properties
P1	Of all the potyvirus proteins, P1 is the least conserved in sequence and the most variable in size. It plays a significant role in virus replication probably due to the stimulation of the gene silencing suppressor HC-Pro. A serine protease domain towards the C-terminus cleaves the P1 from the polyprotein, typically at Tyr/Phe-Ser. It is involved in viral movement and symptomatology.
HC-Pro	The HC-Pro (Helper Component-Protease) protein has roles in suppression of gene silencing and in vector transmission. A cysteine protease domain towards the C-terminus cleaves it from the remainder of the downstream polyprotein, typically at Gly-Gly. It plays roles in systemic movement, interaction with 20S proteasome and metilation of siRNAs.
P3	Involved in virus replication and appears to be significant in host range and symptom development.
PIPO	Cell-to-cell movement.
6K1	The function of this small protein is related to P3 and unknown.
CI	CI (Cylindrical Inclusion protein) has helicase activity and accumulates in inclusion bodies in the cytoplasm of infected plant cells. It binds to RNA and is involved in cell-to-cell movement.
6K2	A small transmembrane protein probably anchoring the replication complex to the ER.
VPg	VPg (Viral Protein genome-linked) is attached to the 5' terminus of the genome and belongs to a class of intrinsically disordered proteins. It plays multiple roles in the viral infection cycle. It is essential for virus replication and translation, interacting with one or several isoforms of the eIF4E translation initiation factor. It is involved in suppression of RNA silencing.
NIa-Pro	Serine-like cysteine protease responsible for cleavage of most sites in the polyprotein, typically at Gln/Glu-(Ser/Gly/Ala).
NIb	The RNA-dependent RNA polymerase.
CP	Viral coat protein that also has roles in virus movement, genome amplification and vector transmission.

The genomic RNA of TEV naturally lacks a 5′ cap structure but is nevertheless efficiently translated. Functionally analogous role to eukaryotic 5’ cap has 5’-covalently attached VPg of potyviruses that binds host initiation factors in order to recruit ribosomes (Gallie 2001; Khan *et al.* 2008; Walsh & Mohr 2011). In addition, the pseudoknot-containing domain within TEV 5′ leader functions as an IRES (Carrington & Freed 1990; Gallie & Browning 2001; Zeenko & Gallie 2005).

1.5. The concept of viral fitness

Fitness is defined as the number of viable progeny produced by a genotype that reproductively contributes to the next generation. The basic viral reproduction rate is a

crucial parameter in many epidemiological models of infectious diseases. Since evolution is always studied in the context of an organism's fitness, herewith, what do we mean by viral fitness? Whether and how selection acts in viruses? Or, in other words, what is a phenotype of a virus and upon which does the selection act? Fitness represents the combined effects of all other phenotypic properties on the capacity for the survival and reproduction by a particular genotype in a particular environment (Lenski 1991). Virus fitness is the ability of a given viral strain to replicate and accumulate within a defined environment, *i.e.*, host. It is a macroscopic property that includes components such as: receptor binding, cell entrance, uncoating, replication, transcription, translation, virion assemblage and encapsidation and cell release (Figure 2), as well as virion stability in the environment, resistance to antiviral responses and transmission or adsorption rates. The majority of these fitness components require interaction with host cellular factors (Ahlquist *et al.*, 2003).

In this Thesis, absolute viral fitness was used to measure the accumulation ability of a virus genotype in a given host, that is, the number of viral genomes produced after a given constant time per unit of total RNA extracted from the whole plant tissue. In Chapter III, absolute viral fitness was approximated as the Malthusian parameter *m*. Since growth is an exponential process, hence, differences in the growth rate should largely determine fitness. Moreover, this approximation using Malthusian growth rate per day ignores interaction between viruses within the same host and it is justifiable in the case when a host is infected with a single virus genotype and if new mutations are generated too rapidly (Smith 1999). Malthusian growth rate per day is calculated as: $= \frac{1}{t} \log \frac{N_t}{N_0}$, where *t* states for "days post infection" (dpi), *i.e.*, the time in days that a virus was left to replicate within a given host. Since the time term is included in the calculus, this measurement of growth allows for comparison between studies carried out for different viruses and over different time scales. N_t is the number of picograms (pg) of TEV RNA per 100 nanograms of total plant RNA quantified at time *t* dpi. N_0 is initial number of pg of TEV RNA used to inoculate the plant, which was 5×10^{-6} in all Chapters except in Chapter IV. TEV RNA is directly infectious when applied to susceptible host in the absence of any virus proteins (although it is about 10^6 times less infectious than virus particles (Cann 2005)). Since $N_0 << N_t$, the denominator term in the equation can be usually neglected, hence, the absolute viral fitness is proportional to decimal logarithm of pg of TEV in 100 ng of total plant RNA. In Chapters IV and V,

the absolute viral fitness is approximated by $W = e^m$. Since the dataset contained synthetic lethal genotypes, that have $N_t = 0$, it turns out that their Malthusian is $-\infty$, and thus could not be used to compute the epistasis coefficient in a classic way by assuming multiplicative effects among mutations (see below). When the trait considered is survival it is often more natural to measure epistasis as deviations from independence rather than from additivity (Crow and Kimura 1970). Survival probabilities are multiplicative if the genes act independently. In Chapter VI, fitness of the mutant genotypes was expressed relative to the wildtype TEV: $W = e^{m-\bar{m}_{wt}}$. Here we decided to use relative fitness in order to emphasize the evolutionary differences in fitness among genotypes.

1.6. Key evolutionary concepts

Here I address the consequences of viral adaptation or, in other words, virus response to the selective pressures imposed by different hosts. Generalist or specialist strategies are widely observed for many different host-parasite systems while issuing the problem of host range (*e.g.*, Turner *et al.* 2010; Bedhomme *et al.* 2012). Specialization leads to better adaptation to a given environmental selective pressure by an increase in fitness and trough the reduction in intraspecific competition (Turner & Elena 2000). Constraints of specialization are reflected in limitation of resource exploration and that even might imply a risk for population extinction (Timms & Read 1999). The existence of genotype-by-environment interaction ($G \times E$) for fitness is prerequisite for local adaptation (Kawecki & Ebert 2004). Out of several causes that can generate $G \times E$, the most important for local adaptation is antagonistic pleiotropy, whereby the mutations have opposite effects on fitness in different hosts (Elena & Lenski 2003). Such antagonistic pleiotropy implies that no single viral genotype is superior in all hosts, leading to trade-offs in adaptation to different hosts.

1.6.1. Generalism *vs.* specialism

The physical environment of a virus, like any other parasite, is its host. As other organisms, viruses also experience environments or resources that are both heterogeneous and dynamic in space and time and thus act as source of selection. Meaning that some viruses are able to replicate only in particular kinds of cells, or tissue types, or single host species, whereas other viruses are able to infect many

different host species even coming from different genera and families (Gibbs *et al.* 1995). Poliovirus and rhinoviruses, for example, infect only humans in nature. Still, majority of both common and emerging pathogens of humans, animals and plants are multi-host pathogens (Cleaveland *et al.* 2001; Woolhouse *et al.* 2001). *Influenza A virus*, for example, causes little or no disease in its natural reservoir host, aquatic birds, but provoke severe respiratory tract infections in poultry, pigs, seals, whales, horses, and humans (Webster *et al.* 1992). Almost all plant viruses have a wide host range, infecting dozens or hundreds of host species, even those belonging to phylogenetically distant families. An extreme example of wide plant host range is *Tomato spotted wilt bunyavirus*, infecting over 600 different species from 70 families (Cann 2005). Still, Malpica *et al.* (2006) were the first to investigate natural virus-plant associations and showed that multi-host plant viruses are not really so, but instead, tend to associate to a particular host. In summary, viral host range is a poorly understood and historically dynamic process, yet comprising very important implications for human, animal and plant health because host-range expansion may lead to the emergence of the new infectious diseases.

Since viruses vary in their host range, they can be arbitrarily divided into: *a*) specialists, *i.e.*, those infecting only one or a few related host species and *b*) generalists, *i.e.*, those that show surprising versatility in the number of hosts where they are able to replicate (Elena *et al.* 2009).

The degree to which parasites adapt to a particular host depends on the balance between within-host selection and among-host gene flow (Lajeunesse & Forbes 2001; Dennehy *et al.* 2006). Encompassing the problem of specialization, Futuyma & Moreno (1988) assumed that the cost of adaptation should be reflected in a negative genetic correlation between alleles associated with the use of different resources. In other words, when a virus population optimizes its fitness in a constant environment (*i.e.*, single host evolution) by fixing adaptive mutations, these host-associated mutations will have negative impact on virus fitness in another host (Kassen 2002). Two mechanisms are responsible for fitness tradeoffs across hosts that provide an advantage to specialist over generalist viruses (reviewed in Elena *et al.* 2009): antagonistic pleiotropy and mutation accumulation. Firstly, antagonistic pleiotropy means that mutations that are beneficial in one host may be deleterious in another host (Duffy *et al.* 2006). Secondly, due to genetic drift, neutral mutations may accumulate in the genes that are not necessary in one host but are essential in another host (Kawecki 1994; Gandon 2004).

Remold *et al.* (2008) were first to investigate the relative contribution of antagonistic pleiotropy *versus* mutation accumulation to the evolution of host specialization in *Vesicular stomatitis rhabdovirus* (VSV).

On the other hand, simultaneous adaptation to multiple host types promotes the evolution of generalist viruses with no fitness tradeoff across hosts. A generalist virus carries mutations associated with different hosts, but never the optimal combination for any one environment: the ''jack of all trades is a master of none'' (Whitlock 1996). Accordingly, a specialist virus may outperform a generalist on a particular host, but would not outperform a generalist on the wider range of hosts. In the case of independent adaptation, a mutation fixed by adapting to one host is beneficial in the host where adaptation is done, but neutral in the other, so there are no fitness tradeoffs across different hosts.

Host homogeneity usually results in specialization (Duffy *et al.* 2007) and leads to virus local adaptation (Woolhouse *et al.* 2001). It is generally believed that interspecific competition promotes the evolution of specialization to particular host species, allowing parasites to avoid competition or become better competitors (Futuyma & Moreno 1988). Key consequences of host specialization are a reduction in genetic variability and fitness canalization into a local optimum that leads to limitation in gene flow among viral populations and results in the limited exploitation of other possible hosts. Thus, specialization is often considered to be an evolutionarily irreversible 'dead end', presuming a potential risk of population extinction, because selection within a single host should in turn compromise the ability of viral populations to adapt to changing environmental conditions (Futuyma & Moreno 1988; Woolhouse *et al.* 2001). In contrast to the aforementioned constraints, a recent work using comparative analyses of host specificity and competition suggests that generalist parasites may even have evolved from host specialists (Johnson *et al.* 2009).

The advantages of generalism, however, are not well understood. It has been suggested that evolution should favor specialists because there are tradeoffs that limit the fitness of generalists in any of the alternative hosts or because evolution proceeds faster with narrower niches (Fry 1996; Whitlock 1996; Kawecki 1998; Woolhouse *et al.* 2001). Still, generalism among viruses is rather an exception. Woelk & Holmes (2002) observed reduced positive selection in vector-borne RNA viruses and further suggested that vector-borne viruses can be thought of as evolutionary generalists, because of evolutionary trade-offs in the vector-host association. But experimental evolution

studies generally do not concur with this assumption of constraint on adaptation imposed by vector (Wallis *et al.* 2007; Deardorff *et al.* 2011). Indeed, simultaneous evolution in multiple host types may result in generalist virus that does not pay fitness cost in any alternative hosts, but instead, performs as well as the specialists (Novella *et al.* 1999; Weaver *et al.* 1999; Turner & Elena 2000; Cooper & Scott 2001; Bedhomme *et al.* 2012), whereby other studies provided evidences in favor of the tradeoff hypothesis and specificity of adaptation to a given host (*e.g.*, Crill *et al.* 2000; Greene *et al.* 2005). Experimental evidences for the cost of being a generalist are scarce (Coffey *et al.* 2008).

The extent of adaptation to a particular host is a function of the genetic variability of the pathogen (Woolhouse *et al.* 2001) and reversely; a variety of different hosts drives parasite diversity (Maclean 2005). The advantage of a high genetic diversity both in parasites and hosts is usually considered in the framework of the Red Queen hypothesis. During the long-term co-evolutionary history and arms-race between viruses and their hosts (Woolhouse *et al.* 2002), fitness trade-offs that occur are function of evolutionary rate (Whitlock 1996). However, both generalist and specialist viruses are known to have become established successfully in new hosts, suggesting that no generalization can be made about the likelihood of either type of virus infecting a previously resistant host to create a new epidemic pathogen (Woolhouse & Gowtage-Sequeria 2005; Parrish *et al.* 2008). To conclude, it has been suggested that antagonistic pleiotropy and fitness tradeoffs may be common in the small and compacted genomes of RNA viruses (Elena 2002; Elena *et al.* 2009). A practical consequence of the cost of host expansion an fitness tradeoffs is the long-standing design and use of attenuated vaccines based on the adaptation of a virus to new hosts, whereby the virus diminishes its virulence in humans and farm animals.

1.6.2. Genotype-by-environment (G×E) interactions

A phenotype is determined by its genotype together with the environment. Yet, it has been widely observed that no single genotype exhibits the same phenotype in different environments; due to a phenomenon termed genotype-by-environment (G×E) interaction (*e.g.*, Remold & Lenski 2001). G×E are commonly reported as mayor compromise in plant and animal breeding experiments, since breeders search for pure-line genotypes that are widely adapted across environments and optimized for high

yield and other desirable characteristics (Lynch & Walsh 1998). Still, the form and the extent of $G \times E$ for fitness is a fundamental issue in evolutionary ecology and pivotal for understanding the process of adaptation (Kondrashov & Houle 1994; Kawecki & Ebert 2004). In relation, $G \times E$ have been explored mainly in quantitative genetics in order to understand the maintenance of genetic variation in complex traits (Lande 1976; Gillespie & Turelli 1989; Fry *et al.* 1996) and the evolution of specialization and generalism (Wright 1977; Via & Lande 1985; Futuyma & Moreno 1988) and adaptation to heterogeneous environments (Via *et al.* 1995).

The goals of quantitative genetics are to partition total phenotypic variation into genetic (G) and environmental (E) components. This information (expressed in terms of variance components) allows predicting the contribution of each component to the total phenotypic response. Imagine a situation, for example, where one is interested to test the fitness (*i.e.*, the phenotype μ) of virus mutants, (*i.e.*, genotypes, G) across different hosts (host represents an environment for a virus; E). If several genotypes are each tested in several environments, the variance of the average fitness of genotypes represents genetic variance (G_i), and the variance of the average fitness in each environment represents environmental variance (E_j). With additivity of G and E, *i.e.*, no $G \times E$, the relative performance of genotypes is independent of the environment $\mu_{ij} = \mu + G_i + E_j + e_{ij}$, where e_{ij} represents the experimental error measure. The term $G \times E$ refers to of joint effects of genetic and environmental factors. In statistical language, $G \times E$ are considered as non-additive effects: the case when the whole is greater than the sum of the parts: $\mu_{ijk} = \mu + G_i + E_j + (G \times E)_{ij} + e_{ijk}$. So, variance due to deviation from linear relation between the genotype and the environment cannot be attributed neither to purely genetic sources, nor to purely environmental sources, but rather to the $G \times E$.

1.6.2.1. The nature and causes of *G*×*E*

There are two equally valid ways of interpreting $G \times E$. On the one hand, it expresses the extent to which genetic variation varies over different environments since the selection may be stronger in some environments than in others (*e.g.*, Via 1984b). So, part of this variation among environments in the quantity of genetic variation that is expressed in them arises because genotypes that have very high relative fitness in one environment are mediocre or even inferior in others. Thus, genotypes may be considered as specialists in particular environments.

On the other hand, $G \times E$ expresses the extent to which environmental variation is differentially expressed by several genotypes. Some genotypes may express very different phenotypes in different environments, and therefore possess a large quantity of environmental variance; others will be less responsive, and express more or less the same phenotype regardless of the environment in which they are raised (*e.g.*, Via 1984). Thus, the quantity of environmental variance, (sometimes referred to as 'plasticity'), is itself a character, and may vary among genotypes, which may be selected as generalists over a range of environments. The main issue raised by $G \times E$ is the balance between generalization and specialization that should evolve in populations that live in a heterogeneous environment.

Such interaction can arise from two not necessarily exclusive mechanisms: (Lynch & Walsh 1998) *i*) a change of scale, such that higher-ranking genotypes in one environment react more (or less) strongly to conditions in the second environment (Figure 8A), and a *ii*) change of ranking (Figure 8B). With multiple genotypes in multiple environments, many more patterns are possible.

The function that relates the mean phenotypic response of a genotype to a change in the environment is called reaction norm. If there is no $G \times E$ these reaction norms are parallel, implying that the same amount of genetic variance is expressed in each environment. So, each genotype performs the same in environment one as it does in environment two, and therefore the genotypes may be considered as generalists (Figure 8A). Recall that generalism occurs, by definition, if the relative performance of each genotype is the same in one environment as it is in the other. If a population exhibits $G \times E$, the slopes of the reaction norms are nonzero, indicating a change in genetic variance, which would be detected by a standard analysis of variance. In the case of the change in rank (Figure 8C) the response of genotypes to the environmental change is specific to a particular genotype so that the genotypes that perform better in one environment do worse in the other. Here, the genotypes can be considered to be specialists, some having an advantage in one environment, while others in the second environment. Therefore, the genotypic variance does not vary between two environments. A change in scale (Figure 9b) gives insight into which genotype is more sensitive to the change in the environment (case of G_2 in Figure 9b). Here, the rank order is preserved, but genotypic variance differs between the two environments.

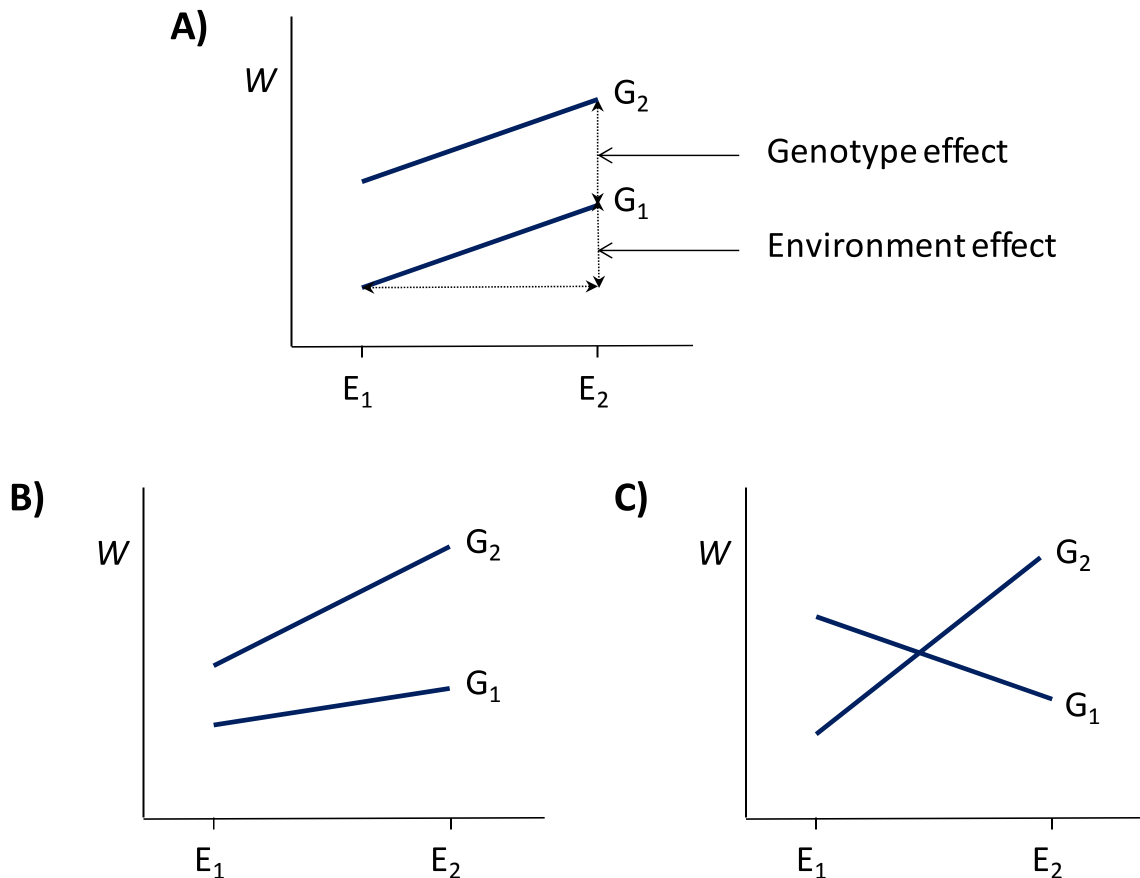


Figure 8. Reaction norms for two genotypes in response to two different environments. A) No *G×E*. Genotype effect (*G*) is due to average differences among genotypes, across environments. Environmental effect accounts for average differences among environments, across genotypes. B) *G×E* is due entirely to a change in scale. C) *G×E* is due to a change in ranking.

The analyses of variance allow assessing how much phenotypic variance is due to the differences in genotype averaged over all environments (*i.e.*, a change in rank or 'plasticity' or antagonistic pleiotropy) and how much in outcome of differences in genotype (averaged over all environments), corresponding to a change in scale. Such analysis from the deviation from the mean gives a result that depends upon the actual distribution of genotypes and environments in a particular population. With the same genetic composition of the population, a shift of the mean of the environmental distribution may reduce the genetic component of the phenotypic variance (Figure 9b). Thus, genetic variance can be drastically reduced or disappear by changing the environment. Conversely, a change in genotypic variance changes the environmental variance (Gupta & Lewontin 1982).

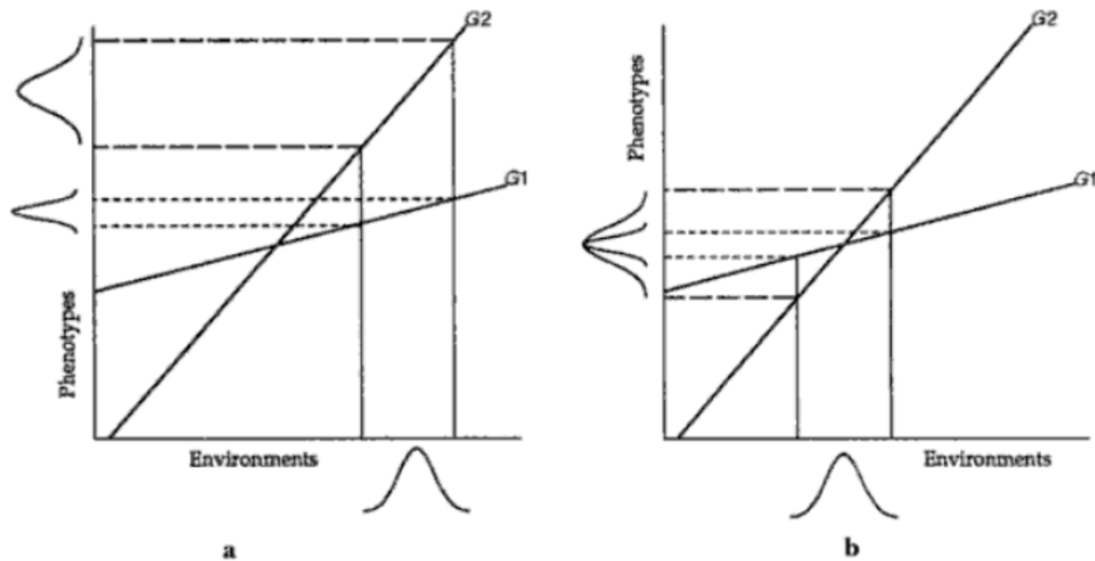


Figure 9. Illustration of how a change in environment affects the genetic variance of two genotypes. The lines G1 and G2 are the norms of reaction of two genotypes whose phenotypes had been ploted on the abscissa in relation to the environmental distribution plotted on the ordinate. The total population will be a mixture of these two phenotypic distributions in proportion to the frequency of the genotypes in the population. Since the means of the two phenotypic distributions are different in panel *a*, the analysis of variance will clearly show the genetic variance. If the environmental distribution shifts to right, as in panel *b*, the genotypes have not changed, but now there is no genetic variance of phenotype because the environmental distribution is centered where the true norms of reaction cross, so there is no average effect of genotype. Thus, genetic variance has been destroyed by changing the environment (adapted from Gupta & Lewontin 1982).

1.6.3. Source of variation: mutation

Mutation is the ultimate source of genetic variability that allows for the appearance of new viral genotypes. Thus, mutational supply rate directly limits adaptation. Still, majority of mutations are deleterious. In a population of large size, the deleterious mutations have a negligible chance of fixation, because purifying selection would eliminate them. Only mutants that have higher fitness compared to the wildtype will be fixed. The rare, favorable mutations must escape accidental loss trough genetic drift. Thus, adaptation requires a mutation to be favorable and that it escapes the stochastic loss (Elena & Sanjuán 2005). The importance of new beneficial alleles, however, depends on the environmental context. In a constant environment, stabilizing selection acts upon already obtained means (fitness optimum), so no new beneficial mutations are produced; the population already contains the appropriate genetic variation. If,

however, selection is commonly directional, as it would be, for example, when a virus faces a new host or antiviral treatment, then new mutations within virus population can be critical for adaptation (Whitlock 1996).

1.6.3.1. Mutation in RNA viruses

Even though the genome of RNA viruses show relatively low complexity in terms of the number of proteins encoded for, still it is a dynamic system from a population genetics' point of view. RNA virus populations contain enormous genetic diversity due to: *i*) mutation, *ii*) homologous and nonhomologous recombination and *iii*) reassortment in the case of viruses with segmented genomes (Rambault *et al.* 2004; Ghedin *et al.* 2005; Duffy *et al.* 2008;). With respect to the lines of investigation pursued in this thesis, only the effects of mutations will be discussed here in more detail.

Mutation rates of RNA viruses are orders of magnitude higher than those of their DNA-based hosts (Figure 10) and in the range of 0.03 – 2 per genome and replication round (Drake *et al.* 1998; Drake & Holland 1999; Chao *et al.* 2002; Sanjuán *et al.* 2010; Tromas & Elena 2010). Positive single-stranded RNA viruses posses the highest mutation rate among all viruses (Figure 11). Moreover, given approximately one mutation per genome per replication round, the lack of the genomic space reflected overlapping reading frames and cistrones encoding for multifunctional proteins (see section 2.1) puts serious constraints on RNA virus genomes (Holmes 2003) being most of the mutations deleterious. Spontaneous mutation rate of TEV was estimated to be within the range of 10^{-6} - 10^{-5} mutations per site and generation (Tromas & Elena 2010). Types of non-lethal mutations observed for plant RNA viruses are nucleotide substitutions or deletions (Malpica *et al.* 2002; Tromas & Elena 2010). Lack of redundancy together with compactness of RNA virus genomes might be reasons that do not allow for mutations of larger size.

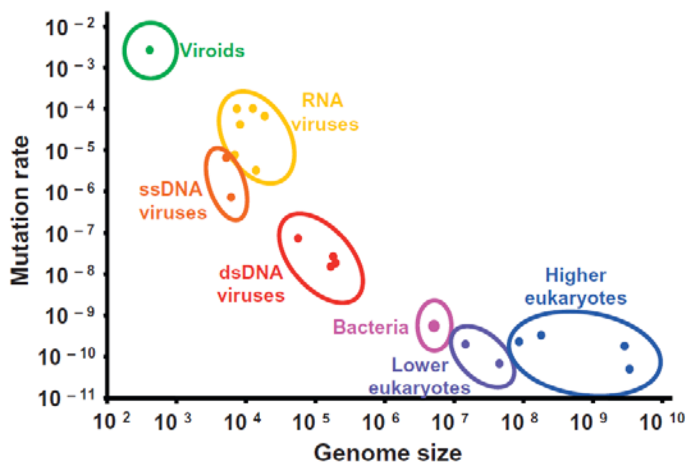


Figure 10. Relationship between mutation rate per nucleotide site and genome size for different genomic systems including viruses. (from Gago *et al.* 2009)

Such elevated mutation rate is attributable to the low replication accuracy of viral RNA-dependent RNA polymerases (RdRP) that lack proof-reading activity (Steinhauer *et al.* 1992); with the only exception being the *Nidovirales* that possess 3'→5' exoribonuclease activity (Gorbalenya *et al.* 2006; Minskaia *et al.* 2006). There is also no mismatch repair, even in double stranded RNA viruses; instead, there are some other mechanisms involved in genome maintenance and repair (Barr & Fearns 2010). Except for some satellite viruses, all RNA viruses encode the RdRp, because host cells possess no enzymes that are capable of replicating long RNA molecules. It is usually thought that low-fidelity of RdRP is an adaptive trait, resulting from a tradeoff with replication speed (Elena & Sanjuán 2005; Furió *et al.* 2005; Belshaw *et al.* 2008). Additionally, some other factors contributing to increased mutation rate have been reported. One of such is the presence of RNA secondary structures (see section 1.2.1.2.) and host or environmental factors (Schneider & Roossinck 2001; Pita *et al.* 2007).

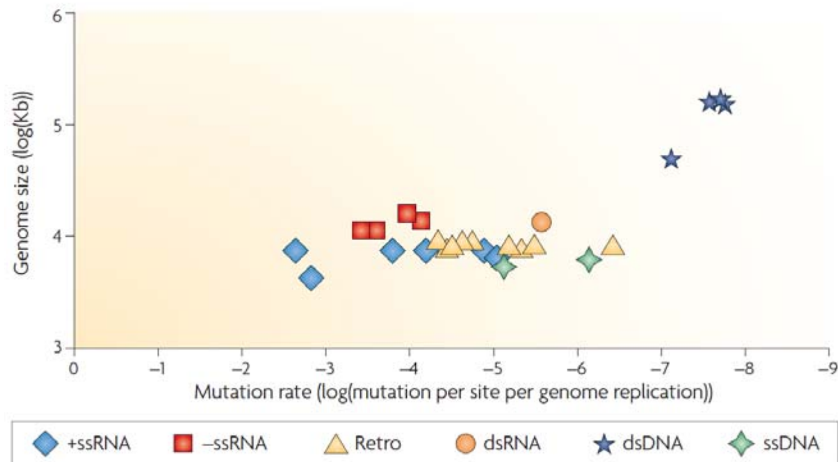


Figure 11. Average mutation rates for different types of viruses (from Duffy *et al.* 2008). Viruses are denoted by different symbols with respect to their genome type, *e.g*. +ssRNA corresponds to positive-sense single-stranded RNA viruses.

1.6.3.2. Mutational fitness effects

Testing theoretical predictions about the genetics of adaptation requires measuring fitness effects of mutations related to the underlying evolutionary dynamics to which a population is exposed (Loewe & Hill 2010) such as *i*) selection, *ii*) population size (or the effects of genetic drift), *iii*) population structure (or demography; *i.e*., migration), *iv*) genetic interactions; such as epistasis, linkage, clonal interference (Gerrish & Lenski 1998) and *v*) environmental interactions.

Most mutations have deleterious impact on organism's fitness (Fisher 1930). The pervasiveness of spontaneous accumulation of deleterious mutations has been demonstrated in mutation-accumulation experiments under minimal effect of purifying selection and strong bottlenecks (Mukai 1964; Chao 1990; Duarte *et al*. 1992; Escarmís *et al.* 1996; Elena & Moya 1999; Yuste *et al.* 1999; De la Peña *et al.* 2000) that onset Muller's ratchet (Mueller 1964). Studies using direct mutagenesis approach confirmed the prevalence of deleterious mutations and additionally revealed the large proportion of lethal mutations (Sanjuán *et al.* 2004; Carrasco *et al.* 2007; Domingo-Calap *et al.* 2009).

Kimura (1968) was the first to recognize the importance of neutral mutations for evolution. Since neutral mutations have no effect on fitness (or their effect in the experimental assay is too small (Burch *et al.* 2007), natural selection does not operate on them, so they can be fixed in a population trough the action of random genetic drift (Kimura 1983; Ohta 1992). Still, Kondrashov & Houle (1994) experimentally

demonstrated that conditionally neutral mutations may be common during adaptation, meaning that the selection might be operative on particular mutations at a given circumstances (Eyre-Walker & Keightley 2007). Even though it is often thought that synonymous sites resemble neutral evolution or are under weak selection, in many instances, it has been demonstrated that synonymous substitutions are not neutral (Plotkin & Kudla 2011) and as such may have a major effect on RNA virus fitness (Carrasco *et al.* 2007; Coleman *et al*. 2008; Marsh *et al.* 2008; Cuevas *et al.* 2011).

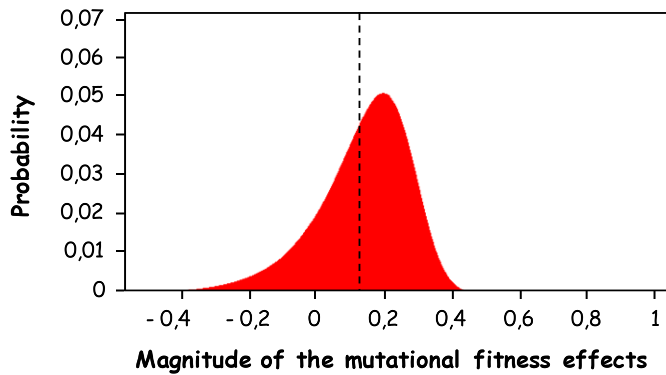
Beneficial mutations are rarely observed, therefore are also thought to be extremely rare. Hence, in a population, a newly arisen beneficial mutation will have a small probability of being fixed in a finite population only by means of natural selection (Patwa & Wahl 2008). It has been argued that the existence of beneficial mutations of small effect rather than large is more probable (Fisher 1930; Orr 1998). Increases in population size and mutation rate can cause larger-effect beneficial mutations to become fixed (Orr 2000). Sniegowski & Gerrish (2010) recently described a deterministic model using data from experimental evolution studies in microbes, which suggests that beneficial mutations may actually become abundant under periodic selection in combination to high mutation rate. In such case, genotypes comprising multiple beneficial mutations may become prevalent in the population.

In spite of high mutation rates, plant RNA virus populations in nature resemble significant genetic stability over time as shown by the analyses of population genetic diversity (*i.e*., nucleotide diversity per site) between different isolates (García-Arenal *et al.* 2001). That observation can be attributed to strong purifying (negative) selection that purges deleterious mutations from virus genomes (Domingo & Holland 1997). More concretely, virus populations show relative stasis in natural host, but evolve rapidly in a new host (Novella *et al.* 1999). This assumes that virus population within its reservoir host has reached the global fitness peak characterized by strong purifying selection that operates within host and eliminates unfit mutants in a stable environment (single-host evolution).

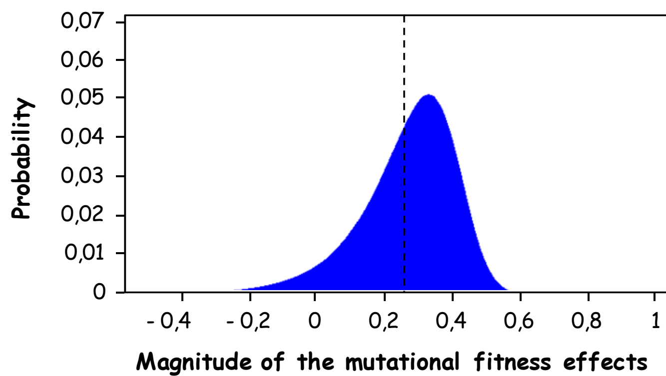
But, what is the fate of virus genetic variation under different selection pressures presented by different hosts? Genetic variation within a viral population can be simulated by a well designed and controlled experiment using a collection of virus genotypes. The analysis of the distribution of genetic variation for fitness of a viral population in its natural or "local" environment (*i.e.*, host), for which the virus is adapted to compared to the "foreign" environments should provide an answer to the

above stated question. If virus population contains mutations that are beneficial in the alternative host(s), an increase in their frequency must ocurr for a virus to adapt to its new host. The fraction of beneficial mutations within a standing genetic variation may be increased in stressfull environments (*i.e.*, new hosts within which a virus population experiences low fitness) by two mutually non-exclusive means. It can be accomplished either by a shift in mean of the distribution of mutational fitness effects (DMFE) towards more positive values while keeping the shape constant (Agrawal & Whitlock 2010), or alternatively, by increasing the variance without affecting the mean of the DMFE (Figure 12; Martin & Lenormand 2006). Additionally, the DMFE can be addressed by testing how the qualitative effect of mutations on viral fitness (*e.g.*, lethal, deleterious, neutral and beneficial) changes from the reservoir to potential new hosts. A mutation beneficial in one host may not be so in an alternative one, so this antagonistic pleiotropy among mutations may change the rank order of mutations across hosts. Moreover, while still retaining the rank order of fitness effects, $G \times E$ can be generated by altering the genetic component of phenotypic variance (Remold & Lenski 2001). Till now, DMFE have been characterized for a handful of viruses in their reservoir hosts (reviewed in Sanjuán 2010), but whether and how these distributions change across potential hosts has never been experimentally addressed.

To conclude, selection and counterselection of high mutation rates depends on many factors: the number of mutations required for adaptation, the population size, competition with other viruses, transmission and temporal and spatial environmental heterogeneity (Steinhauer & Holland 1987).

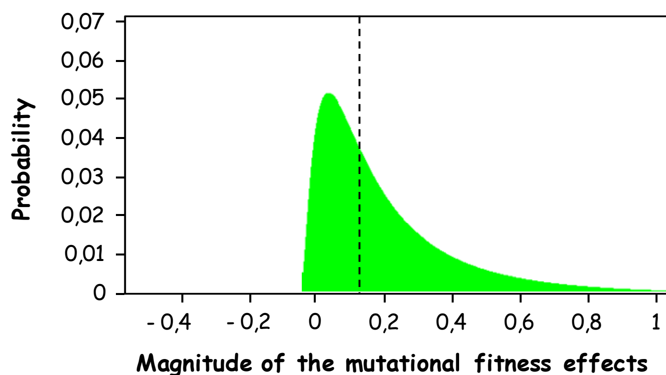


a) **Reservoir host**



b) **New host**

Change in mean, no change in shape



c) **New host**

No change in mean, change in shape

Figure 12. Two possible effects of host switching on the distribution of mutational effects on viral fitness. Panel a) represents the DMFE in the reservoir host. Panel b) represents a change in the mean mutational effect effect (vertical dashed line) but retaining the same shape. Panel c) represents a change in the shape without affecting the mean. The area under the curve to the right of the dashed line corresponds to the fraction of possible beneficial mutations in the new host (*i.e.*, host-range mutants).

1.6.4. Epistasis

Over a century ago, it was demonstrated that genes do not act independently of all others (Bateson 1909). Since then, indeed, the existence of numerous interactions among genes or mutations in determining phenotypes, *i.e.*, epistasis, has vastly been demonstrated (Phillips 2008). Analogously to $G \times E$ interactions, where the expression

of a phenotype of different genotypes under different environmental conditions can be viewed as a reaction norm curves, the interaction between different loci within a single genotype can be viewed as the reaction norms of different genotypes at one locus plotted against the different genotypes at the second locus (Wolf *et al.* 2000). In this sense, epistasis can be referred to as genotypes genotype-by-genotype or *G*×*G* interactions.

Herewith, it is important to distinguish between epistasis and genetic complementation. The main difference is that genetic complementation occurs between two physically separated (unlinked) genes or mutations. For instance, complementation occurs between virus particles within the single host cell wherein one is defective for a particular function or a protein product and this function is provided *in trans* by a virion carrying functional allele or protein.

RNA viruses are characterized by their small genomes containing secondary structures that encode for multifunctional proteins, hence single mutations are likely to have multiple and antagonistic effects, such as compensatory mutations (Sanjuán *et al.* 2005). Moreover, genome compactness and the existence of overlapping reading frames may impose limitations on the number of available adaptive solutions, which in turn may explain the commonness of convergent evolution RNA viruses (Cuevas *et al.* 2002; Agudelo-Romero *et al.* 2008; Remold *et al.* 2008).

In the first instance, it is important to differentiate between biological and statistical meaning of epistasis. In biological terms, epistasis occurs when the fitness differences of single mutations depend on the presence or absence of other mutations. In statistical genetics, epistasis is the deviation (or interaction) of fitness value of a genotype carrying multiple mutations from the additive combination of single mutations (Falconer 1989). A detailed description of epistasis can be found in the introduction part of Chapter IV of this thesis; here epistasis is elaborated in a more simple and general way.

In the absence of epistasis, the effects of mutations combine in a linear way; so by knowing the effect (*i.e.*, fitness value) of each mutation separately, the expected fitness value of genotype carrying both mutations can be easily calculated (Figure 13) using multiplicative or additive epistasis model. For some traits, such as fertility, an additive scale is thought to be more natural, whereas for other traits, such as fitness or mortality, the multiplicative approach is more appropriate (Phillips 2008).

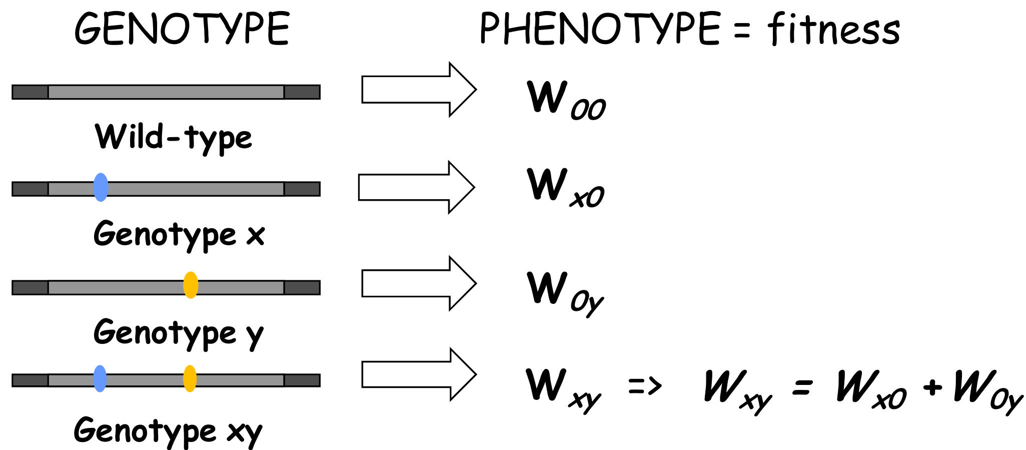
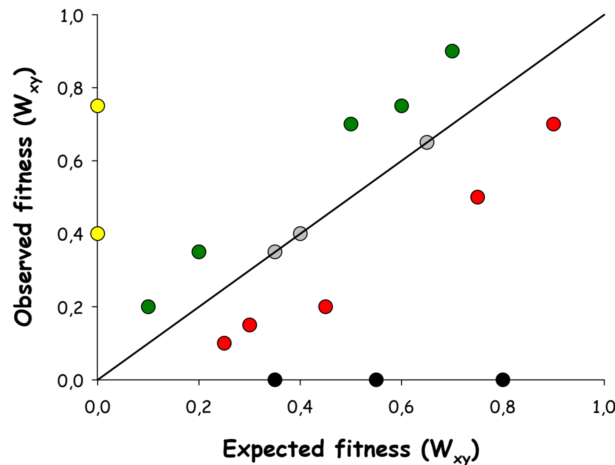


Figure 13. An example of multiplicative genetic effects: the absence of epistasis. Grey bars represent a genome of a virus. The fitness value of the wildtype virus is represented by W_{00}. Mutation x or y (blue and yellow point, respectively) may change the virus fitness: W_{x0} and W_{0y}, respectively. Under multiplicative mutational effects the fitness of the double mutant xy equals the product of the effects of individual mutations.

Alternatively, if the observed fitness of a double mutant deviates from the sum or the product (depending on the epistasis model used) of individual mutations, there is an interaction between the two mutations, namely $\varepsilon_{xy} = W_{00}W_{xy} - W_{x0}W_{0y}$ (in the multiplicative scale). This ε_{xy} parameter is, therefore, magnitude epistasis (Figure 14; Kouyos *et al.* 2007). Negative epistasis for fitness on a multiplicative scale is known as synergistic epistasis, meaning that combination of two deleterious mutations result in fitness of double mutant that is lower than expected by multiplying the fitness of single mutations. Positive or synergistic epistasis for fitness on a multiplicative scale results in larger fitness of a double mutant than expected under the null hypothesis of multiplicative fitness effects.

Epistasis may be studied by another approach: by its sign. Sign is referred to qualitative effect of mutations: a mutation can be beneficial in one genetic background and deleterious in another (and *vice versa*). In other words, the sign of mutational fitness effect is conditioned upon the genetic background (Weinreich *et al.* 2005). Reciprocal sign epistasis occurs when two mutations are deleterious *per se* but together augment the fitness of the double mutant (Poelwijk *et al.* 2007).

The deleterious mutational impact on fitness can be buffered by positive or antagonistic epistasis (Bonhoeffer *et el.* 2004; Burch & Chao 2004; Sanjuán *et al.* 2004a; Desai *et al.* 2007) where two deleterious mutations jointly result in milder fitness effect.



H_0: Expected: $E(W_{xy}) = W_x \times W_y$ -> NO EPISTASIS ○

H_A: Observed: $O(W_{xy}) = E(W_{xy}) + \varepsilon_{xy}$ -> ε_{xy} = **EPISTASIS**

$O(W_{xy}) > E(W_{xy})$ => $\varepsilon_{xy} > 0$ **positive** (antagonistic) ●

$W_x \mid W_y = 0$ -> compensatory viability ○

$O(W_{xy}) < E(W_{xy})$ => $\varepsilon_{xy} < 0$ **negative** (synergistic) ●

$W_{xy} = 0$ -> synthetic lethality ●

Figure 14. The relationshiop between observed and expected fitness of a double mutant picturing types of magnitude epistasis. Grey circles correspond to the absence of epistasis. Positive or negative epistasis implies that the fitness of the double mutant is higher or lower than expected (green or red circles), respectively, *i.e.*, corresponds to magnitude epistasis. Extreme forms of magnitude epistasis are illustrated by black circles ('synthetic lethality') and yellow circles ('compensatory viability') (adapted from Kouyos *et al.* 2007).

1.6.5. Robustness

Robustness and adaptability are two counterpoised properties of living organisms. The ability to preserve a constant phenotype in spite of genetic or environmental perturbations (De Visser *et al.* 2003) is referred to as mutational (or genetic) or environmental robustness, respectively. On the other hand, adaptability or evolvability is an organism's capacity to generate heritable phenotypic variation in order to adapt in response to selection (Wagner & Altenberg 1996; Kirschner & Gerhart 1998). Alternatively, adaptability may be defined as the ability of an organism to acquire novel functions (Burch & Chao 2000; De Visser *et al.* 2003). Thus, a prerequisite for evolvability is the existence of genetic variability. Genetic diversity in RNA virus populations is maintained due to the counteraction of mutation accumulation *vs.*

selective forces acting on the virus population. Under appropriate conditions (*e.g.*, selection coming from host resistance genes or antiviral treatments) RNA viruses are capable of undergoing rapid evolution (Steinhauer & Holland 1987). How do both robustness and evolvability coexist and evolve is an open debate, especially in the case of virus populations which need to be robust enough in order to cope with their own high mutation rates, as well as evolvable enough to adapt to ever fluctuating environments imposed by their hosts or antiviral treatments by creating novel phenotypes.

Robustness may be examined at many different scales of biological organization; from biochemical level to populations and ecosystems. The hallmarks of robustness are redundancy of component parts that compensate for functionality and the existence of negative feedbacks that actuate by decreasing the magnitude of perturbation (Lenski *et al.* 2006). In concrete, two genes are redundant if each can partially or fully substitute for the function of the other (Thomas 1993). Biological systems have evolved mechanisms that insure their integrity by purging or buffering the damage. In concrete, examples of robustness are: *i*) the existence of alternative metabolic pathways; or, *ii*) the presence of multiple genes or alleles that contribute to the same function; or, *iii*) biochemical buffering mechanisms trough regulation of gene expression or chaperon proteins that compensate for or alleviate the mutational effects; or *iv*) the existence of proofreading and repair mechanisms that purge deleterious mutations from the genome (Elena *et al.* 2006; Lenski *et al.* 2006). So, redundancy is prevalent among eukaryotes. Effective genetic redundancy, where several paralogous genes or alleles contribute in performing the same single function, gives a space for the accumulation of deleterious mutations in any of them since the selection operates on phenotype that is unaffected because the deleterious mutational effect (or loss of function of a gene) is masked by the functional allele (Figure 15).

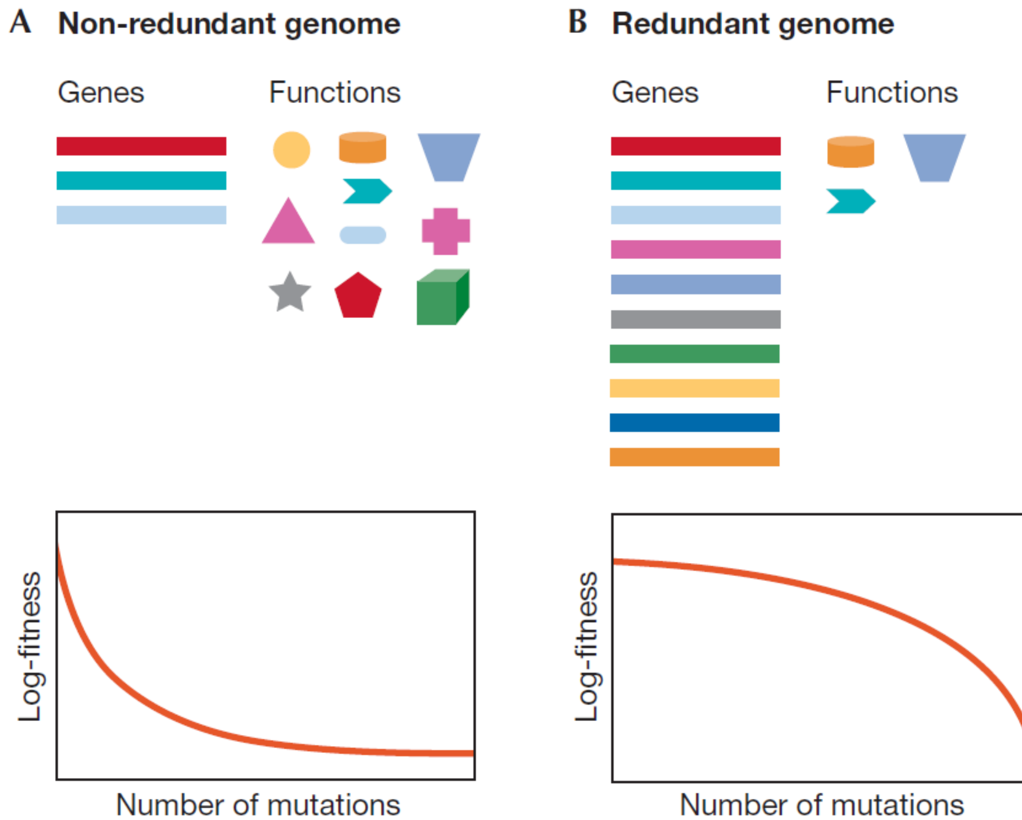


Figure 15. (**A**) Non-robust genomes should have fewer genes than functions (upper panel) so the first mutation should lead to a strong decline in fitness. Subsequent mutation accumulation does not lead to a significant fitness decline resulting in positive epistasis (lower panel). (**B**) Robust genomes should have more genes than functions (upper panel) so initial mutations have mild fitness effects. Increasing number of mutations leads to strong decrease fitness giving rise to negative epistasis (lower panel; from Elena *et al.* 2006).

Conferring two redundant genes with no selective advantage associated with having both genes, in the long term, should be an unstable condition (Thomas 1993). Since, by time, each gene would inevitably suffer a mutation and become non-functional. In the absence of selection for both genes or due to genetic drift, a non-functional allele would probably become fixed in the population by chance and the redundancy would be lost. Thus, if not essential, how are redundant genes maintained by the natural selection? Nowak *et al* (1997) modeled such a scenario by using different selection pressures that act on redundant genes and found that redundancy is an evolutionary stable trait. The proposed, inverse relation between genetic redundancy and the impact of deleterious mutations (Nowak *et al.* 1997) is also seen in dominance of antagonistic epistasis among pairs of deleterious mutations. Furthermore, Krakauer

& Plotkin (2009) used a mathematical model in order to investigate the impact of genomic redundancy on mean fitness as a function of the effective population size, the mutation rate, and the size of the genome. They found that redundancy is more common among phenotypes of higher organisms that experience low mutation rates and small population sizes, and reversely; redundancy is less common among organisms with high mutation rates and large populations, such as bacteria or viruses.

As for their genome architecture, RNA viruses are thought to comprise low mutational robustness; their compact genomes comprising overlapping reading frames clearly indicate the lack genetic redundancy and modularity (Elena *et al.* 2006; Belshaw *et al.* 2008). Moreover, since RNA viruses lack proofreading and repair mechanisms, mutations in their genomes have strong fitness effects, most of which being deleterious (Sanjuán 2010).

Selection should promote genetic robustness when mutation rates are high, as in RNA viruses (De Visser *et al.* 2003). This has been vastly demonstrated by theoretical works (Wilke & Adami 2003; Wagner 2005; Félix & Wagner 2008; Draghi *et al.* 2010). A study carried out in digital organisms (*i.e.*, self-replicating computer programs that can evolve) showed that elevated mutation rates can select for the increase in genetic robustness in order to tolerate the accumulated mutations, even at the expense of population's fitness (Wilke *et al.* 2001). Thus, robust digital organisms residing on lower but flatter (*i.e.*, low-fitness) regions of the fitness landscape should outcompete the non-robust populations located at a higher but narrower fitness peak in high mutation rate environments (Figure 16). This phenomenon of mutation-induced movement across the landscape whereby robust genotypes produce equally fit phenotypes while non-robust genotypes suffer from mutational fitness effects is known as "survival of the flattest" in contrary to Darwin's "survival of the fittest" paradigm. The "survival of the flattest" was confirmed experimentally for viroids (Codoñer *et al.* 2006) and VSV (Sanjuán *et al.* 2007).

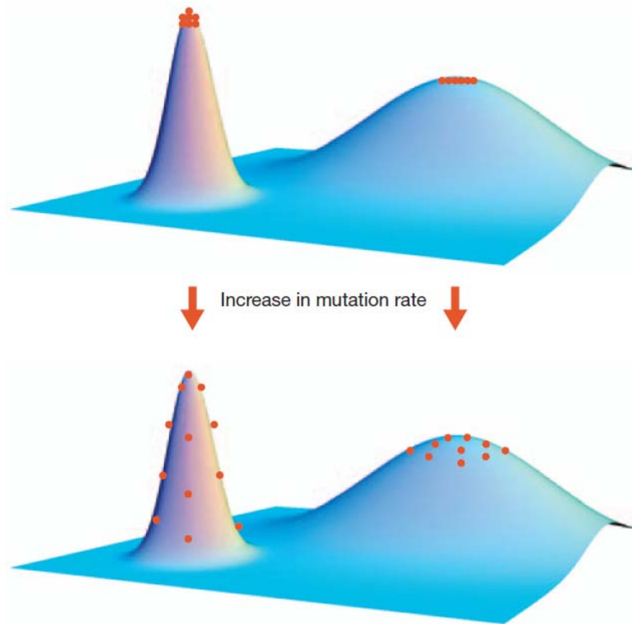


Figure 16. Schematic representation of a landscape characterized by a peak of high fitness but low robustness and another one of low fitness but high neutrality (that is, robustness). At a high mutation rate, populations at the high peak are pushed down, whereas those at the low peak remain unchanged (from Elena *et al.* 2006).

Even though viruses, at the individual level comprise low mutational robustness, at the level of the whole population, mutational robustness may be accomplished by complementation where functionality is provided by other virus particle(s). Different multiplicity of infection (*MOI*) are thought to be associated with differences in opportunity for genetic complementation (*i.e.*, fewer opportunities when the *MOI* is low), that in final, should influence the population's ability to retain fitness in the presence of deleterious mutations fixed through genetic bottlenecking. More robust viruses are expected to evolve under low *MOI* and have lower fitness and smaller average change in fitness (*i.e.*, lower variance) than the brittle ones. Under high *MOI*, a single virion can utilize protein products from other virus particles, so the inherent mutational robustness of a population subjected to high *MOI* would not be manifested. The first study that demonstrated a measurable difference in robustness between two different populations was in the RNA bacteriophage Φ6 (Montville *et al.* 2005). The authors have confirmed the aforementioned predictions by showing the decreased mutational robustness for virus lineages evolved at high *MOI*, where co-infecting particles provided the functional redundancy. A complementing study conducted *in silico* had investigated the interactions between complementation and mutation,

selection and epistasis (Gao & Feldman 2009). They found that strong complementation slightly reduces the fitness of a virus population but substantially enhances its diversity and robustness, especially if epistasis among deleterious mutations is antagonistic.

Variable environments require robust organisms. As noted before, RNA viruses, especially those that are generalists or vector-borne, experience wide variety of host types, so it is to be expected that they are environmentally robust. Still, the empirical evidences for the environmental robustness are scarce. Turner *et al.* (2010) explored how environmental robustness influences evolvability measured as a successful host shift using VSV. They showed that environmentally robust viruses (*i.e.*, generalists) have higher fitness and and lower variance for fitness than the brittle specialists across the novel hosts. This observation has implications of linking environmental robustness to the likelihood of viral emergence.

Finally, the relationship between robustness and adaptability remains unclear. Mutational robustness can either impede or facilitate adaptation, depending on the population size, the mutation rate, the time frame of observation and the structure of the fitness landscape (Elena & Sanjuán 2008; Wagner 2008; Draghi *et al.* 2010). A single empirical study provided the evidence that genetic robustness increases evolvability of thermotolerance in RNA virus Φ6 (McBride *et al.* 2008). Altogether, these evidences suggest that RNA virus populations, but not single individuals, comprise inherent environmental and mutational robustness.

1.6.6. Adaptive fitness landscapes

> *It is almost impossible with any brevity to exemplify the notion of adaptation. Just because adaptation consists, even in the simplest cases, in a multiplicity of correspondences between one sufficiently complicated system, the organism itself, and another equally complicated, the environment in which it finds itself. It is, indeed, just this multiplicity that makes the thing recognizably adaptive.* (R. A. Fisher 1934)

Adaptation in constantly changing environments is a complex process influenced by a large number of genetical, environmental, developmental and other factors, thus, is a central problem of evolutionary theory. Visualization of the process of evolution as

the movement of populations on landscapes consisting of adaptive valleys and peaks was proposed by Wright (1932). The stepwise adaptation of organisms to new environments by natural selection results from the intricate relationship between genotype and phenotype and between phenotype and fitness that is denominated as fitness (or adaptive) landscape (Wright 1932). The landscape is an *N*-dimensional plot that describes the fitness of all possible genotypes composing the so-called genotypic space. In a two-dimensional projection, the vertical axis, the height of a peak, indicates the fitness of each genotype, which is defined by the horizontal axis (Figure 17). Thus different peaks represent alternative solutions to the problem of survival (Gavrilets 2010). Spatially neighboring genotypes mutually differ by one mutational change, so that fitness associated to a genotype sequence yields a fitness surface. Genotypes that have high fitness occupy landscape peaks, whereas unfit genotypes occupy valleys. To infer a fitness landscape empirically means to (re)construct all possible genotypic intermediates that led to adaptation to a new environment and measure their fitness. Genotypes should bear all possible combinations of mutations fixed by the adapted genotype. If the order of the appearance of the n mutations during evolution is unknown, there are $n!$ mutational trajectories in which the adaptation might have occurred. Still, they all may not be equally accessible to natural selection (Weinreich *et al.* 2006). The number of combinations of n distinct genotypes taken per class k, can be calculated as the combinatory number $\binom{n}{k}$.

Evolution in a constant environment and by natural selection alone is always a hill-climbing process where a population is pushed towards the top of the nearest peak. In this framework, evolution occurs trough a continuous network of intermediate genotypes bearing each of the mutational steps that lead to increase in fitness in a particular environment, without passing trough low fitness valleys. Low fitness valleys formed by intermediate genotypes compromise evolution that always proceeds uphill. Hence, selection prevents genotypes from descending from the adaptive peak, moving across the valley of low fitness and climbing up to a new peak, even if the new peak has a higher fitness. In this scenario, the adaptive landscape is smooth and single peaked (Figure 17B). A change in the environment changes the fitness surface and shifts the peak away from the wild type. In a haploid system, as is our focus here, mutations on the wildtype background allow the population to explore the sequence space and thereby climb a fitness peak.

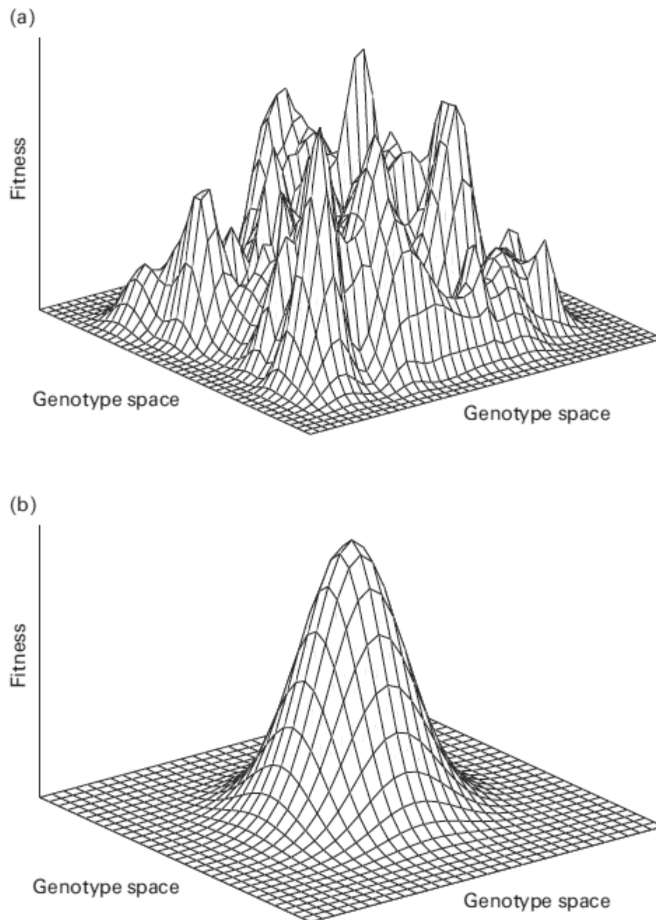


Figure 17. Two-dimensional representation of a population's fitness in a particular environment as a function of its genetic composition. Two types of classical fitness landscapes are presented: (a) a rugged landscape and (b) a single-peaked landscape (from Gavrilets 2010)

However, the probability that a random mutation will be beneficial is very small. Fisher (1930) in his geometrical model of adaptation predicted that the evolution occurs in small steps, *i.e.* the accumulation of many mutations of small beneficial effects. The smaller the mutational phenotypic effect, the higher will be the chance that the mutation is beneficial. Later on, Kimura (1983) showed that mutations, apart from being beneficial, must escape accidental loss when rare and mutations of larger effect are more likely to escape such loss. In other words, mutations of intermediate size are the most likely to contribute to adaptation (Kimura 1983). Using Fisher's geometric model of adaptation, Orr (1998, 1999) showed that adaptation is characterized by exponentially distributed phenotypic effects of mutations. Apart from mutation and selection, the fitness of a population, and consequently its underlying fitness landscape, can be influenced by the actions of genetic drift, migration, epistasis, etc., therefore giving rise to the landscape with multiple adaptive peaks (Figure 17A). The existence of the multiple peaks gives rise to a peak shift problem: how does the population move

between peaks? In small populations, genetic drift plays a more prominent role in comparison to selection, so a deleterious mutation can be fixed if followed by a compensatory mutation that restores fitness (Kimura 1985; Burch & Chao 1999). In this manner, genetic drift or periods of small population size allows a population to traverse fitness valleys and move from one peak to another (Wright 1931; 1932). Experimental evolution studies with bacteria (Lenski *et al.* 1991; Lenski & Travisano 1994; Elena *et al.* 1996a) and viruses (Novella *et al.* 1995, 1996, 1999; Elena *et al.* 1996, 1998; Burch & Chao 1999; Miralles *et al.* 1999, 2000; Wichman *et al.* 1999) have shown that the probability of fixation of beneficial mutations depends upon both the population size and mutation rate, thus confirming Fisher's geometrical model of adaptation. The increase in mutation rate or in the population size results in larger fitness increases in the initial steps of adaptation; however, the rate of adaptation decreases with time. On the other hand, if the mutation rate is small, like in complex organisms, the theory predicts slower increases in fitness during adaptation (Orr 2000).

Wright (1932) was the first to theoretically reason that the genetic effects on fitness are not properties of genes but depend upon genetic background. Indeed, epistasis determines the architecture of adaptive landscapes as well as the accessibility of adaptive pathways throughout the landscape, thus, is central to understanding the course of evolution (Figure 18). In absence of epistasis or in the case of magnitude epistasis, mutations give rise to either zero, positive or a negative fitness effect, regardless of the genetic background. This results in adaptive landscapes that are smooth and single peaked. As already mentioned, in a smooth fitness landscape the evolution will always proceed uphill towards the single global optimum. The curvature of the fitness trajectory is informative about the sign and strength of epistasis in the fitness landscape. In the presence of sign epistasis, the sign of the fitness effect of a mutation depends on the genetic background, such that only a fraction of the total paths to the optimum are selectively accessible, *i.e.*, contain only steps that confer a performance increase. Reciprocal sign epistasis is a particular case of sign epsitasis in which two mutations are individually deleterious but jointly advantageous. Both, sign epistasis and especially the reciprocal sign epistasis give rise to rugged landscape with multiple local optima (*i.e.*, peaks; Figure 17A). The ruggedness of adaptive landscapes is critical to predict whether the evolving populations may reach the global optima or, by contrast, trough alternative evolutionary pathways, may get stuck into suboptimal fitness peaks (Whitlock *et al.* 1995; Weinreich 2005; Poelwijk *et al.* 2011; Kvitek &

Sherlock 2011; Poelwijk *et al.* 2007). Rugged landscapes are also known as uncorrelated, meaning that adaptive mutations are independently drawn form an unknown (usually assumed exponential) function (Orr 2002; Rokyta *et al.* 2006). In other words, there is no correlation in fitness between similar genotypes (*i.e.*, those sharing sequence similarities).

For multihost parasites, such as RNA viruses, different adaptive landscapes are predicted because each host imposes different genetic constraints to the virus that result in fitness differences among hosts (*i.e.*, fitness trade-offs or antagonistic pleiotropy). Generalist viruses must have evolved regions where fitness peaks on different hosts coincide or overlap sufficiently to avoid demoting and movement trough low-fitness valley during the switch between different hosts. The same notion could be applied for the emerging viruses. If the hosts' landscapes are independent and fitness peaks are rare relative to low-fitness valleys (*i.e.*, most mutations are deleterious), overlapping peaks will be rare relative to peaks for either individual host. Therefore, the transition from one overlapping peak to another will involve movement trough even wider low-fitness valleys than it would be required for single-host landscapes. Thus, reaching coinciding fitness peaks in multiple hosts is less likely than reaching a single-host single high-fitness peak.

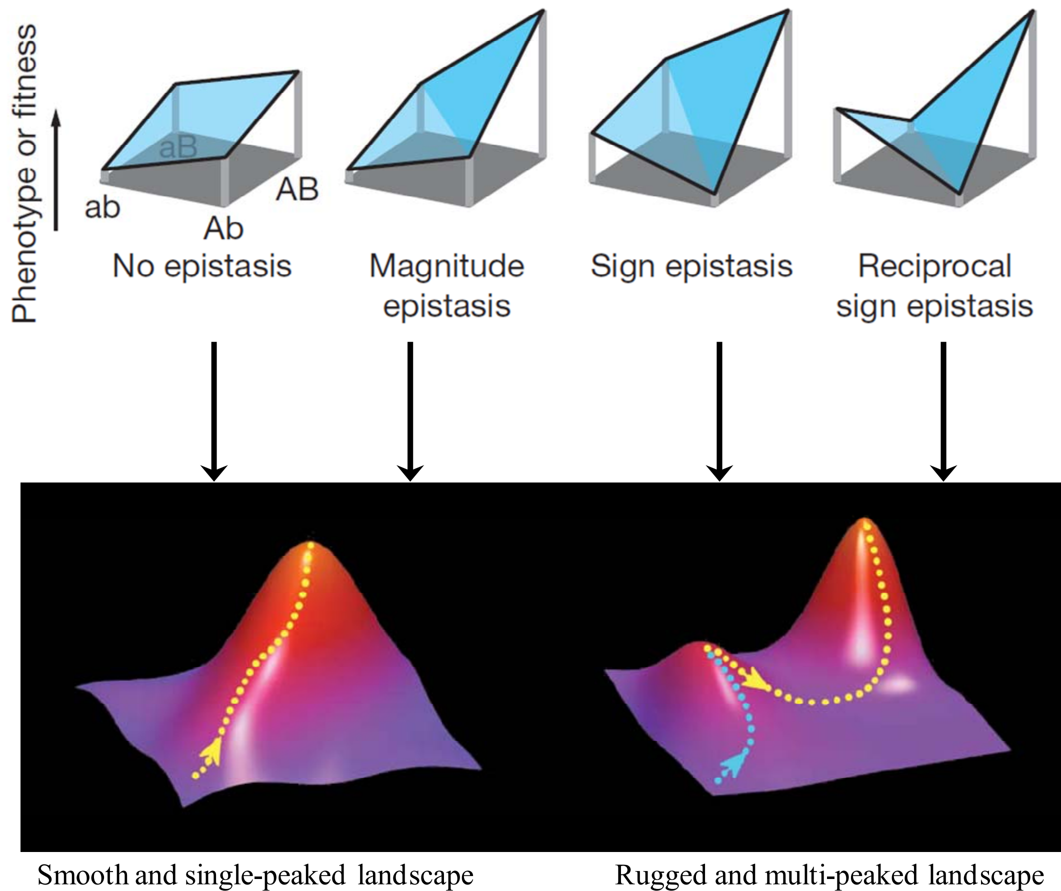


Figure 18. Illustrative relation between forms of epistasis and the topography of fitness landscapes (adapted from Poelwijk *et al.,* 2007). Consider an evolutionary transition of a haploid population of individuals with ancestral genotype *ab* to the adapted genotype *AB*. In the absence of epistasis or in the case of magnitude epistasis, the intermediate genotypes: *Ab* and *aB* have fitnesses somewhere between *ab* and *AB*, so the evolutionary transition of a population from the ancestral genotype *ab* to the adapted genotype *AB* is favored by natural selection because both of the intermediate genotypes (*Ab* and *aB*) have higher fitnesses than the ancestral genotype. This reflects in fitness landscapes being smooth and single-peaked. In a smooth fitness landscape evolution will always proceed uphill towards the single global optimum. Evolutionary constraints happen in the cases of sign and reciprocal sign epistasis when either or both of the intermediate (*ab* or *aB* genotypes) have lower fitness(es) compared to the ancestral *ab* genotype, respectably. such as in cases of sign epistasis. In this case(s), the intermediate genotype(s) of lower fitness(es) than the ancestral will be selected against thus giving rise to rugged fitness landscape.

CHAPTER II - OBJECTIVES

This Thesis explores the multidimensionality of genetic and environmental interactions using sets of mutant genotypes of a plant positive-sense single-stranded RNA virus, TEV, and measuring their fitness effects across different hosts and/or different genetic backgrounds. In concrete, the objectives were following and as such, correspond to different Chapters:

III. To generate 20 random, single-nucleotide substitution mutants of TEV and characterize the distribution of mutational fitness effects across a panel of eight different hosts, varying in their genetic relatedness to the natural one *Nicotiana tabacum*. This study will inform on the importance of $G \times E$ interactions in determining TEV fitness.

IV. To generate 53 double mutants of TEV, by randomly combining pairs of mutations used in Chapter III, and further estimate their mutational fitness effects in the natural host *N. tabacum* to characterize the patterns of epistasis (or $G \times G$) for TEV genome.

V. To randomly select ten double mutants of TEV from Chapter IV and characterize the epistasis across four different hosts, varying in genetic relatedness to the natural one. This study will provide light on the possible dependence of epistasis on environmental variation (*i.e.*, $G \times G \times E$)

VI. To construct all combinations of the five mutations fixed by an experimentally evolved isolate of TEV adapted to *Arabidopsis thaliana* L*er*-0 named TEV-*At*17 (*i.e.*, $2^5 = 32$ genotypes) quantify their fitness in the new host *A. thaliana* L*er*-0 and construct and analyze the empirical fitness landscape for this emerging plant RNA virus.

CHAPTER III - $G \times E$ INTERACTIONS

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

Jasna Lalić[1], José M. Cuevas[1], Santiago F. Elena[1,2]*

1 Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas–Universidad Politécnica de Valencia, València, 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

Editor: David S. Guttman, University of Toronto, Canada

Received August 3, 2011; **Accepted** September 22, 2011; **Published** November 17, 2011

Copyright: © 2011 Lalić et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by the Spanish Ministry of Science and Innovation grant BFU2009-06993 to SFE. JL and JMC were supported by the JAE program from CSIC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: santiago.elena@csic.es

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 [5–8]. 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*, *Hantan virus*, 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 naïve 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

Author Summary

Mutations are the raw material on which natural selection operates to optimize the fitness of populations. The occurrence of selection and its strength depend on the effect that mutations may have on the survival and reproduction of individuals: mutations can be lethal, deleterious, neutral, or beneficial. Thus, determining how many mutations belong to each of these categories is of importance for predicting the evolutionary fate of a population. For emerging infectious diseases, this distribution determines the likelihood that a pathogen crosses the species barrier and successfully infects a new host. We characterized such distributions across a panel of alternative hosts for a plant virus and found that fitness effects of individual mutations varied across hosts in an unpredictable way and that many mutations considered deleterious in the natural host may turn out to be beneficial in other hosts.

fitness effects in the primary host. DMFE have been characterized in recent years for a handful of single-stranded DNA [16,17] and RNA viruses [16,18–20] in their primary hosts. All these studies but one [18] took a similar experimental approach to the characterization of DMFEs. In all cases, site-directed mutagenesis was performed on infectious clones, generating collections of random single-nucleotide substitution mutants. The fitness of these mutants was then measured by means of competition experiments against the parental non-mutated virus. In [18] (and in some experiments described in [16]), an undetermined number of mutations were fixed by genetic drift in the absence of purifying selection (Muller's ratchet). Three commonalities can be found in these studies [21], which we will briefly summarize. First, all viruses examined show very low tolerance to mutation, as demonstrated by a large fraction of lethal mutations (between 20% and 40%). Second, for non-lethal mutations, the mean fitness loss associated to a single nucleotide substitution is about 10%. Third, DMFEs characterized are left-skewed (i.e., containing more negative values than the Gaussian distribution) and leptokurtic (i.e., comprising less central values than the Gaussian and having longer tails). Accordingly, the probability density functions that better fitted the data were from the heavy-tailed family (Log-normal or Weibull) or highly skewed ones (Gamma or Beta). Still, probably due to the overwhelming amount of work associated with these studies, the effect of host heterogeneity on the properties of DMFE have not been experimentally addressed; with the exception of the work done by Van Opijnen et al. [22] with HIV-1. However, this study was limited to a few single nucleotide-substitution mutations that were not randomly scattered along 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 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 lycopersicum*, *Capsicum annuum*, and *Datura stramonium*), all belonging to the same plant family, the *Solanacea*. 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 (Table 1). Our set of mutants consisted in changes that were randomly dispersed throughout the TEV genome (Table 1). Selected mutants were all viable in the natural host *N. tabacum*. The absolute fitness effects of these genotypes were evaluated in eight susceptible host species. The observed DMFEs for the 21 genotypes in all eight hosts are shown in Figure 1. A quick look at these histograms suggests that in the natural host *N. tabacum* and in its close relative *N. benthamiana* (both species belong to the same genus of the *Nicotianoideae* subfamily) most mutants have absolute fitness indistinguishable from or below the value of the wildtype (indicated by the vertical dashed line; enumerated in Table 2). Indeed, the average absolute fitness values for all mutant genotypes on these two hosts were significantly smaller than the values estimated for the wildtype (Table 2; one-sample *t*-tests, $P \leq 0.019$ in both cases). Also supporting this excess of deleterious effects, the distributions had significant negative skewness values (Table 2; *t*-test comparing to the Gaussian null expectation,

Table 1. TEV genotypes used in this study and some of their properties.

Genotype	Protein	Location	Nucleotide substitution	Amino acid change	Polarity change
DQ986288, wild-type isolate					
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	NIa-VPg	6012	U→G	I→M	
PC69	NIa-VPg	6044	C→A	T→N	
PC70	NIa-VPg	6197	U→G	M→R	apolar→basic
PC72	NIa-VPg	6251	U→C	F→S	apolar→polar
PC76	NIa-Pro	6519	U→C	synonymous	
PC83	NIb	7315	A→G	I→V	
PC95	NIb	8501	A→C	E→A	acid→polar

doi:10.1371/journal.pgen.1002378.t001

$P<0.001$ in both cases). The average absolute fitness effect of all genotypes together was undistinguishable in these two hosts (Mann-Whitney test, $P=0.232$). Both distributions are also significantly leptokurtic (Table 2; *t*-test comparing to the Gaussian null expectation, $P<0.001$ in both cases), indicating that many mutations have mild fitness effects and, therefore, the DMFEs are more peaked than expected for a Gaussian distribution. When the absolute fitness of the different TEV mutants was evaluated in hosts whose genetic relatedness to *N. tabacum* decreased, while still belonging to the *Solanaceae* (*Solanoideae* subfamily: *D. stramonium*, *C. annuum* and *S. lycopersicum*), the average value of the distributions did not shift significantly compared to *Nicotianoideae* (Mann-Whitney test, $P=0.348$). In addition, it remained skewed towards the left tail, that is, the values were smaller than the median of the distribution (Table 2; *t*-test, $P\leq0.026$). In *D. stramonium* and *S. lycopersicum*, a few mutations were lethal (see below the arguments supporting the lethality of these mutants), thus making the distributions even more negatively skewed. The change in shape of DMFE noticeably affected the kurtosis parameter. In the three *Solanoideae* hosts DMFEs have no significant kurtosis (Table 2; *t*-tests, $P\geq0.195$ in all cases), and thus they are effectively mesokurtic (e.g., Gaussian-like). In general, DMFE dramatically change in several aspects within non-*Solanaceae* hosts. First, the distribution mean shifts towards lower values; a comparison of absolute fitness values between *Solanaceae* and non-*Solanaceae* hosts indicates that the difference is highly significant (Mann-Whitney test, $P<0.001$). Second, the distributions become positively skewed, although the asymmetry was significant only for *S. oleracea* (Table 2; *t*-test, $P=0.008$). Positive skewness means that the tail of the distribution containing fitness effects higher than the mean is significantly heavier than the negative tail. This finding is particularly interesting when observed that the fitness of the wildtype is always in the negative tail of the distribution.

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 *Nicotianeae* towards more beneficial and lethal in the three *Solanoideae*, 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, $P\geq0.131$ 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

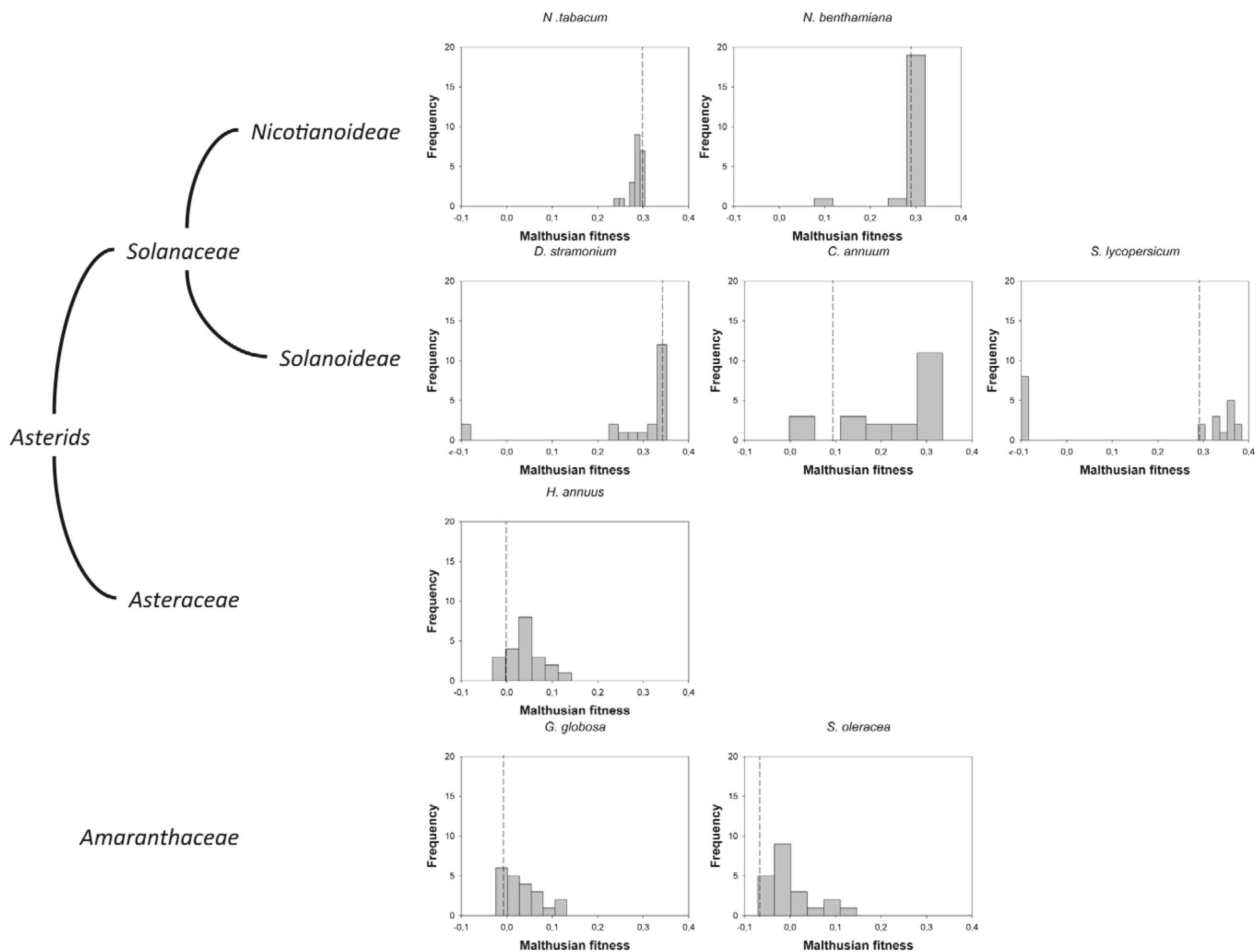


Figure 1. DMFEs across different host species. Host species belong to the taxonomic families *Solanaceae*, *Asteraceae* and *Amaranthaceae*. The first two families belong to the Asterids class. In nature, TEV is found infecting members of the *Solanaceae* family. The ancestral isolate used in this study was obtained from and subsequently passed in *N. tabacum* plants. Lethal mutations (which have a Malthusian fitness of $-\infty$) 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	0	9	11
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

t-test significance levels for skewness and kurtosis:
*$0.05>P\geq0.01$,
**$0.01>P\geq0.001$;
***$P<0.001$.
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. Out 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 ($21 \times 0.007 = 0.153$) 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 λ and the shape κ, 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 Akaike'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 $E(m) = \alpha c/(\alpha - 1)$, where α is the shape

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 $b = 0.104 \pm 0.019$				
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 $b = 0.032 \pm 0.014$				
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 $\alpha = 22.189 \pm 1.493$				

[a] ±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

parameter and c the threshold value. For the two non-Asterids hosts (e.g., *G. globosa* and *S. oleracea*) the expected fitness values were negative, whereas in all other cases the expected fitness values were positive and in the range 0.02–0.311.

The Akaike's weight informs about which one among a set of competing models is best supported by the data, after ranking them according to their *AIC* values. However, given the uncertainties associated to the small sample size here used (21 TEV genotypes), one may be interested in evaluating how much better performs the best fitting model relative to any other model. 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 was ranked as the second best fitting, performing only ~5.5 times worse than the Laplace pdf. For *H. annuus*, 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 $ER = 190935.254$, indicative of a very poor fit compared to the best fitting Pareto pdf.

The phylogenetic distance between natural and naïve hosts influence 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., $P = 0.018$) 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 $G \times E$ 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 ca. 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 $G \times E$ interaction term is also highly significant, and explains ca. 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 $G \times E$ 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 $G \times E$ interaction term.

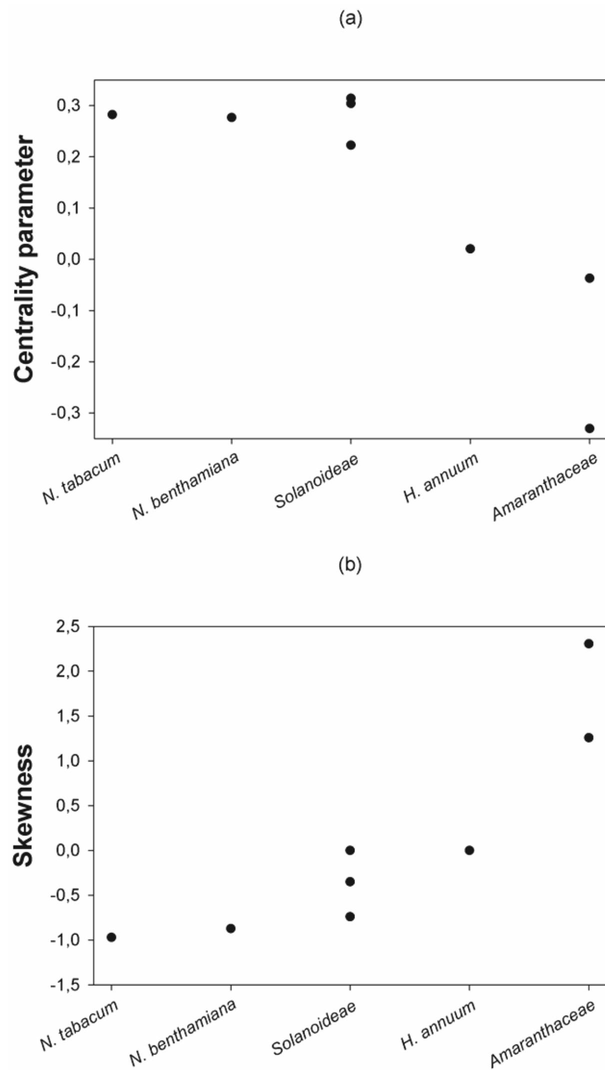


Figure 2. Changes in the centrality and shape parameters of the DMFE with increasing genetic distance among hosts. (a) The centrality parameter of the best fitting pdf shifts from positive to negative Malthusian fitness, indicating that the average effect of single mutations is stronger as the host genetic relatedness with the natural host *N. tabacum* decreases. (b) Distributions become more positively skewed with increasing host genetic distance from *N. tabacum*, suggesting that more mutations have positive effect in the new hosts.
doi:10.1371/journal.pgen.1002378.g002

Table 4. Two generalized lineal models testing the effect of TEV genetic background (G), host species (E), and their interaction ($G \times E$).

Source of variation	χ^2	d.f.	P	Variance component[a]	Percentage of variance[b]
Model I ($AIC^c = -2328.299$)					
G (TEV genotype)	2783.062	20	<0.001	4.48×10^{-3}	4.29%
E (Host species)	6467.415	7	<0.001	2.73×10^{-2}	26.13%
$G \times E$	7282.589	140	<0.001	6.99×10^{-2}	66.82%
Model II ($AIC = -2412.799$)					
G (TEV genotype)	2783.062	20	<0.001	4.32×10^{-3}	4.17%
Host class	1371.172	1	<0.001	8.56×10^{-3}	8.25%
E (species within *Host class*)	3177.883	6	<0.001	1.81×10^{-2}	17.47%
$G \times E$	7282.589	140	<0.001	6.99×10^{-2}	67.33%

Both variables were treated as random sources.
[a]Maximum-likelihood estimators.
[b]For *Model I*, computed using a value of error variance equal to 2.88×10^{-3}, which is equivalent to a 2.76% of unexplained variance. For Model II, computed with an error variance 2.88×10^{-3} (2.77%).
[c]Akaike information criterion.
doi:10.1371/journal.pgen.1002378.t004

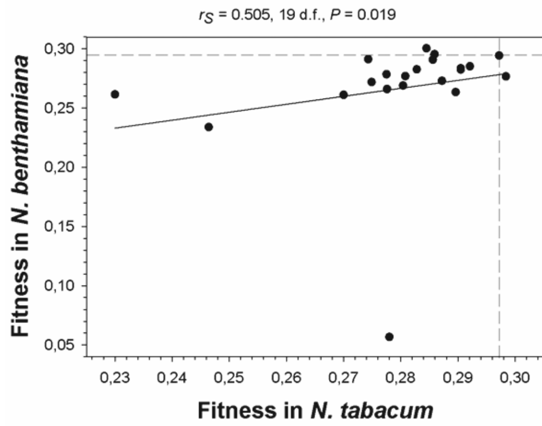
The causes of $G \times E$

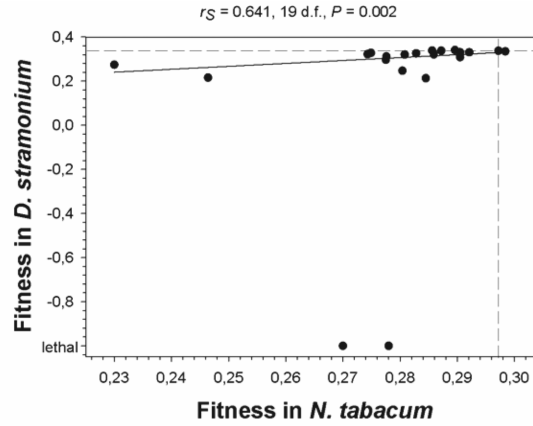
A significant $G \times E$ interaction can be produced by two non-mutually exclusive mechanisms [26]. First, pleiotropic effects may change the rank order of mutations across environments (e.g., a mutation beneficial in one environment may not be so in an alternative one). Second, while still retaining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

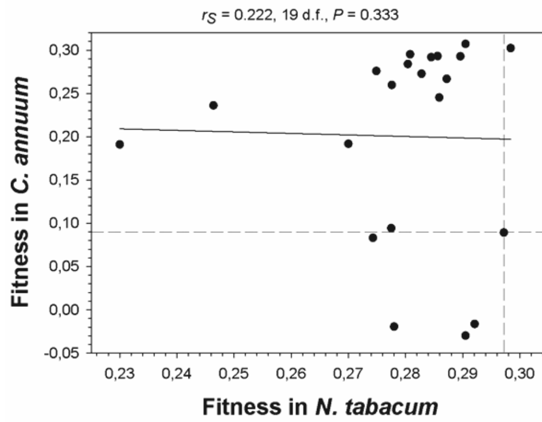
As a first statistical test, we computed Spearman's rank correlation coefficients between absolute fitness values in the primary host *N. tabacum* and the values estimated on each alternative host (Figure 3). Lethal mutations were assigned to the lowest rank. A negative correlation would indicate negative 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 were viable in *N. tabacum* but lethal in *D. stramonium* and *S. lycopersicum*, all being from the same family.

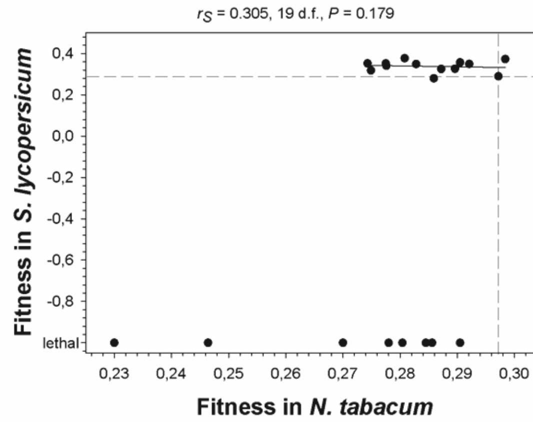
A non-significant correlation test, however, cannot be taken as an evidence of a lack of pleiotropic effects across hosts. For instance, one can imagine a situation in which, in a given host, some mutations may have negative pleiotropic effects, some others positive ones and some even being independent on the host. In such situation, the correlation would turn out to be non-significant while still some mutations may be pleiotropic. To overcome this drawback, we performed a second statistical test based on the frequency of mutations that changed the sign of its fitness effects (compared to that of the wildtype TEV) across hosts. For each mutation on each host, we recorded whether fitness was lower (negative sign) or higher (positive sign) than the wildtype TEV. Then we counted the number of cases for which the sign changed between the primary host, *N. tabacum*, and each alternative one. If a mutation has the same sign both in the primary and in the alternative hosts, it is considered not to be pleiotropic. By contrast, if sign changes, then it is considered as pleiotropic. Under the null hypothesis of no excess of pleiotropic effects, mutations would distribute evenly across both categories. Departures from this null hypothesis were evaluated using Binomial tests. Only in *N. benthamiana* ($x = 2$) and *D. stramonium* ($x = 4$) the number of observed mutations with putative pleiotropic effects was not significantly larger than expected under the null expectation (probability of having x or more cases of pleiotropic mutations than expected by sheer chance: $P < 0.001$ and $P = 0.006$, respectively). By contrast, the number of mutations whose fitness effects switched signs were significantly larger than expected by chance in all other hosts: $x = 18$ in *C. annuum* ($P > 0.999$), $x = 19$ in *S. lycopersicum* ($P > 0.999$), $x = 14$ in *H. annuus* ($P = 0.942$), 15 in *G. globosa* ($P = 0.979$), and 17 in *S. oleracea* ($P > 0.999$). Therefore, this second test of antagonistic pleiotropy confirmed the conclusions drawn from the Spearman's correlation test. Moreover, it showed that antagonistic pleiotropy also made an important contribution to the fitness variability observed in the two hosts (*C. annuum* and *S. lycopersicum*) in which no overall tendency was observed in Figure 2.

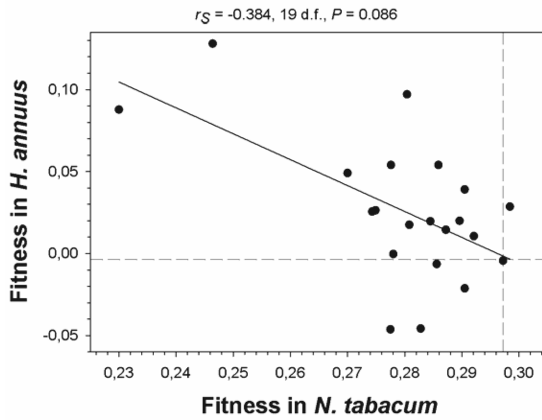
Next, to evaluate the importance of changes in genetic variance, σ_G^2, for absolute fitness as a source of $G \times E$ we computed it for each of the eight host species. Table 5 shows the estimates of σ_G^2, of error variance (σ_e^2) as well as the broad sense heritability (H^2) that indicates the percentage of total phenotypic variance explained by genetic differences among TEV genotypes. For the five *Solanaceae* hosts, σ_G^2 ranged from 0.051 to 0.115, with an average value of 0.083, and σ_G^2 explaining >95% of the observed phenotypic variance. By contrast, σ_G^2 within the non-*Solanacea* hosts was

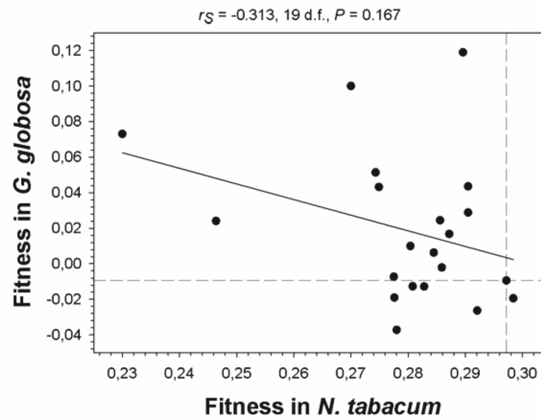

r_S = 0.505, 19 d.f., P = 0.019
Fitness in N. benthamiana
Fitness in N. tabacum


r_S = 0.641, 19 d.f., P = 0.002
Fitness in D. stramonium
lethal
Fitness in N. tabacum


r_S = 0.222, 19 d.f., P = 0.333
Fitness in C. annuum
Fitness in N. tabacum


r_S = 0.305, 19 d.f., P = 0.179
Fitness in S. lycopersicum
lethal
Fitness in N. tabacum


r_S = -0.384, 19 d.f., P = 0.086
Fitness in H. annuus
Fitness in N. tabacum


r_S = -0.313, 19 d.f., P = 0.167
Fitness in G. globosa
Fitness in N. tabacum

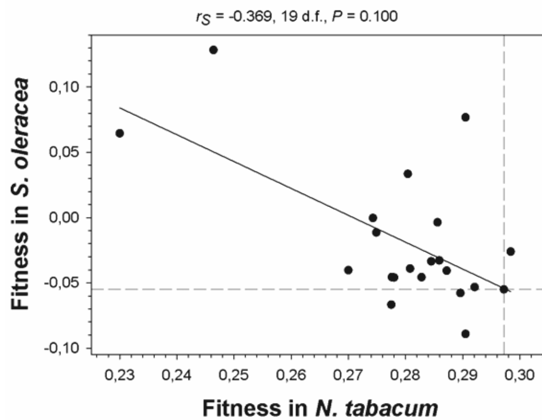

r_S = -0.369, 19 d.f., P = 0.100
Fitness in S. oleracea
Fitness in N. tabacum

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 ~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 a 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 ϕ6 [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).

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

Host species	σ_G^2 (×10^{-2})	σ_e^2 (×10^{-4})	H^2
N. tabacum	7.858±0.059	3.524±0.000	0.996
N. benthamiana	7.323±0.051	16.052±0.000	0.979
D. stramonium[a]	9.462±0.097	40.160±0.006	0.959
C. annuum	5.162±0.028	61.520±0.015	0.894
S. lycopersicum[a]	11.475±0.203	6.204±0.000	0.995
H. annuus	0.148±0.000	48.061±0.006	0.236
G. globosa	0.109±0.000	47.062±0.006	0.188
S. oleracea	0.195±0.000	46.762±0.005	0.294

[a]Lethal alleles were removed from the computations because they have absolute fitness −∞.
doi:10.1371/journal.pgen.1002378.t005

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 Qβ [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 ~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 Qβ DMFE was well described by a Gamma pdf. In the case of DNA phages ϕX174 [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]. ϕX174 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

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 *Saccharomyces cerevisiae* and *Cryptococcus neoformans*, the nematode *Caenorhabditis elegans*, and the fruitfly *Drosophila melanogaster*, Martin and Lenormad [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 $G\times E$ interactions

The existence of $G\times E$ 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 $G\times E$ 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 $G\times E$ interaction. Mutations involved in significant $G\times E$ 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 $G\times E$ is a general norm. In the context of emerging viral infections, the existence of a significant $G\times E$ 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 $G\times E$: 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 $G\times E$ and found that both indeed coexist. Antagonistic pleiotropy does not contribute significantly to $G\times E$ 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 $G\times E$ interaction can be explained both by antagonistic pleiotropy and by changes in the genetic component of variance. Previous studies with *E. coli* showed that $G\times E$ 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 $G\times E$ 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. tabacum* 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. tabacum* (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. Daròs, was used to generate both the wildtype virus and the mutant genotypes. Single-nucleotide substitution mutants were generated by site-directed mutagenesis using QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) as described in [20] and following the manufacturer's instructions. The kit incorporates PfuUltra 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 *Nicotianoideae* subfamily whereas the other three belong to the *Solanoideae* 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. globosa* 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 ($9 \times 8 \times 21 = 1512$) 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:isoamylic 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 hosts 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 II (TaKaRa) following the instructions provided by the manufacturer. The primers forward TEV-CP 5′-TTGGTCTTGATGGCAACGTG and reverse TEV-CP 5′-TGTGCCGTTCAGTGTCTTCCT 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

42°C followed by 10 s at 95°C; and PCR phase of 40 cycles of 5 s at 95°C and 31 s at 60°C. Quantification results were examined using SDS7000 software v. 1.2.3 (Applied Biosystems).

Statistics

Absolute fitness was estimated as Malthusian growth rate per day, according to expression $m = \frac{1}{t} \log Q$, where Q is the number of pg of TEV RNA per 100 ng of total plant RNA quantified at $t = 10$ dpi.

Unless otherwise indicated, all statistical tests were performed using SPSS version 19. Generalized linear models (GLM) were used to explore the effect of the different factors on TEV fitness. We assumed that m was distributed either as a Gaussian pdf or as a more stretched Gamma pdf. In both cases an identity link function 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 Àngels Pròsper for excellent technical assistance and Stéphanie Bedhomme, José A. Daròs, Guillaume 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. Woolhouse MEJ, 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 epidemic diseases. Microbiol Mol Biol Rev 72: 457–470.
3. Holmes EC (2009) The evolutionary genetics of emerging viruses. Annu Rev Ecol Evol Syst 40: 353–372.
4. Elena SF, Froissart R (2010) New experimental and theoretical approaches towards the understanding of the emergence of viral infections. Phil Trans R Soc B 365: 1867–1869.
5. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, et al. (2004) Emerging infectious diseases of plants: pathogen pollution, 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-Microb 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. Gentics 156: 1465–1470.
11. Duffy S, Turner PE, Burch CL (2006) Pleiotropic costs of niche expansion in the RNA bacteriophage ϕ6. Genetics 172: 751–757.
12. Ferris MT, 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 potyvirus*. Infect Genet Evol 8: 806–814.
14. Gandon S (2004) Evolution of multihost parasites. Evolution 58: 455–469.
15. Remold SK, Rambaut A, Turner PE (2008) Evolutionary genomics of host adaptation in *Vesicular stomatitis virus*. Mol Biol Evol 25: 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 f1. Genetics 185: 603–609.
18. Elena SF, Moya A (1999) Rate of deleterious mutations and the distribution of its effects on fitness in *Vesicular stomatitis 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: 8396–8401.
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 Opijnen T, Boerlijst MC, Berkhout 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 *G*×*E*. Trends Ecol Evol 24: 649–658.
24. Futuyma DJ, Moreno G (1988) 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) Contribution of individual random mutations to genotype-to-environment interactions in *Escherichia coli*. Proc Natl Acad Sci USA 98: 11388–11393.
27. Soltis DE, Soltis 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, Ebendick-Corpus BE (2004) Positive selection of synonymous mutations in *Vesicular stomatitis virus*. J Mol Biol 342: 1415–1421.
29. Ohta 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, Villareal 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 258: 221–227.
37. Via S, Gomulkiewicz R, 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. Auld JR, Agrawal AA, Relyea 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, Codoñer FM, Martín S, et al. (2008) Experimental evolution of plant RNA viruses. Heredity 100: 478–483.
42. Ayme V, Souche S, Caranta C, Jacquemond M, Chadoeuf J, et al. (2006) Different mutations in the genome-linked VPg of *Potato virus Y* confers virulence on the *pvr2(3)* resistance in pepper. Mol Plant-Microbe Interact 19: 557–563.
43. Charron C, Nicolaï M, Gallois JL, Robaglia C, Moury B, et al. (2008) Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J 54: 56–68.
44. Bedoya LC, Daròs JA (2010) Stability of *Tobacco etch virus* infectious clones in plasmid vectors. Virus Res 149: 234–240.
45. Carrasco P, Daròs JA, Agudelo-Romero P, Elena SF (2007) A real-time RT-PCR assay for quantifying the fitness of *Tobacco etch virus* in competition experiments. J Virol Meth 139: 181–188.
46. Lalić J, Agudelo-Romero P, Carrasco P, Elena SF (2010) Adaptation of *Tobacco etch potyvirus* to a susceptible ecotype of *Arabidopsis thaliana* capacitates it for systemic infection of resistant ecotypes. Phil Trans R Soc B 65: 1997–2008.
47. Pfaffl MV (2004) Quantification strategies in real-time PCR. In *A-Z of Quantitative PCR* (ed. SA Bustin). International University Line: La Jolla, USA. pp 87–112.

CHAPTER IV - $G \times G$ INTERACTIONS

Heredity

Heredity (2012) 109, 71–77
© 2012 Macmillan Publishers Limited. All rights reserved 0018-067X/12
www.nature.com/hdy

ORIGINAL ARTICLE

Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus

J Lalić[1] and SF Elena[1,2]

How epistatic interactions between mutations determine the genetic architecture of fitness is of central importance in evolution. The study of epistasis is particularly interesting for RNA viruses because of their genomic compactness, lack of genetic redundancy, and apparent low complexity. Moreover, interactions between mutations in viral genomes determine traits such as resistance to antiviral drugs, virulence and host range. In this study we generated 53 *Tobacco etch potyvirus* genotypes carrying pairs of single-nucleotide substitutions and measured their separated and combined deleterious fitness effects. We found that up to 38% of pairs had significant epistasis for fitness, including both positive and negative deviations from the null hypothesis of multiplicative effects. Interestingly, the sign of epistasis was correlated with viral protein–protein interactions in a model network, being predominantly positive between linked pairs of proteins and negative between unlinked ones. Furthermore, 55% of significant interactions were cases of reciprocal sign epistasis (RSE), indicating that adaptive landscapes for RNA viruses maybe highly rugged. Finally, we found that the magnitude of epistasis correlated negatively with the average effect of mutations. Overall, our results are in good agreement to those previously reported for other viruses and further consolidate the view that positive epistasis is the norm for small and compact genomes that lack genetic robustness.
Heredity (2012) **109,** 71–77; doi:10.1038/hdy.2012.15; published online 11 April 2012

Keywords: epistasis; fitness landscapes; genome architecture; virus evolution

INTRODUCTION

Epistasis has been the focus of intensive research since the beginning of genetics as a scientific discipline (Phillips, 2008). In general, epistasis is the interaction between genes or mutations in determining phenotypes. The direction, magnitude and prevalence of epistasis is central to theories seeking to explain the origin of characteristics of genetic systems, such as sex and recombination (De Visser and Elena, 2007), dominance (Bagheri and Wagner, 2004), ploidy (Kondrashov and Crow, 1991), phenotypic plasticity (Remold and Lenski, 2004), robustness (De Visser *et al.*, 2003), the ruggedness of adaptive landscapes (Weinreich *et al.*, 2006; Poelwijk *et al.*, 2007), or attempting to mechanistically explain dynamic biological processes such as the accumulation of mutations in finite populations (Kondrashov, 1994), and speciation by reproductive isolation (Coyne, 1992). Very recently, the evolutionary causes of epistasis, and not only their evolutionary consequences, have also attracted attention (Sanjuán and Nebot, 2008; De Visser *et al.*, 2011; Macía *et al.*, 2012).

Broadly speaking, epistatic interactions can be classified as *uni-* or *multi*-dimensional (Kondrashov and Kondrashov, 2001). Uni-dimensional epistasis is defined as deviations from a linear relationship between mean multiplicative fitness and the number of mutations affecting fitness. By contrast, multi-dimensional epistasis includes all the possible individual interactions among a set of mutations. Multi-dimensional epistasis provides a more complete description of the interactions within the fitness landscape defined by a set of mutations. Interactions can be further classified as *magnitude* or as *sign* epistasis. Magnitude epistasis (ME) occurs when that the fitness value associated to a mutation, but not its sign, changes upon the genetic background wherein it appears (Weinreich *et al.*, 2005; Poelwijk *et al.*, 2007). Moreover, ME can be positive or negative, depending on whether the fitness of the double mutant is larger or smaller than expected under the multiplicative null model, respectively. ME is a widespread phenomenon observed in organisms of different complexity (Sanjuán and Elena, 2006). Sign epistasis (SE) refers to cases in which the sign of the fitness effect of a mutation is under epistatic control; thus, such a mutation is beneficial in some genetic backgrounds and deleterious in others (Weinreich *et al.*, 2005; Poelwijk *et al.*, 2007). In few instances where it has been sought, SE seems to be quite common, although perhaps not as ubiquitous as ME (Weinreich *et al.*, 2006 Poelwijk *et al.*, 2007; Franke *et al.*, 2011; Kvitek and Sherlock, 2011).

Epistasis is particularly relevant for our understanding of adaptive evolution, as it determines the ruggedness of the adaptive landscape (Withlock *et al.*, 1995; Poelwijk *et al.*, 2011) as well as the accessibility of adaptive pathways throughout the landscape (Weinreich, 2005 Welch and Waxman, 2005; Franke *et al.*, 2011). Evolutionary trajectories may end at suboptimal fitness peaks due to the ruggedness of the fitness landscape. Epistasis can therefore hamper the efficiency of natural selection and thus slow down the rate of adaptation (Withlock *et al.*, 1995). Moreover, epistasis can make certain evolutionary pathways towards higher fitness genotypes selectively inaccessible because of troughs and valleys in the fitness landscape: intermediate genotypes have reduced fitness compared with surrounding genotypes. Weinreich *et al.* (2005) were the first to notice this evolutionary constraint and to postulate that such limitation would arise only as a consequence of SE. Indeed, a particular type of

[1]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, València, Spain and and [2]Santa Fe Institute, Santa Fe NM, USA
Correspondence: Professor SF Elena, Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, Campus UPV CPI 8E, Ingeniero Fausto Elio s/n, 46022 València, Spain.
E-mail: santiago.elena@csic.es
Received 3 October 2011; revised 21 February 2012; accepted 28 February 2012; published online 11 April 2012

SE known as RSE, that is, the sign of the fitness effect of a mutation is conditional upon the state of another locus and *vice versa*, has been shown to be a necessary condition for an adaptive landscape to be rugged (Poelwijk *et al*., 2011).

RNA viruses are ideal experimental systems for exploring the nature of epistatic interactions: their compact genomes often code for overlapping reading frames, contain functional RNA secondary structures and encode for multi-functional proteins. Altogether, these properties are expected to lead to strong epistasis. Indeed, recent studies exploring uni- and multi-dimensional epistasis have provided empirical evidences that ME is common for RNA viruses such as *Foot-and-mouth disease virus* (Elena, 1999), bacteriophage ϕ6 (Burch and Chao, 2004), *Vesicular stomatitis virus* (VSV, Sanjuán *et al*., 2004), *Human immunodeficiency virus* type 1 (Bonhoeffer *et al*., 2004; Van Opijnen *et al*., 2006; Parera *et al*., 2009; Da Silva *et al*., 2010; Martínez *et al*., 2011), *Rous sarcoma virus* (Sanjuán, 2006), or *Tobacco etch virus* (TEV; De la Iglesia and Elena, 2007b), among others, as well as for ssDNA bacteriophages such as ID11 (Rokyta *et al*., 2011) or ϕX174 (Pepin and Wichman, 2007). Furthermore, in most of these studies positive epistasis is more abundant than negative epistasis, although variability exists within each virus. Positive epistasis may appear as a consequence of individual mutations having a large negative impact on fitness such that any additional mutation that still produces a viable virus must necessarily exert a minor impact (Elena *et al*., 2010). SE, by contrast, has been detected only among compensatory mutations for ϕX174 (Poon and Chao, 2006) and among pairs of beneficial mutations for ID11 (Rokyta *et al*., 2011). By contrast, no evidence of SE was found for combinations of beneficial mutations in the RNA bacteriophage MS2 (Betancourt, 2010).

In this study we sought to characterize the patterns of multi-dimensional epistasis for the RNA plant virus TEV (genus *Potyvirus*, family *Potyviridae*). TEV has a single-stranded positive-sense RNA genome of ca. 9.5 Kb that encodes for a single polyprotein that self-processes into 10 mature peptides. An additional peptide is translated from an overlapping ORF after +2 frameshifting. To this end, we generated a collection of 53 double mutants by randomly combining 20 individual mutations whose deleterious fitness effect had been previously quantified (Carrasco *et al*., 2007b). The fitness of all single and double mutants was evaluated in the primary host *Nicotiana tabacum*. We characterized the statistical properties of the distribution of epistatic interactions and found a mixture of positive and negative effects (including some examples of synthetic lethals (SLs)). Next, we found that RSE was the most common type of epistasis. We also explored the negative association between the average fitness effect of deleterious mutations and the strength of the epistastic interaction in which they were involved. Finally, we tried to frame the observed epistatic effects within a model of the protein–protein interaction network (PPIN) formed by all 11 TEV proteins.

There are many novelties within our study. First, this is the first description of extensive SE, particularly of the reciprocal type, contributing to the architecture of fitness of an RNA virus. Second, we contextualize epistasis in the network of interactions among viral proteins. Third, it is the first report of epistasis for a eukaryotic virus in its natural host rather than in *in vitro* cell cultures, which represent an artificial and oversimplified environment. Last, but not least, this is the first analysis of multidimensional epistasis for any plant pathogen.

MATERIALS AND METHODS

Virus genotypes

A subset of 20 mutants non-lethal in *N. tabacum* (Supplementary Table 1) was randomly chosen from a larger collection generated in a previous study (Carrasco *et al*., 2007b). Six were synonymous mutations, whereas the rest were nonsynonymous. Plasmid pMTEV (Bedoya and Daròs, 2010) was used to reconstitute the wild-type TEV and to generate the mutant genotypes. These 20 mutations were randomly combined to generate a set of 53 double mutants (Supplementary Table 2) by site-directed mutagenesis using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene, Santa Clara, CA, USA) as described by Carrasco *et al*. (2007a). The kit incorporates PfuUltra 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 mutated nucleotide.

Infectious RNA of each genotype was obtained by *in vitro* transcription after BglII linearization of the corresponding plasmid as described in Carrasco *et al*. (2007a).

Inoculation experiments

All *N. tabacum* plants were inoculated at an identical growth stage to minimize variations in defense response to infection with developmental stage. All inoculations were done in a single experimental block. Nine plants per TEV genotype 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^{-1}).

Ten days post-inoculation, the whole infected plant, except the inoculated leaf, was collected. The collected tissue was frozen in liquid nitrogen and grounded 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: isoamylic alcohol (25:25:1) was added, thoroughly vortexed and centrifuged at 14 000 g for 5 min at 25 °C. Ca. 160 μl of the upper aqueous phase was mixed with 80 μl of a solution containing 7.5 M LiCl and 50 mM EDTA and incubated overnight on ice at 4 °C. The precipitated RNA was centrifuged at 14000 g for 15 min at 4 °C, washed once with 70% ice-cold ethanol, dried in a SpeedVac (Thermo Fisher Scientific, Waltham, MA, USA) 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^{-1}.

Within-plant virus accumulation was measured by absolute RT–qPCR using an external standard (Pfaffl, 2004). Standard curves were constructed using five serial dilutions of TEV RNA produced by *in vitro* transcription and diluted in RNA obtained from the host plant.

RT–qPCR reactions were performed in 20 μl volume using One Step SYBR PrimeScript RT–PCR Kit II (TaKaRa, Bio Inc, Otsu, Japan) following the instructions provided by the manufacturer. The forward TEV–CP 5′-TTGGTCTTGATGGCAACGTG and reverse TEV–CP 5′-TGTGCCGTTCAGTGTCTTCCT primers amplify a 71 nt fragment within the TEV CP cistron. CP was chosen because it is located in the 3′ end of TEV genome and hence would only quantify complete genomes. Each RNA sample was quantified three times in independent experiments. Amplifications were done using the ABI PRISM Sequence Analyzer 7000 (Applied Biosystems, Carlsbad, CA, USA). The thermal profile was: RT phase consisted of 5 min at 42 °C followed by 10 s at 95 °C; and PCR phase of 40 cycles of 5 s at 95 °C and 31 s at 60 °C. Quantification results were examined using SDS7000 software v. 1.2.3 (Applied Biosystems).

For each genotype, a Malthusian growth rate per day was computed as $m = (1/t)\log(Q_t)$, where Q_t are the pg of TEV RNA

per 100 ng of total plant RNA quantified at $t = 10$ days post-inoculation. Absolute fitness was then defined as $W = e^m$ (Crow and Kimura, 1970).

Estimation of epistasis among pairs of mutations

Epistasis among pair of mutations x and y, ϵ_{xy}, was calculated as $\epsilon_{xy} = W_{00}W_{xy} - W_{x0}W_{0y}$ (Kouyos *et al.*, 2007), where W_{00}, W_{xy}, W_{x0}, and W_{0y} correspond to the absolute fitness of the wild-type, the double mutant and each single mutant, respectively. A value of $\epsilon_{xy} > 0$ corresponds to the case of positive (antagonistic) epistasis, whereas a value of $\epsilon_{xy} < 0$ is indicative of negative (synergistic) epistasis. Values of ϵ_{xy} not significantly deviating from zero were qualified as multiplicative (that is, non-epistatic) mutational effects.

In all cases, reported error intervals correspond to ± 1 s.e.m. All statistical analyses were performed using IBM SPSS v. 19 (Armonk, NY, USA).

RESULTS

Epistasis among pairs of deleterious mutations

Figure 1 shows the relationship between observed and expected fitness values for the set of 53 double mutant genotypes synthetized for this study (Supplementary Table 2). The solid line represents the null hypothesis of non-epistatic fitness effects. The observed fitness values of 20 double mutant genotypes significantly departed from this null expectation (Supplementary Table 2; *t*-tests, in all cases $P \leqslant 0.049$). Nine of these cases were SLs, which means that two mutations that were viable by themselves become lethal when combined. These SLs represent an extreme case of negative epistasis. All other significant cases corresponded to positive epistasis. Therefore, we found variability in the sign and strength of epistasis. However, only the nine SLs remained significant after applying the more stringent sequential Bonferroni correction for multiple tests of the same hypothesis (Rice, 1989). Nonetheless, for all analyses presented below, we used all 20 significant cases, unless otherwise indicated. This decision represents a compromise between reducing the data set to only the nine SLs (which precludes running any additional analysis) and using the whole data set irrespective of the significance of observed fitness values.

Three double mutants contained two synonymous mutations, 22 combined one synonymous and one nonsynonymous mutation and 28 carried two nonsynonymous mutations. No differences existed, however, in the magnitude of epistasis among these genotypic classes (Kruskal–Wallis test: $\chi^2 = 0.405$, 2 df, $P = 0.817$).

Using the whole data set, we sought to test whether the distribution of SL and viable mutations were homogeneous among pairs of mutations within the same cistron or among affecting different cistrons. In two out of nine SLs both mutations were at the same cistron (22.2%), whereas in the case of viable double mutant genotypes, only one genotype out of 44 had both mutations in the same cistron (2.3%), a significant difference ($\chi^2 = 5.569$, 1 df, $P = 0.018$) despite the small sample size. Furthermore, the average epistasis coefficient computed for mutations within the same cistron was -1.142 ± 0.617, whereas it was reduced to -0.171 ± 0.090 for pairs of mutations affecting different cistrons. This 85.1% relaxation in the strength of epistasis was also significant ($t_{51} = 2.477$, $P = 0.017$). Therefore, we can conclude that a tendency exists for mutations affecting the same cistron to generate a SL phenotype and to interact in a stronger and more negative manner, whereas mutations affecting different viral proteins presented weaker interactions.

Statistical properties of the epistasis distribution

Figure 2 illustrates the distribution of epistasis parameters for all pairs of point mutations analyzed. The distribution had a bimodal shape, with SLs representing the left probability mass and the viable genotypes being on the right side of the distribution. The average epistasis was $\bar{\epsilon} = -0.226 \pm 0.095$, a value that departs from the null hypothesis of multiplicative effects ($t_{52} = 2.376$, $P = 0.021$). Furthermore, the distribution had a significant negative skewness ($g_1 = -1.806 \pm 0.327$; $t_{52} = 5.515$, $P < 0.001$), that is, the tail containing negative epistasis is heavier than the Gaussian and thus asymmetric. Similarly, the distribution was significantly leptokurtic ($g_2 = 1.326 \pm 0.644$, $t_{52} = 2.058$, $P = 0.045$), indicating that it had a more acute peak around the mean value compared with the Gaussian.

Given that lethal mutations are largely irrelevant for evolutionary dynamics, we sought to reanalyze the epistasis distribution after removing SLs. The main consequence of this removal was that the average epistasis then becomes significantly positive ($\bar{\epsilon} = 0.084 \pm 0.005$; $t_{43} = 17.438$, $P < 0.001$). Regarding the shape of the distribution, it still remained asymmetric with significant negative

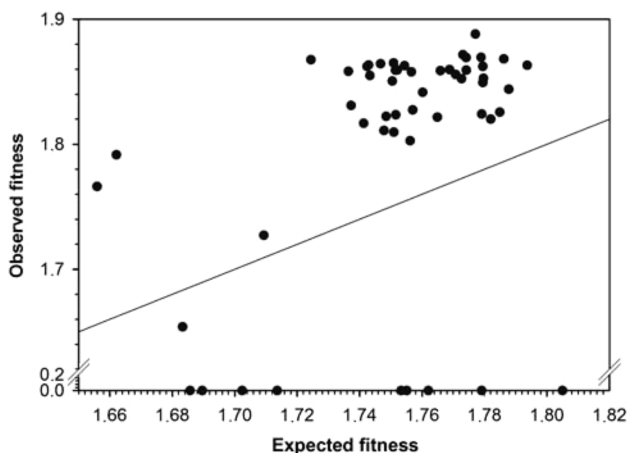


Figure 1 Relationship between observed and expected multiplicative fitness for 53 TEV genotypes carrying pairs of nucleotide substitutions. The solid line represents the null hypothesis of multiplicative fitness effects. Deviations from this line arise as a consequence of the existence of epistatic fitness effects.

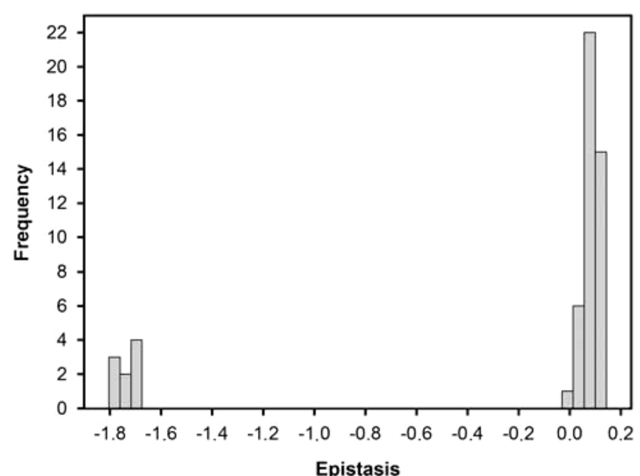


Figure 2 Distribution of epistasis. Epistasis, ϵ, was computed as the difference between the observed fitness of the double mutant ($W_{00}W_{xy}$) and the value expected from subtracting the effects of each single mutant from the wild-type value ($W_{x0}W_{0y}$).

skewness ($g_1 = -1.050 \pm 0.358$; $t_{43} = 2.936$, $P = 0.005$), although the skewness parameter was 41.9% smaller than when SLs were included in the data set. In contrast, the distribution became 77.1% more leptokurtic ($g_2 = 2.348 \pm 0.702$, $t_{43} = 3.346$, $P = 0.002$), as a consequence of the removing the cases from the left tail extreme.

Pervasive RSE

We were interested in evaluating the extent to which SE was present in our data set. Poelwijk *et al.* (2011) defined mathematically the condition for SE as

$$|W_{x0} - W_{00} + W_{xy} - W_{0y}| < |W_{x0} - W_{00}| + |W_{xy} - W_{0y}|$$

Twelve out of the twenty TEV double-mutant genotypes for which we had detected significant epistasis (Supplementary Table 2) fulfilled this condition and thus can be classified as cases of SE. The other eight, hence, correspond to cases of ME. Is this 3:2 proportion expected given the observed fitness values of individual mutations and of the double mutants? To tackle this question we applied the above inequality to the 33 non-epistatic pairs of mutations, founding that 26 fulfilled it, despite not being significant. A Fisher's exact test failed to detect significant differences among epistatic and non-epistatic pairs fulfilling the inequality (1-tailed $P = 0.124$), thus confirming that the observed proportion of ME and of SE was not significantly enriched in the later class. Therefore, we conclude that SE makes a major contribution (60%) to all cases of significant epistasis.

Next, we specifically evaluated the contribution of RSE to the observed pattern of SE. According to Poelwijk *et al.* (2011) the following additional condition must be met by a pair of mutations showing SE in order to be considered as cases of RSE:

$$|W_{0y} - W_{00} + W_{xy} - W_{x0}| < |W_{0y} - W_{00}| + |W_{xy} - W_{x0}| \ .$$

Herewith, this condition was fulfilled by 11 out of 12 cases of SE (91.7%). Indeed, only synthetic lethal PC6/PC76 did not match it. As before, given the fitness of single and double mutants, we tested whether this extremely high prevalence of RSE among cases of SE is to be expected. We counted the number of cases that fulfilled this second inequality (25) among the 26 non-epistatic cases that matched the first one. A Fisher's exact test also showed no significant enrichment in cases of RSE among cases of SE (1-tailed $P = 0.538$). Therefore, we conclude that RSE is common in TEV genome.

Correlation between fitness effects and epistasis

It has been observed that average deleterious mutational effects and the strength of positive epistasis are not independent parameters but, instead, are negatively correlated (Wilke and Adami, 2001 You and Yin, 2002; Wilke *et al.*, 2003; Bershtein *et al.*, 2006; Sanjuán *et al.*, 2006; De la Iglesia and Elena, 2007). We sought to investigate if this negative relationship holds for TEV. Figure 3 shows the relationship between the mean fitness of the two mutations combined and the estimated epistasis for all 53 double mutants. A first observation is that two different and significant relationships exist in correspondence to different phenotypic classes: one for the nine SLs (Spearman's $r_S = -1.000$, 7 df, $P < 0.001$) and another one for the viable genotypes ($r_S = -0.416$, 42 df, $P = 0.005$). However, overall a significant negative correlation existed after controlling for the difference within two phenotypic classes (partial $r = -0.331$, 50 df, $P = 0.017$). The slope for the viable genotypes was significantly smaller than the slope for the SLs (analysis of covariance test for the homogeneity of slopes in Figure 3: $F_{1,49} = 9.212$, $P = 0.004$), suggesting that the underlying mechanisms for the observed relationships were different for each phenotypic class. Indeed, the correlation observed for the SLs is trivial

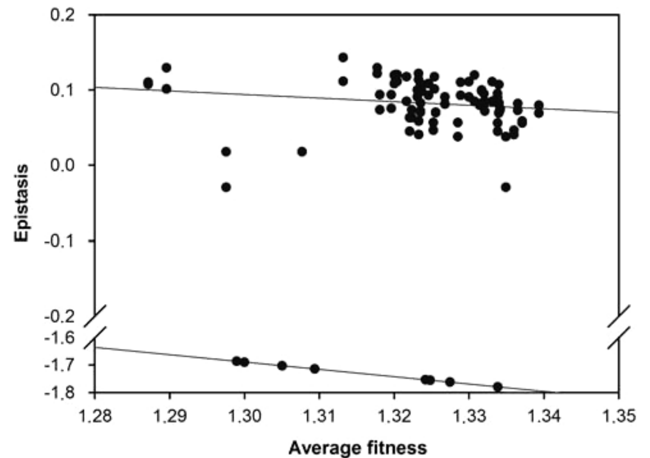


Figure 3 Association between average mutational effects and the magnitude of epistasis. Two apparent relationships exist: one for pairs of mutations generating viable genotypes (upper cloud) and a different one associated to the SLs (lower cloud). The regression lines are included to illustrate the difference in the underlying relationship between epistasis and average mutational effects between both types of phenotypes.

because it is expected based on the definition of epistasis used here. If the observed fitness of the double mutant is $W_{xy} = 0$, then $\epsilon_{xy} = -W_{x0}W_{0y} = -\widetilde{W}^2$, where $\widetilde{W}$ is the geometric mean fitness of mutations x and y. The validity of this explanation was confirmed by the fact that linear regression throughout the origin of epistasis on $\widetilde{W}^2$ for the SLs data rendered the expected slope of -1.000 ± 0.000.

These correlations suggest that mutational effects and epistasis are not independent traits, but instead, they may evolve hand in hand. Stronger mutational effects are associated with more positive interactions, whereas milder effects are associated with more relaxed positive interactions. Therefore, a reduction in the magnitude of mutational effects translates into a relaxation of the positive epistasis.

Epistasis in the context of TEV PPIN

Mutations were grouped according to the mature protein they affect. By doing so, we focused on the analyses of interaction among proteins rather than among individual nucleotide residues. Rodrigo *et al.* (2011) inferred the undirected PPIN shown in Figure 4 using a compendium of physical interactions experimentally determined by the yeast two-hybrid method. We were interested in correlating the network properties with the characteristics of the distribution of epistasis inferred in this study.

First, we sought to test whether the number of significant epistatic and non-epistatic interactions was homogeneously distributed among pairs of proteins directly linked in the PPIN graph or unlinked (Figure 4). A Fisher's exact test failed to reveal a significant association ($P = 0.151$), thus rejecting the hypothesis that a direct interaction between two proteins is a necessary condition to generate significant epistasis.

Second, we explored whether the number of pairs of proteins involved in positive and negative epistatic interactions was evenly distributed among pairs directly connected in the PPIN and those that are not (Figure 4). It has been argued for modularly organized PPINs that mutations affecting independent modules would show a pattern of positive epistasis, although PPINs organized as a single-functional module would be more sensitive to the effect of mutations and show a pattern of negative epistasis (Sanjuán and Elena, 2006; Sanjuán and Nebot, 2008; Macía *et al.*, 2012). In agreement to this expectation,

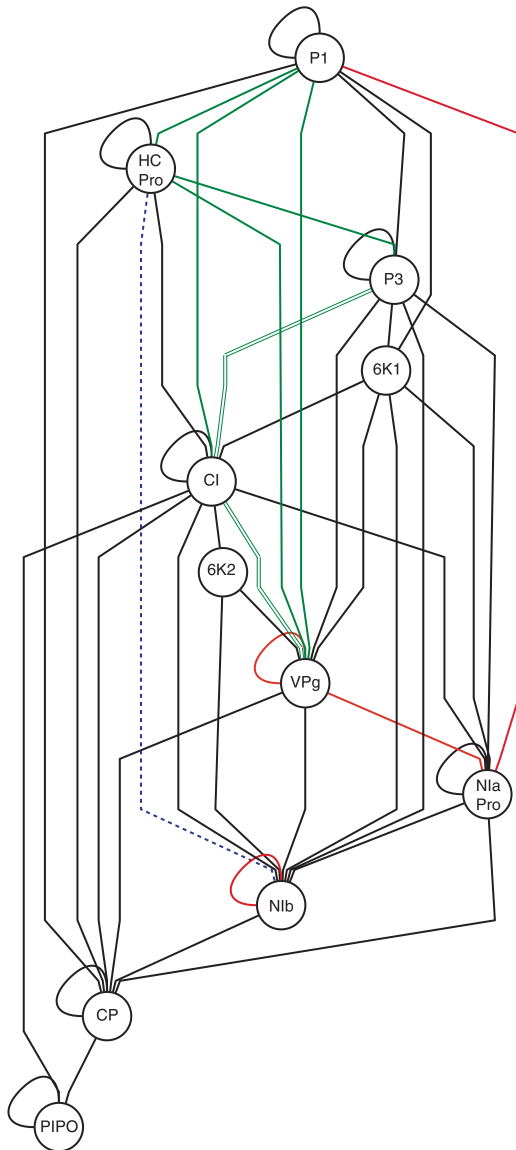


Figure 4 TEV PPIN inferred from yeast two-hybrid data published elsewhere. The 11 mature peptides are indicated as nodes. Black edges correspond to interactions for which we did not detect significant epistasis. Red edges correspond to cases of negative epistasis and green edges correspond to cases of positive epistasis. Double green lines correspond to two pairs of mutations affecting the same proteins. The dashed blue line corresponds to a case in which a first pair of mutations showed positive epistasis (PC19/PC95) but a second pair had negative epistasis (PC22/PC95). The PPIN was drawn using Cytoscape (Killcoyne *et al.*, 2009).

we found that nine out of fourteen (64.3%) positive interactions between linked elements, whereas only two out of six (33.3%) interactions between unlinked elements were positive. Thus, we conclude that mutations affecting connected elements in the PPIN tend to be involved in more positive epistatic interactions than those affecting non-connected components.

Finally, we hypothesized that highly linked nodes would have a stronger tendency to be epistatic, whereas less connected nodes will be less so. To test this hypothesis, we first computed the tendency of a protein to be involved in significant epistasis interactions (that is, epistasiness) for each protein as the absolute value of the average epistasis coefficient computed across all interactions in which this protein was involved and using the whole data set. Absolute values were used because we tested for the tendency for involvement in significant interactions regardless of their sign. Then, we computed the connectivity of each node as the number of links it has in Figure 4. A non-parametric correlation coefficient failed to detect a significant association between these two variables ($r_S = -0.221$, 6 df, $P = 0.599$). Therefore, we conclude that the tendency of a protein to be involved in epistatic interactions is not a necessary consequence of the amount of interactions itself.

DISCUSSION

In this study, the distribution of epistatic interactions on fitness for a plant RNA virus has been evaluated by constructing genotypes carrying pairs of single-nucleotide substitutions, each having a deleterious fitness effect. We detected cases of both positive and negative epistasis, although positive epistases were significantly more abundant than negative ones, such that the combined effect of mutations is significantly less harmful than expected under the null multiplicative model. This dominance of positive epistasis is particularly frequent among mutations affecting two different proteins, whereas, on average, mutations affecting the same protein interact in a negative manner. These findings are in good agreement with observations accumulated in recent years for other RNA viruses, including retroviruses, and small ssDNA viruses (reviewed by Elena *et al.*, 2010), both using experimental approaches to characterize uni- and multi-dimensional epistasis. Given this heterogeneity in viral systems, it thus seems highly likely that positive epistasis among deleterious mutations is a general feature of most small viruses. What may be the mechanistic reason for this excess of positive epistasis? Several reasons can be brought forward. First, the compactness of viral genomes, many of which even had adopted the strategy of overlapping genes and multifunctional proteins, necessarily implies that the deleterious effects of different mutations can partially overlap, hence producing positive epistasis. Indeed, this expectation is well fulfilled by our finding that interactions are, on average, more positive when mutations occur in two different proteins than when they both occur in the same one. Second, positive epistasis can also be a consequence of the existence of elements of RNA secondary structure. Indeed, it was shown by computer simulations of RNA folding that mutations affecting the same structural element may restore it and thus generate positive epistasis (Wilke *et al.*, 2003 Sanjuán *et al.*, 2006). Another observation that seems to be common among experiments of multi-dimensional epistasis in RNA viruses is the existence of frequent cases of synthetic lethality, for example, for the mammalian rhabdovirus VSV (Sanjuán *et al.*, 2004).

The dominance of positive epistasis among deleterious mutations and the existence of frequent cases of synthetic lethality are both fingerprints of another phenomenon: the low genetic robustness of viral genomes. It has been postulated that epistasis and robustness are two sides of the same coin and that negative epistasis must be a hallmark for genetic robustness (De Visser *et al.*, 2003, 2011; Proulx and Phillips, 2005; Desai *et al.*, 2007). Indeed, the observed negative correlation between epistasis and mutational effects shown in Figure 3 provides additional support for this hypothesis and is consistent with observations made in systems as diverse as artificial life (Wilke and Adami, 2001; Edlund and Adami, 2004), computer simulations of genetic systems (You and Yin, 2002; Macía *et al.*, 2012), RNA (Wilke *et al.*, 2003; Sanjuán *et al.*, 2006) and protein folding (Bershtein *et al.*, 2006), and in a mutation–accumulation experiment done with TEV

(De la Iglesia and Elena, 2007). The negative correlation between epistasis and mutational effects means that the milder the average mutational effect is, the more negative the epistatic interactions between mutations will be. This results in a genotype that is more mutationally robust against genetic perturbations. In contrast, positive epistasis reflects strong mutational effects and, therefore, low genetic robustness. Sanjuán and Elena (2006) postulated that robustness would scale up with genetic complexity and that it may result from the fact that more complex genetic systems may contain more redundant structures capable of buffering the effect of mutations. Very recently, Macía *et al.* (2012) tested this hypothesis by simulating the evolution of genetic circuits under variable selection for robustness. They found that, as predicted, negative epistasis was caused by the existence of genetic redundancy in complex networks and not due to complexity itself, as the correlation disappeared when the formation of redundant structures was not allowed during the evolution of complex networks. In this sense, RNA viruses will occupy the lower side of the complexity spectrum and, therefore, would be highly sensitive (that is, non-robust) to mutations.

Within cases of significant epistatic interactions, we found a large contribution of SE relative to the contribution of ME. This represents the first description of SE for an RNA virus, as previous studies of multi-dimensional epistasis in RNA viruses did not explicitly look for SE (for example, Bonhoeffer *et al.*, 2004; Sanjuán *et al.*, 2004; Sanjuán, 2006; Van Opijnen *et al.*, 2006) or simply failed to find them (Betancourt, 2010). In contrast, SE has been shown to be common during adaptation of β-lactamase to cefotaxime (Weinreich *et al.*, 2006; Salverda *et al.*, 2011), in evolution experiments compensating for the cost of antibiotic resistance in bacteria (Schrag *et al.*, 1997; Maisnier-Patin *et al.*, 2002) and viruses (Molla *et al.*, 1996; Cong *et al.*, 2007; Martínez-Picado and Martínez, 2009), in experimental evolution of asexual *Saccharomyces cerevisiae* (Kvitek and Sherlock, 2011), and in multi-dimensional tests of epistasis in *Aspergillus niger* (Franke *et al.*, 2011). All but one cases of SE detected in TEV corresponded to RSE, perhaps making this observation even more interesting. This type of epistasis is particularly relevant from the perspective of describing fitness landscapes. Poelwijk *et al.* (2011) have shown that the existence of multiple adaptive peaks in a fitness landscape, that is, ruggedness, requires RSE. Furthermore, Kwitek and Sherlock (2011) experimentally confirmed that RSE caused the ruggedness of a fitness landscape. The ruggedness of adaptive landscapes is critical to predict whether evolving populations may reach the global optima or may get stuck into suboptimal peaks (Weinreich, 2005; Withlock *et al.*, 1995). Our finding of a predominance of RSE suggests that the fitness landscape for TEV, and maybe for other RNA viruses, must be highly rugged.

In conclusion, the results reported here, together with previous findings, contribute to the perspective that viral genomes are dominated by positive epistasis, which may result from their compactness and lack of genetic redundancy. In addition, we provide the first direct proof that SE, in particular RSE, contributes in a large extent to the architecture of viral fitness. The high frequency of RSE suggests that adaptive landscapes for RNA viruses maybe highly rugged. This ruggedness may impose harsh constraints on the often-invoked but not empirically grounded limitless adaptability of RNA viruses.

DATA ARCHIVING

Data have been archived at Dryad: doi: 10.5061/dryad.bq4pp7f9.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Francisca de la Iglesia and Àngels Pròsper for their excellent technical assistance, Stéphanie Bedhomme and Mark P Zwart for the discussion and Mario A Fares for statistical advice. José A Daròs generously gifted us the pMTEV plasmid. This research was supported by the Spanish Ministry of Science and Innovation grant BFU2009-06993 to SFE. JL was supported by the JAE program from CSIC.

Bagheri HC, Wagner GP (2004). Evolution of dominance in metabolic pathways. *Genetics* **168**: 1716–1735.
Bedoya LC, Daròs JA (2010). Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res* **149**: 234–240.
Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**: 929–932.
Betancourt AJ (2010). Lack of evidence for sign epistasis between beneficial mutations in an RNA bacteriophage. *J Mol Evol* **71**: 437–443.
Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ (2004). Evidence for positive epistasis in HIV-1. *Science* **306**: 1547–1550.
Burch CL, Chao L (2004). Epistasis and its relationship to canalization in the RNA virus ϕ6. *Genetics* **167**: 559–567.
Carrasco P, Daròs JA, Agudelo-Romero P, Elena SF (2007a). A real-time RT-PCR assay for quantifying the fitness of *Tobacco etch virus* in competition experiments. *J Virol Meth* **139**: 181–188.
Carrasco P, de la Iglesia F, Elena SF (2007b). Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J Virol* **81**: 12979–12984.
Cong M, Heneine W, García-Lerma JG (2007). The fitness cost of mutations associated with *Human immunodeficiency virus* type 1 drug resistance is modulated by mutational interactions. *J Virol* **81**: 3037–3041.
Coyne JA (1992). Genetics and speciation. *Nature* **355**: 511–515.
Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*. Harper and Row: New York.
Da Silva J, Coetzer M, Nedellec R, Pastore C, Mosier DE (2010). Fitness epistasis and constraints on adaptation in a *Human immunodeficiency virus* type 1 protein region. *Genetics* **185**: 293–303.
Desai MM, Weissman D, Feldman MW (2007). Evolution can favor antagonistic epistasis. *Genetics* **177**: 1001–1010.
De la Iglesia F, Elena SF (2007). Fitness declines in *Tobacco etch virus* upon serial bottleneck transfers. *J Virol* **81**: 4941–4947.
De Visser JAGM, Elena SF (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. *Nat Rev Genet* **8**: 139–149.
De Visser JAGM, Hermisson J, Wagner GP, Ancel-Meyers L, Bagheri-Chaichian H, Blanchard JL *et al.* (2003). Perspective: Evolution and detection of genetic robustness. *Evolution* **57**: 1959–1972.
De Visser JAGM, Cooper TF, Elena SF (2011). The causes of epistasis. *Proc R Soc B* **10**: 3617–3624.
Edlund JA, Adami C (2004). Evolution of robustness in digital organisms. *Artif Life* **10**: 167–179.
Elena SF (1999). Little evidence for synergism among deleterious mutations in a nonsegmented RNA virus. *J Mol Evol* **49**: 703–707.
Elena SF, Solé RV, Sardanyés J (2010). Simple genomes, complex interactions: epistasis in RNA virus. *Chaos* **20**: 026106.
Franke J, Klözer A, de Visser JAGM, Krug J (2011). Evolutionary accessibility of mutational pathways. *PLoS Comp Biol* **7**: e1002134.
Killcoyne S, Carter GW, Smith J, Boyle J (2009). Cytoscape: a community-based framework for network modeling. *Meth Mol Biol* **563**: 219–239.
Kondrashov AS (1994). Muller's ratchet under epistatic selection. *Genetics* **136**: 1469–1473.
Kondrashov AS, Crow JF (1991). Haploidy or diploidy: which is better. *Nature* **351**: 314–315.
Kondrashov FA, Kondrashov AS (2001). Multidimensional epistasis and the disadvantage of sex. *Proc Natl Acad Sci USA* **98**: 12089–12092.
Kouyos RD, Silander OK, Bonhoeffer S (2007). Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol Evol* **6**: 308–315.
Kvitek DJ, Sherlock G (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet* **7**: e1002056.
Macía J, Solé RV, Elena SF (2012). The causes of epistasis in genetic networks. *Evolution* **66**: 586–596.
Maisnier-Patin S, Berg OG, Lijas L, Andersson DI (2002). Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol Microbiol* **46**: 355–366.

Martínez JP, Bocharov G, Ignatovich A, Reiter J, Dittmar MT, Wain-Hobson S *et al*. (2011). Fitness ranking of individual mutants drives patterns of epistatic interactions in HIV-1. *PLoS ONE* **6**: e18375.
Martínez-Picado J, Martínez MA (2009). HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and *ex vivo*. *Virus Res* **134**: 104–123.
Molla A, Korneyeve M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM *et al*. (1996). Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med* **2**: 760–766.
Parera M, Pérez-Álvarez N, Clotet B, Martínez MA (2009). Epistasis among deleterious mutations in the HIV-1 protease. *J Mol Biol* **392**: 243–250.
Pepin KM, Wichman HA (2007). Variable epistatic effects between mutations at host recognition sites in ϕX174. *Evolution* **67**: 1710–1724.
Pfaffl MV (2004). Quantification strategies in real-time PCR. In Bustin SA (ed *A-Z of Quantitative PCR, International University Line*. La Jolla: USA, pp 87–112.
Phillips PC (2008). Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**: 855–867.
Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**: 383–386.
Poelwijk FJ, Tanase-Nicola S, Kiviet DJ, Tans SJ (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J Theor Biol* **272**: 141–144.
Poon AFY, Chao L (2006). Functional origins of fitness effect-sizes of compensatory mutations in the DNA bacteriophage ϕX174. *Evolution* **60**: 2032–2043.
Proulx SR, Phillips PC (2005). The opportunity for canalization and the evolution of genetic networks. *Am Nat* **165**: 147–162.
Remold SK, Lenski RE (2004). Pervasive joint influence of epistasis and plasticity on mutational effects in Escherichia coli. *Nat Genet* **36**: 423 426.
Rice WR (1989). Analyzing tables of statistical tests. *Evolution* **43**: 223–225.
Rodrigo G, Carrera J, Ruiz-Ferrer V, Del Toro FJ, Llave C, Voinnet O *et al*. (2011). Characterization of the *Arabidopsis thaliana* interactome targeted by viruses. *Santa Fe Institute Working Paper* 11-10-049.
Rokyta DR, Joyce P, Caudle B, Miller C, Beisel CJ, Wichman HA (2011). Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS Genet* **7**: e1002075.
Salverda MLM, Dellus E, Gorter FA, Debets AJM, Van der Oost J, Hoekstra RF *et al*. (2011). Initial mutations direct alternative pathways of protein evolution. *PLoS Genet* **7**: e1001321.
Sanjuán R (2006). Quantifying antagonistic epistasis in a multifunctional RNA secondary structure of the *Rous sarcoma virus*. *J Gen Virol* **87**: 1595–1602.
Sanjuán R, Elena SF (2006). Epistasis correlates to genomic complexity. *Proc Natl Acad Sci USA* **103**: 14402–14405.
Sanjuán R, Forment J, Elena SF (2006). *In silico* predicted robustness of viroids RNA secondary structure. II. Interaction between mutation pairs. *Mol Biol Evol* **23**: 2123–2130.
Sanjuán R, Moya A, Elena SF (2004). The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc Natl Acad Sci USA* **101**: 15376–15379.
Sanjuán R, Nebot MR (2008). A network model for the correlation between epistasis and genomic complexity. *PLoS ONE* **3**: e2663.
Schrag SJ, Perrot V, Levin BR (1997). Adaptation to the fitness cost of antibiotic resistance in *E. coli*. *Proc R Soc B* **264**: 1287–1291.
Van Opijnen T, Boerlijst MC, Berkhout B (2006). Effects of random mutations in the *Human immunodeficiency virus* type 1 transcriptional promoter on viral fitness in different host cell environments. *J Virol* **80**: 6678–6685.
Weinreich DM (2005). The rank ordering of genotypic fitness values predicts genetic constraints on natural selection on landscapes lacking sign epistasis. *Genetics* **171**: 1397–1405.
Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**: 111–114.
Weinreich DM, Watson RA, Chao L (2005). Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**: 1165–1174.
Welch JJ, Waxman D (2005). The *nk* model and population genetics. *J Theor Biol* **234**: 329–340.
Wilke CO, Adami C (2001). Interaction between directional epistasis and average mutational effects. *Proc R Soc B* **298**: 1469–1474.
Wilke CO, Lenski RE, Adami C (2003). Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evol Biol* **3**: 1–14.
Withlock MC, Phillips PC, Moore FBG, Tonsor SJ (1995). Multiple fitness peaks and epistasis. *Annu Rev Ecol Evol Syst* **26**: 601–629.
You L, Yin J (2002). Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage T7. *Genetics* **160**: 1273–1281.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)

Supplementary Table 1. TEV single mutant genotypes used in this study and some of their properties.

Genotype	*Protein*	*Location*	*Nucleotide substitution*	*Amino acid change*	*Polarity change*	*Fitness (±1 SEM)*
DQ986288, wild-type isolate						1.3461±0.0118
PC2	P1	158	U→G	F→C	apolar→polar	1.3310±0.0101
PC6	P1	375	A→G	L→M		1.3477±0.0100
PC7	P1	475	A→C	K→Q	basic→polar	1.3198±0.0088
PC12	P1	872	A→C	M→L		1.3371±0.0084
PC19	HC-Pro	1503	A→G	synonymous		1.3308±0.0049
PC22	HC-Pro	1655	A→G	N→S		1.2795±0.0081
PC26	HC-Pro	2119	A→U	synonymous		1.2586±0.0115
PC40	P3	3238	T→C	synonymous		1.3291±0.0150
PC41	P3	3406	C→A	Q→K	polar→basic	1.3100±0.0077
PC44	P3	3468	U→G	synonymous		1.3237±0.0099
PC49	CI	4418	G→C	S→T		1.3164±0.0065
PC60	CI	5349	U→C	synonymous		1.3200±0.0108
PC63	6K2	5582	A→G	K→R		1.3205±0.0092
PC67	NIa-VPg	6012	U→G	I→M		1.3327±0.0093
PC69	NIa-VPg	6044	C→A	T→N		1.3156±0.0092
PC70	NIa-VPg	6197	U→G	M→R	apolar→basic	1.3268±0.0084
PC72	NIa-VPg	6251	U→C	F→S	apolar→polar	1.3359±0.0072
PC76	NIa-Pro	6519	U→C	synonymous		1.3392±0.0059
PC83	NIb	7315	A→G	I→V		1.3371±0.0099
PC95	NIb	8501	A→C	E→A	acid→polar	1.3306±0.0051

Supplementary Table 2. TEV double mutant genotypes synthesized for this study and some of their properties. Epistasis values marked with an asterisk were significant (*t*-test, $P \leq 0.0491$ in all cases).

Mutation 1	*Mutation 2*	*Fitness (±1 SEM)*	*Epistasis (±1 SEM)*
PC2	PC69	1.3445±0.0089	0.0587±0.0533
PC6	PC7	1.3889±0.0035	0.0908±0.0461
PC6	PC19	1.3842±0.0047	0.0694±0.0424
PC6	PC49	1.3813±0.0065	0.0851±0.0468
PC6	PC63	1.3764±0.0063	0.0730±0.0485
PC6	PC69	1.3904±0.0022	0.0985±0.0448
PC6	PC76	0	−1.7136±0.0183*
PC7	PC19	1.3802±0.0040	0.1014±0.0399*
PC7	PC63	1.3843±0.0026	0.1205±0.0420*
PC7	PC69	1.3806±0.0033	0.1141±0.0418*
PC12	PC19	1.3740±0.0055	0.0699±0.0413
PC12	PC83	1.3698±0.0056	0.0562±0.0481
PC12	PC95	1.3553±0.0039	0.0451±0.0392
PC19	PC40	1.3817±0.0033	0.0909±0.0471
PC19	PC41	1.3781±0.0029	0.1117±0.0369*
PC19	PC69	1.3856±0.0051	0.1219±0.0444*
PC19	PC70	1.3810±0.0032	0.0930±0.0383
PC19	PC83	1.3836±0.0010	0.0828±0.0372
PC19	PC95	1.3786±0.0000	0.0850±0.0294*
PC22	PC63	0	−1.7533±0.0245*
PC22	PC69	1.2288±0.0306	0.0179±0.0635
PC22	PC72	1.2831±0.0210	−0.0293±0.0780
PC22	PC76	0	−1.7791±0.0198*
PC22	PC95	0	−1.8050±0.0212*
PC26	PC63	1.3310±0.0052	0.1296±0.0478*
PC26	PC69	1.3122±0.0003	0.1103±0.0426
PC26	PC76	0	−1.6896±0.0208*

PC40	PC63	0	−1.6857±0.0228*
PC40	PC69	1.3538±0.0051	0.0737±0.0548
PC40	PC83	1.4027±0.0000	0.0814±0.0436
PC41	PC49	1.3875±0.0040	0.1432±0.0404*
PC41	PC83	1.3811±0.0054	0.1077±0.0469
PC44	PC49	1.3835±0.0064	0.1199±0.0453*
PC44	PC63	1.3454±0.0094	0.0632±0.0509
PC44	PC69	1.3497±0.0084	0.0755±0.0510
PC44	PC76	1.3762±0.0114	0.0821±0.0528
PC49	PC67	1.3840±0.0051	0.1085±0.0441*
PC49	PC70	1.3851±0.0035	0.1177±0.0406*
PC49	PC83	1.3682±0.0044	0.1111±0.0496
PC49	PC95	1.3547±0.0072	0.0720±0.0409
PC60	PC83	1.3534±0.0238	0.0568±0.0754
PC60	PC95	1.3392±0.0149	0.0465±0.0567
PC63	PC69	1.3603±0.0215	0.0938±0.0676
PC63	PC70	1.3814±0.0057	0.1073±0.0456
PC63	PC95	1.3577±0.0058	0.0704±0.0411
PC67	PC69	0	−1.7551±0.0303*
PC67	PC76	1.3564±0.0095	0.0408±0.0490
PC69	PC76	0	−1.7024±0.0172*
PC69	PC95	1.3886±0.0087	0.1001±0.0466
PC70	PC83	1.3887±0.0039	0.0952±0.0459
PC72	PC83	1.3881±0.0101	0.0798±0.0514
PC76	PC95	1.3522±0.0227	0.0381±0.0609
PC83	PC95	0	−1.7620±0.0200*

CHAPTER V - *G*×*G*×*E* INTERACTIONS

Downloaded from rsbl.royalsocietypublishing.org on July 28, 2012

biology letters

Biol. Lett.
doi:10.1098/rsbl.2012.0396
Published online

Evolutionary biology

Epistasis between mutations is host-dependent for an RNA virus

Jasna Lalić[1] and Santiago F. Elena[1,2,*]

[1]*Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, 46022 València, Spain*
[2]*Santa Fe Institute, Santa Fe NM 87501, USA*
**Author for correspondence (santiago.elena@csic.es).*

How, and to what extent, does the environment influence the way mutations interact? Do environmental changes affect both the sign and the magnitude of epistasis? Are there any correlations between environments in the variability, sign or magnitude of epistasis? Very few studies have tackled these questions. Here, we addressed them in the context of viral emergence. Most emerging viruses are RNA viruses with small genomes, overlapping reading frames and multifunctional proteins for which epistasis is abundant. Understanding the effect of host species in the sign and magnitude of epistasis will provide insights into the evolutionary ecology of infectious diseases and the predictability of viral emergence.

Keywords: emerging viruses; deleterious mutations; epistasis; genotype-by-environment; virus evolution

1. INTRODUCTION

The large majority of emerging viruses are RNA viruses [1]. However, their compact genomes comprising overlapping reading frames and multifunctional proteins and their high mutation rates may impose severe adaptive constraints [2]. Understanding the mechanistic basis of these constraints is central to explaining why some RNA viruses are more able than others to cross species boundaries. Epistasis is thought to be important in the evolution of host range [3,4]. Moreover, it has been suggested that the sign of epistasis depends on environmental severity, switching from positive to negative as environments become stressful [5]. Yet, few studies have empirically examined this possibility.

To evaluate the effect that different hosts exert on the distribution of epistatic interactions, we tested the fitness of *Tobacco etch virus* (TEV) genotypes carrying two single-nucleotide substitutions, whose independent effects were previously evaluated [6], across susceptible hosts of increasing genetic divergence from the primary host. TEV naturally infects *Solanaceae* plants, and the strain used here was isolated from *Nicotiana tabacum* [7]. Previously, we have shown that the deleterious effects of mutations were stronger as the host (i.e. the virus's environment, E) was more genetically diverged from tobacco, and the proportion of lethal, deleterious, neutral and beneficial mutations was also altered [6]. We also found that this host dependence (i.e. plasticity or $G \times E$) had two origins: antagonistic pleiotropy and changes in genetic variance for fitness across hosts [6]. Furthermore, we recently found that the fitness effect of a given mutation depended on the genetic background where it was evaluated (i.e. epistasis or $G \times G$) [8]. Variation was observed both in the sign and the strength of epistasis, being negative on average and with abundant cases of reciprocal sign epistasis [8]. If $G \times E$ and $G \times G$ play major roles in determining TEV fitness, it is logical to expect that epistasis may also vary depending on environmental severity [9], that is, a $G \times G \times E$ component may exist. Quantifying the extent to which $G \times G \times E$ determines viral fitness is central to predicting the fate of viral genotypes across hosts and, ultimately, the likelihood that viruses will cross host species barriers. Epistatic interactions allowing RNA viruses to infect new hosts have been widely observed. For example, interactions between five amino acids in the coat protein of *Pelargonium flower break virus* are necessary for improving fitness in the new host *Chenopodium quinoa* [10]. Similarly, the ability of *Potato virus Y* to infect resistant pepper plants depends both on the alleles at the VPg and at the CI genes [11].

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsbl.2012.0396 or via http://rsbl.royalsocietypublishing.org.

One contribution to a special feature on 'Experimental evolution' organized by Paul Sniegowski, Thomas Bataillon and Paul Joyce.

2. MATERIAL AND METHODS

(a) *Virus genotypes*

The 10 double mutants ([8]; electronic supplementary material, table S2) were generated by randomly combining pairs of 12 single mutations ([12]; electronic supplementary material, table S1). The particular 10 double mutants generated were randomly chosen. Mutant genotypes were generated by site-directed mutagenesis of plasmid pMTEV [7] using QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). Infectious RNAs were obtained by *in vitro* transcription of the corresponding plasmids [13].

(b) *Host species and inoculation experiments*

Nicotiana tabacum and *Datura stramonium* are *Solanaceae*. *Helianthus annuus* is an *Asteraceae*. *Solanaceae* and *Asteraceae* are Asterids [14]. *Spinacea oleracea* is an *Amaranthaceae*. All families are eudicots [14].

All plants were inoculated in a single block and at similar developmental stages. Nine plants per host per virus genotype were rub-inoculated at the first true leaf with 5 μg of RNA of each genotype and 10 per cent carborundum. *Solanaceae* hosts show symptoms when infected; non-*Solanaceae* hosts do not, and infections were confirmed by RT-PCR [15]. Ten days post-inoculation (dpi), the whole infected plant, except the inoculated leaf, was collected. Tissue was frozen in liquid nitrogen and ground.

(c) *RNA purification, virus fitness and epistasis estimation*

Viral RNA was purified as described elsewhere [6]. Total plant RNA concentration was measured spectrophotometrically and the samples were diluted to a final concentration of 50 ng μl^{-1}. Within-plant virus accumulation was measured by absolute RT-qPCR [6].

For each genotype, a Malthusian growth rate per day was computed as $m = 1/t \log Q_t$, where Q_t is the number of pg of TEV RNA per 100 ng of total plant RNA quantified at $t = 10$ dpi. Absolute fitness was defined as $W = e^m$ (electronic supplementary material, table S1).

Epistasis between mutations x and y was calculated as $\varepsilon_{xy} = W_{00}W_{xy} - W_{x0}W_{0y}$, where W_{00}, W_{xy}, W_{x0}, W_{0y} stand for the fitness of wild-type, double and single mutants, respectively (electronic supplementary material, table S2). Qualitatively identical results are obtained using the scaled epistasis [16].

3. RESULTS

First, we sought to determine whether the number of epistatic pairs was affected by the host species. Table 1 shows the pairs of mutations evaluated on each host classified as: (i) independent effects $\varepsilon_{xy} = 0$, (ii) positive epistasis, and (iii) negative interactions (for each host, one-sample t-tests controlling for

Received 25 April 2012
Accepted 21 June 2012

This journal is © 2012 The Royal Society

Downloaded from rsbl.royalsocietypublishing.org on July 28, 2012

Table 1. Epistasis of double mutants in each host. Average epistasis was computed after excluding lethal combinations. Sign epistasis refers to cases in which the sign of the fitness effect depends on the genetic background. Reciprocal (recip.) sign epistasis means that the sign of the fitness effect of a mutation is conditional upon the state of another locus and *vice versa*. Last row shows the significance test for the average epistasis. Red numbers indicate significant changes in epistasis from the primary host (*N. tabacum*) to alternative ones (paired *t*-tests corrected for multiple comparisons; figure 1). Errors represent ±1 s.e.m.

genotype	*N. tabacum*	*D. stramonium*	*H. annuus*	*S. oleracea*	average epistasis
PC6/PC63	0.0730	1.5520[a]	−0.0725	−0.0828	0.3674 ± 0.3965
PC6/PC76	−1.8050[a] (sign)	−0.6233 (sign)	−0.1178	−0.0055	−0.6379 ± 0.4116
PC19/PC41	0.1117[a] (recip. sign)	0	−0.0245	−0.0263	0.0152 ± 0.0327
PC22/PC69	−0.0293	−1.7129[a]	−0.2147	−0.2106	−0.5419 ± 0.3927
PC22/PC72	0.0179	−0.3213[a]	−0.2172	−0.1414	−0.1633 ± 0.0698
PC22/PC95	−1.7024[a]	−0.4537	−0.1855	−0.1474	−0.6222 ± 0.3665
PC40/PC83	0.1111	−0.2108	−0.0829	−0.0535	−0.0590 ± 0.0662
PC67/PC76	0.0408	−0.5341[a] (sign)	−1.0253[a] (recip. sign)	0.1158	−0.3507 ± 0.2677
PC69/PC76	−1.7620[a]	−0.5057[a] (sign)	−0.1112	0.0221	−0.5892 ± 0.4067
PC76/PC95	0.0381	−0.5955[a]	0.0127	0.0496	−0.1238 ± 0.1574
average epistasis	0.0519 ± 0.0193	−0.2834 ± 0.3187	−0.2185 ± 0.1043	−0.0480 ± 0.0316	
t-test (9 d.f.)	0.0358	0.4034	0.0695	0.1630	

[a]Epistasis significantly departs from zero within the host.

multiple comparisons). The distribution of counts for these three categories differs among hosts ($\chi^2 = 14.157$, 6 d.f., $p = 0.028$), with the difference being driven by an excess of non-epistatic cases in the non-*Solanaceae* (table 1). The difference is further enhanced if counts are pooled together for *Solanaceae* and non-*Solanaceae* (Fisher's exact tests, $p = 0.003$). However, this classification into multiplicative versus epistatic pairs has to be taken with caution since a weak yet significant negative correlation exists between the absolute value of ε_{xy} and its error (see electronic supplementary material, table S2; partial correlation coefficient controlling for host: $r = -0.282$, 37 d.f., 1-tailed $p = 0.041$), suggesting that the smaller the ε_{xy}, the larger its uncertainty, resulting in less power to reject the null hypothesis of independent effects.

The above classification is just one of several possible. An alternative classification distinguishes between magnitude and sign epistasis. For magnitude epistasis, the fitness value associated with a mutation, but not its sign, changes upon the genetic background [17]. For sign epistasis, the sign of the fitness effect itself is under epistatic control [17]. Table 1 indicates which pairs match these categories. For pairs involved in significant sign epistasis, those of reciprocal type (i.e. the sign of the fitness effects change for both mutations) are also indicated. A significant difference among hosts holds if mutations are sorted according to this classification ($\chi^2 = 14.927$, 6 d.f., $p = 0.021$; *Solanaceae* versus non-*Solanaceae*: Fisher's exact test, $p = 0.004$). With this classification scheme, the excess of independent fitness effects for non-*Solanaceae* also drives the difference among hosts. From these analyses, we can conclude that the host species has an effect on the number of epistatic interactions in TEV, with the number of independent fitness effects being significantly larger in hosts distantly related to the primary host.

Next, we identified the effect of hosts on epistasis for each pair of mutations. Figure 1 shows the change in ε_{xy} from *N. tabacum* to alternative hosts. A horizontal line means that epistasis among a pair of mutations is host-independent. Lines with positive or negative slopes indicate host-dependent epistasis. In *D. stramonium* (figure 1*a*), epistasis became more negative in one case, less negative in three, more positive in one and less positive in four instances. In *H. annuus* (figure 1*b*), one case was significantly more negative than in tobacco and the less negative cases were the same as in *D. stramonium*. Finally, for *S. oleracea*, significant changes were detected only for the same three pervasive genotypes (table 1). Interestingly, pairs PC6/PC76, PC22/PC95 and PC69/PC76, each of which carries viable mutations when tested individually in *N. tabacum*, are not viable in this host when combined. This synthetic lethality (SL) is an extreme case of negative epistasis. However, these three genotypes are viable in the alternative hosts. By contrast, genotypes PC22/PC69 and PC67/PC76 represent cases of SL only in *D. stramonium* and *H. annuus*, respectively. These observations indicate that SL is also host-dependent. In all these cases, mutations affect different proteins (see electronic supplementary material, table S1). PC19, affecting HC-Pro, was previously described as lethal in *D. stramonium* [6], and the same lethal phenotype was observed for PC19/PC41. Conversely, PC63, affecting 6K2, also previously described as lethal in this host [6], is compensated by PC6 in protein P1, rendering a viable PC6/PC63.

When SLs are included, no host departed from the expectation of independent effects (table 1, one-sample *t*-tests; $p \geq 0.052$), although significant differences among hosts exist ($F_{3,177} = 33.660$, $p < 0.001$). Since SLs are irrelevant in terms of evolutionary dynamics, we re-evaluated average epistasis after removing them. In this case, the average ε_{xy} becomes significantly positive in *N. tabacum* ($p = 0.036$) but remains non-significant in the alternative hosts ($p \geq 0.070$). Therefore, we conclude that the intensity of epistasis decreases as the genetic divergence between the primary host and alternative hosts increases. However, this trend may be a spurious consequence of our reduced statistical power to detect small epistasis values.

Downloaded from rsbl.royalsocietypublishing.org on July 28, 2012

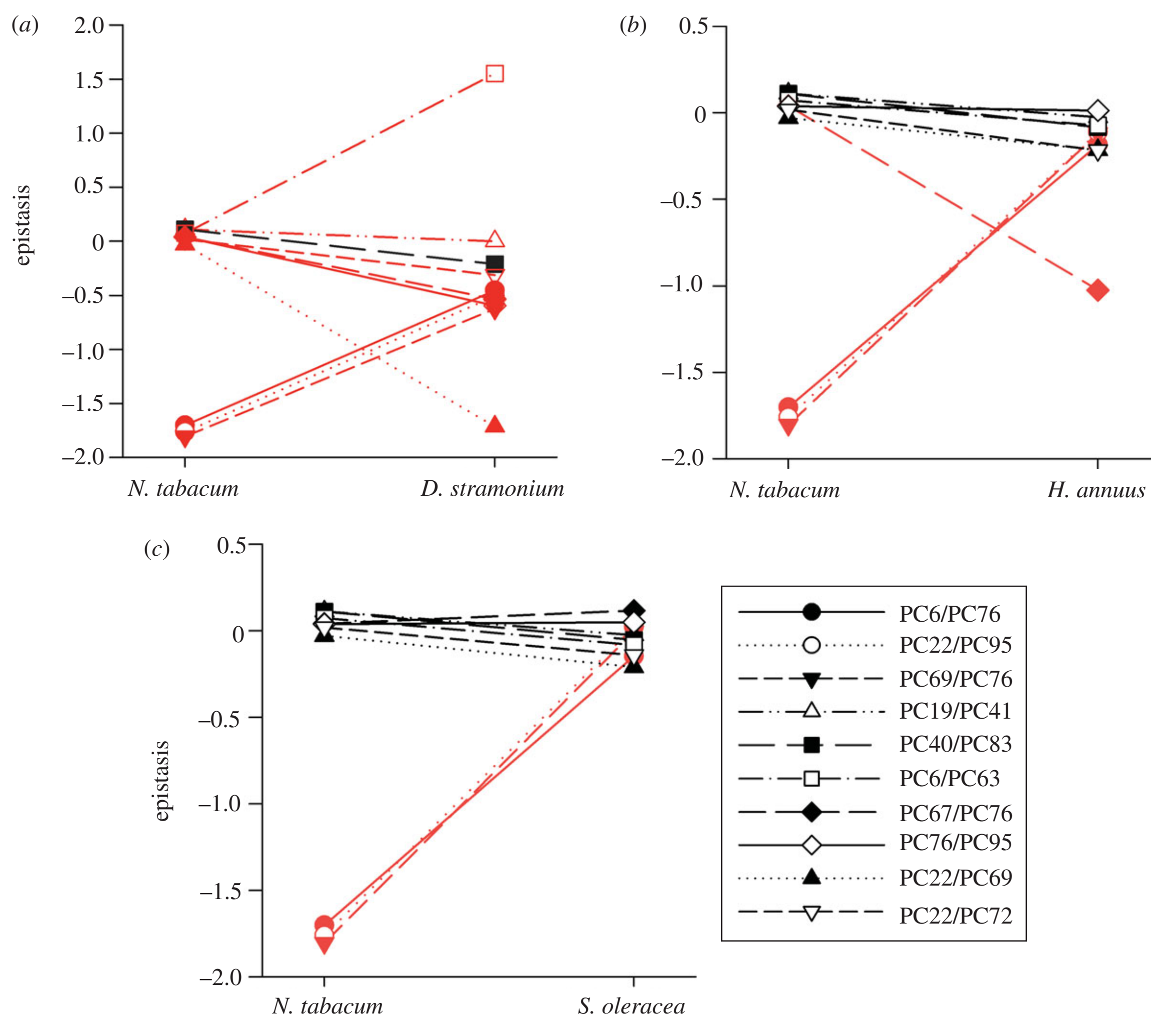


Figure 1. Changes in epistasis from the primary host to alternative hosts ((*a*) *D. stramonium*, (*b*) *H. annuus* and (*c*) *S. oleracea*). Significant differences are indicated in red (paired *t*-tests corrected for multiple comparisons).

The last column of table 1 shows epistasis for each double mutant averaged across hosts. A significant overall genotype effect exists ($F_{9,177} = 168.593$, $p < 0.001$), with epistasis ranging from negative to weakly positive. A significant genotype-by-host effect has been detected ($F_{27,177} = 1.55 \times 10^5$, $p < 0.001$), providing support for the importance of $G \times G \times E$ in the architecture of viral fitness.

This ANOVA treated epistasis values as independent observations. However, this raises two statistical concerns: (i) the same mutations are involved in multiple pairs and (ii) the fitness of the wild-type on a host (W_{00}) has been used to compute ε_{xy} for each genotype in this host. We circumvented these problems as follows: (i) the effect of using the same mutation on different combinations was removed by running the analyses for each genotype independently and making inferences valid only for each individual genotype and (ii) the non-independence introduced by re-using W_{00} was minimized using a bootstrap approach. The results from these extra analyses (see electronic supplementary material, table S3) confirm the significant $G \times G \times E$.

4. DISCUSSION

Our experiments show that the fitness value of a given mutation depends on the genotypic background wherein it appears and on the infected host. This observation has implications for predicting the fate of viral genotypes under different and variable environments and, consequently, for the development of successful antiviral strategies based on the use of attenuated vaccines. We stress the importance of evaluating candidate attenuating mutations in multiple genetic backgrounds and across the widest possible panel of hosts, especially in close relatives to the ones for which the vaccine is intended. Otherwise, attenuating mutations may be easily compensated by second-site changes that are viable, or even beneficial, in alternative hosts.

Our results indicate that host effects on epistasis are modulated by the degree of genetic divergence between the primary and alternative hosts. It was previously shown that point mutations had more deleterious effects as the genetic divergence from the primary host increased [6]. This observation agreed with the results of a simulation study of phage T7 showing that mutations were more severe in poor environments and milder in rich ones [5]. Furthermore, mild mutations showed negative epistasis in poor environments but weak positive epistasis in rich ones, while severe mutations showed either no epistasis or weak positive epistasis in poor environments and positive epistasis in rich ones [5]. We have shown here that epistasis was positive in the primary host (after removing SLs) but switched to no epistasis in

Downloaded from rsbl.royalsocietypublishing.org on July 28, 2012

other hosts. Together, these observations suggest that *N. tabacum* (and to a minor extent *D. stramonium*) represent rich environments for TEV, while the alternative 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.

We thank Francisca de la Iglesia and Àngels Pròsper for technical assistance, Mark Zwart for help with R, Stéphanie Bedhomme and Susanna Remold for fruitful discussion. This research was supported by Ministerio de Ciencia e Innovación grant BFU2009-06993 (S.F.E.) and by CSIC JAE-Pre program (J.L.).

1 Holmes, E. C. 2009 The evolutionary genetics of emerging viruses. *Annu. Rev. Ecol. Evol. Syst.* **40**, 353–372. (doi:10.1146/annurev.ecolsys.110308.120248)
2 Holmes, E. C. 2003 Error threshold and the constraints to RNA virus evolution. *Trends Microbiol.* **11**, 543–546. (doi:10.1016/j.tim.2003.10.006)
3 Holmes, E. C. & Rambaut, A. 2004 Viral evolution and the emergence of SARS coronavirus. *Phil. Trans. R. Soc. Lond. B* **359**, 1059–1065. (doi:10.1098/rstb.2004.1478)
4 Pepin, K. M. & Wichman, H. A. 2007 Variable epistatic effects between mutations at host recognition sites in ϕX174 bacteriophage. *Evolution* **67**, 1710–1724. (doi:10.1111/j.1558-5646.2007.00143.x)
5 You, L. & Yin, J. 2002 Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage T7. *Genetics* **160**, 1273–1281.
6 Lalić, J., Cuevas, J. M. & Elena, S. F. 2011 Effect of host species on the distribution of mutational effects for an RNA virus. *PLoS Genet.* **7**, e1002378. (doi:10.1371/journal.pgen.1002378)
7 Bedoya, L. C. & Daròs, J. A. 2010 Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* **149**, 234–240. (doi:10.1016/j.virusres.2010.02.004)
8 Lalić, J. & Elena, S. F. In press. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. *Heredity.* (doi:10.1038/hdy.2012.15)
9 Agrawal, A. F. & Withlock, M. C. 2010 Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? *Trends Ecol. Evol.* **25**, 450–458. (doi:10.1016/j.tree.2010.05.003)
10 Rico, P., Ivars, P., Elena, S. F. & Hernández, C. 2006 Insights into the selective pressures restricting *Pelargonium flower break virus* genome variability: evidence for host adaptation. *J. Virol.* **80**, 8124–8132. (doi:10.1128/JVI.00603-06)
11 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. *Mol. Plant Pathol.* **12**, 911–920. (doi:10.1111/j.1364-3703.2011.00724.x)
12 Carrasco, P., de la Iglesia, F. & Elena, S. F. 2007 Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J. Virol.* **81**, 12 979–12 984. (doi:10.1128/JVI.00524-07)
13 Carrasco, P., Daròs, J. A., Agudelo-Romero, P. & Elena, S. F. 2007 A real-time RT-PCR assay for quantifying the fitness of *Tobacco etch virus* in competition experiments. *J. Virol. Meth.* **139**, 181–188. (doi:10.1016/j.jviromet.2006.09.020)
14 Soltis, D. E. & Soltis, P. S. 2000 Contributions of plant molecular systematics to studies of molecular evolution. *Plant Mol. Biol.* **42**, 45–75. (doi:10.1023/A:1006371803911)
15 Lalić, J., Agudelo-Romero, P., Carrasco, P. & Elena, S. F. 2010 Adaptation of *Tobacco etch potyvirus* to a susceptible ecotype of *Arabidopsis thaliana* capacitates it for systemic infection of resistant ecotypes. *Phil. Trans. R. Soc. B* **65**, 1997–2008. (doi:10.1098/rstb.2010.0044)
16 Segrè, D., DeLuna, A., Church, G. M. & Kishony, R. 2005 Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83. (doi:10.1038/ng1489)
17 Poelwijk, F. J., Tanase-Nicola, S., Kiviet, D. J. & Tans, S. J. 2011 Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–144. (doi:10.1016/j.jtbi.2010.12.015)

Table S1. TEV single mutant genotypes used in this study, some of their properties and their absolute fitness estimated in the four experimental hosts.

Genotype	Protein	Location	Nucleotide substitution	Amino acid change	Polarity change	Fitness ± 1 SEM			
						N. tabacum	***D. stramonium***	***H. annuum***	***S. oleracea***
DQ986288, wild-type isolate						1.3461±0.0118	1.4017±0.0030	0.9957±0.0145	0.9465±0.0154
PC6	P1	375	A→G	L→M		1.3478±0.0099	1.3981±0.0160	1.0329±0.0353	0.9743±0.0147
PC19	HC-Pro	1503	A→G	synonymous		1.3099±0.0078	0	1.0503±0.0112	0.9605±0.0121
PC22	HC-Pro	1655	A→G	N→S		1.2795±0.0081	1.2408±0.0108	1.1365±0.0301	1.1370±0.0687
PC40	P3	3238	T→C	synonymous		1.3291±0.0150	1.2380±0.1557	1.0199±0.0259	0.9670±0.0204
PC41	P3	3406	C→A	Q→K	polar→basic	1.3309±0.0049	1.3778±0.0083	1.0176±0.0186	0.9618±0.0141
PC63	6K2	5582	A→G	K→R		1.3205±0.0079	0	0.9997±0.0206	0.9551±0.0217
PC67	VPg	6012	U→G	I→M		1.3327±0.0093	1.4023±0.0017	1.0146±0.0155	0.9602±0.0210
PC69	VPg	6044	C→A	T→N		1.3156±0.0092	1.3805±0.0103	1.0259±0.0656	0.9997±0.0217
PC72	VPg	6251	U→C	F→S	apolar→polar	1.3359±0.0072	1.4063±0.0169	1.0202±0.0223	0.9438±0.0248
PC76	NIa-Pro	6519	U→C	synonymous		1.3306±0.0050	1.4028±0.0000	0.9936±0.0264	0.9964±0.0139
PC83	NIb	7315	A→G	I→V		1.3371±0.0099	1.3616±0.0093	1.0399±0.0328	1.0798±0.0658
PC95	NIb	8501	A→C	E→A	acid→polar	1.3306±0.0050	1.3923±0.0081	0.9936±0.0264	0.9964±0.0139

Table S2. TEV double mutant genotypes used in this study, their absolute fitness and epistasis coefficient measured in the four experimental hosts. Errors represent ± 1 SEM. Significant epistatic interactions are marked with asterisks.

	N. tabacum		*D. stramonium*		*H. annuus*		*S. oleracea*	
	Fitness	**Epistasis**	**Fitness**	**Epistasis**	**Fitness**	**Epistasis**	**Fitness**	**Epistasis**
PC6/PC63	1.3765±0.0063	0.0730±0.0485	1.1073±0.0432	1.5520±0.0639*	0.9643±0.0392	-0.0725±0.1096	0.8954±0.0055	-0.0828±0.0542
PC6/PC76	0	-1.8050±0.0212*	0.9441±0.0208	-0.6233±0.0655	0.9301±0.0148	-0.1178±0.0917	0.9703±0.0328	-0.0055±0.0903
PC19/PC41	1.3781±0.0029	0.1117±0.0369*	0	0	1.0489±0.0000	-0.0245±0.0462	0.9482±0.0000	-0.0263±0.0398
PC22/PC69	1.2288±0.0306	-0.0293±0.0780	0	-1.7129±0.0278*	0.9554±0.0069	-0.2147±0.1262	0.9784±0.0352	-0.2106±0.1418
PC22/PC72	1.2831±0.0211	0.0179±0.0635	1.0221±0.0060	-0.3123±0.0476*	0.9463±0.0026	-0.2172±0.0724	0.9843±0.0139	-0.1414±0.1214
PC22/PC95	0	-1.7024±0.0172*	0.9182±0.0476	-0.4537±0.0847	0.9478±0.0159	-0.1855±0.0896	1.0412±0.0462	-0.1474±0.1441
PC40/PC83	1.4028±0.0000	0.1111±0.0496	1.0523±0.0000	-0.2108±0.2267	0.9818±0.0008	-0.0829±0.0754	1.0468±0.0080	-0.0535±0.1095
PC67/PC76	1.3564±0.0095	0.0408±0.0490	1.0119±0.0241	-0.5341±0.0506*	0	-1.0253±0.0429*	1.0844±0.0386	0.1158±0.1031
PC69/PC76	0	-1.7620±0.0200*	1.0105±0.0018	-0.5057±0.0312*	0.9296±0.0000	-0.1112±0.1074	1.0249±0.0000	0.0221±0.0675
PC76/PC95	1.3522±0.0227	0.0381±0.0609	0.9687±0.0000	-0.5955±0.0143*	1.0212±0.0577	0.0127±0.1257	1.0506±0.0204	0.0496±0.0797

Table S3. Bootstrap one-way ANOVAs for the effect of host species on epistasis. Fitness values of single mutants W_{x0} and W_{0y}, of the wild-type W_{00} and of the double mutant W_{xy} are sampled from the experimental replicates to compute pseudo-values of ε_{xy}. The number of such pseudo-values equals the actual experimental sample size for each host. Then, a one-way ANOVA testing for differences in epistasis among hosts was computed for this pseudo-sample and the associated significance level P was recorded. This procedure was repeated 10000 times to estimate the median P and to construct 95% CI for P. These confidence intervals represent a measure of the statistical power associated to each test. Using Fisher's combined probability test of the same hypothesis, we can summarize the results into a single test and conclude that an overall significant host effect on epistasis exists (χ^2 = 61.238, 20 d.f., $P < 0.001$).

Genotype	**Lower *P***	**Median *P***	**Upper *P***
PC6/PC76	2.207×10^{-12}	1.299×10^{-8}	1.176×10^{-4}
PC76/PC95	3.878×10^{-11}	1.758×10^{-7}	1.251×10^{-3}
PC22/PC95	1.315×10^{-10}	7.428×10^{-7}	4.734×10^{-3}
PC69/PC76	2.363×10^{-10}	1.489×10^{-6}	3.472×10^{-3}
PC19/PC41	8.496×10^{-7}	3.215×10^{-4}	0.0298
PC22/PC72	5.632×10^{-4}	0.1443	0.9377
PC6/PC63	0.1469	0.3629	0.7913
PC40/PC83	0.0089	0.3949	0.9654
PC67/PC76	0.0838	0.6611	0.9858
PC22/PC69	0.3104	0.7512	0.9876

CHAPTER VI - Empirical fitness landscapes

Empirical Fitness Landscape for the Adaptation of an RNA Virus to *Arabidopsis thaliana*

Jasna Lalić[1] and Santiago F. Elena[1,2]

[1]*Instituto de Biología Molecular y Celular de Planta (CSIC-UPV), València, Spain and* [2]*The Santa Fe Institute, New Mexico, USA*.

RNA viruses are of considerable concern in evolution because their high mutation rates, great population sizes and short generation times allow them to rapidly explore fitness landscapes and consequently, emerge and adapt to new hosts. Here we partially reconstructed the fitness landscape describing the adaptation of *Tobacco etch potyvirus* (TEV) to *Arabidopsis thaliana*. The adapted virus differed from the wildtype in five single-nucleotide substitutions. We reconstructed the majority of intermediate genotypes of adaptation and measured two fitness components: the infectivity and within host virus growth rate relative to the wildtype, the later being considered as an approximate to relative fitness. We found prevailing epistatic effects between mutations in the first three steps of adaptation, while independent fitness effects became more common with the increase in the number of fixed mutations. Epistatic interactions included both positive and negative deviations from the expected values. Furthermore, we characterized the landscape topography approximated by the number of the plants infected by a particular virus genotype and found that more than a half of possible evolutionary paths were neutral and the landscape was single peaked. The smoothness of TEV landscape in *A. thaliana* suggests that virus easily adapted to its new host. This is an important result from the perspective of emerging viral diseases and in concordance to common observation that majority of emerging viruses are indeed RNA viruses.

Evolutionary dynamics of adaptation is a complex process depending on the complexity of the organism, as well as on many genetic, developmental, ecological and environmental factors. Sewall Wright's metaphor of the fitness landscape (Wright 1932) pictures the process of adaptation as a surface in a multidimensional space that represents the relationship between mean fitness and the frequency of alleles within a population. Epistasis determines the topography of adaptive landscapes (Wright 1932; Whitlock *et al*. 1995; Poelwijk *et al.* 2011) as well as the accessibility of adaptive pathways throughout the landscape, thus, is central to understanding the course of evolution (Weinreich, 2005; Welch & Waxman, 2005; Franke *et al.* 2011). In absence of epistasis or in the case of magnitude epistasis, mutations give rise to either zero, positive or a negative fitness effect, regardless of the genetic background. This results in adaptive landscapes that are smooth and single peaked. In a smooth fitness landscape the evolution will always proceed uphill towards the single global optimum. In the case of sign epistasis, the sign of the fitness effect of a mutation depends on the genetic background, such that only a fraction of the total paths to the optimum are selectively accessible, *i.e.*, contain only steps that confer a performance increase. Reciprocal sign epistasis, in which two mutations are individually deleterious but jointly advantageous, gives rise to rugged landscape with multiple local optima (*i.e.,* peaks). The ruggedness of adaptive landscapes is critical to predict whether the evolving populations may reach the global optima or, by contrast, trough alternative evolutionary pathways, may get stuck into suboptimal fitness peaks (Whitlock *et al.* 1995, Weinreich 2005, Poelwijk *et al.* 2011, Kvitek & Sherlock, 2011, Poelwijk *et al.* 2007.)

To infer a fitness landscape empirically means to reconstruct all possible genotypic intermediates that led to adaptation to a new environment and measure their fitness. Genotypes should bear all possible combinations of mutations fixed by the adapted genotype. Till now, simple empirical fitness landscapes have been characterized for either a single gene or promotor in: bacteria (Weinreich *et al*. 2006; Lunzer *et al*. 2005; Poelwijk *et al*. 2007; Dawid *et al*. 2010; Chou *et al*. 2011), protozoa (Lozovsky *et al*. 2009), fungi (De Visser *et al*. 2009) and HIV-1 (Da Silva *et al.* 2010; Kouyos *et al.* 2012) being the only virus for which such landscape has been inferred. The majority of these studies are based on two genotypes: the ancestral and the one adapted to a given environment, differing mutually in a small set of known mutations. The largest empirical fitness landscape derived so far was for an evolved genotype that bared five mutations, that is, 32 genotypes. In all cases, the landscape is rugged and the number of accessible evolutionary pathways driving to the highest fitness peak is quite limited.

Here, we aimed to construct the first empirical adaptive landscape for an emerging plant RNA virus. Our work is based on previous experimental evolution of host switching. Recently, Agudelo-Romero *et al.* (2008) simulated the emergence of TEV in a population of a partially susceptible host, *A. thaliana* L*er*-0. *A. thaliana* belongs to family *Brassicaceae*, whereas TEV primary host species come from a completely distinct family: the *Solanaceae*. Evolution of TEV on *A. thaliana* was done by serial passaging of the virus within several independent evolution lineages. After 17 serial passages, the resulting evolved virus, hereafter named TEV-*At*17, showed ca. 10-fold higher infectivity, 2-logs greater viral load, increased virulence and more severe symptoms in the new host compared to the ancestral virus (Agudelo-Romero *et al.* 2008). Transcriptomic analyses have shown that the evolved virus was able of evading host's defense mechanisms by shutting down the expression of several defense pathways compared to its ancestor TEV. During adaptation TEV-*At*17 fixed five mutations among which, two were synonymous (Table 1). Each of the mutations was in different virus cistrons (Table 1).

Table 1. Mutations fixed in the evolved virus TEV-*At*17. The adaptiveness of two synonymous single-nucleotide substitutions is inferred form codon usage database: http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=3702 and denoted in the last two columns on the right.

						***A. thaliana* codon usage frequency (per thousand)**	
Locus name	**Mutation**	**Gene**	**Amino acid change**	**Wildtype codon**	**Mutated codon**	**Wildtype codon**	**Mutated codon**
1	U357C	P1	synonymous (119H)	CAU	CAC	13,8	8,7
2	C3140U	P3	A1047V	GCG	GUG		
3	C3629U	6K1	T1210M	ACG	AUG		
4	C6037U	VPg	L2013F	CUU	UUU		
5	C6906U	NIa-Pro	synonymous (2302F)	UUC	UUU	20,7	21,8

In order to picture the accessible evolutionary pathway(s) for TEV-*At*17 in *A. thaliana* L*er*-0, we constructed all experimentally possible intermediate genotypes of adaptation and quantifyied the fitness and the infectivity of each viral genotype in *A thaliana* L*er*-0. The set of genotypes consisted of all $2^5 = 32$ combinations of the five mutations fixed by TEV-*At*17 during its adaptation to the new host. The parental class consisted of only one genotype: the wildtype TEV, following by a group of five simple mutants, ten both double and triple mutants, five quadruple mutants and one quintuple (Figure 1). Our dataset consisted of 22 genotypes in total; missing four double mutants, three triple mutants and two quadruple mutants (Figure 1).

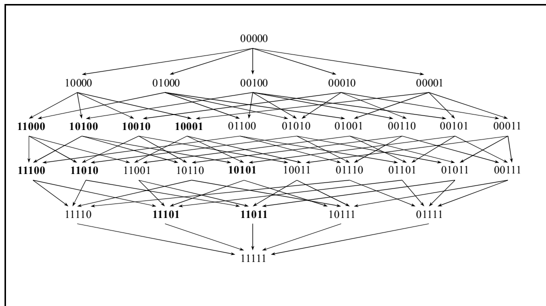


Figure 1. Graphical representation of fitness landscape of size *m* = 5 for the evolution of TEV-*At*17 in *A. thaliana* L*er*-0. The presence/absence of a given mutation is indicated by 1/0, respectively. Genotypes are ordered top-down; starting from the ancestral, wildtype virus; whereby each row represents each of five classes of genotypes; ending with the adapted TEV-*At*17 genotype. Arrows show all possible mutation-fixation pathways. The genotypes marked in bold are missing from our empirical dataset.

With this experiment we sought to answer important questions about the adaptive landscape and its topography. First, we aimed to characterize the prevalence of epistasis and characterize different types of epistasis that may be present in the architecture of fitness. Next, we question whether the differences in fitness are large relative to differences between genotypes, so that many changes are needed to obtain a high-fitness peak? In other words, how rugged is the landscape and how this topography influences the evolutionary potential of a virus population in the new host environment? If small genetic differences are associated to large differences in fitness, the landscape would be rugged. Conversely, if small differences in fitness are associated with large differences between genotypes, the landscape would be smooth.

MATERIALS AND METHODS

Virus genotypes

Virus genotypes were constructed by successive rounds of site-directed mutagenesis starting from template plasmid containing TEV genome: pMTEV (Bedoya & Daròs 2010) using mutagenic primers with specific single-nucleotide mismatch and Phusion® High-Fidelity DNA Polymerase (Finnzymes) following manufacturer's manual. PCR mutagenesis profile consisted of 30 s denaturation at 98ºC, followed by 30 cycles of 10 s at 98ºC, 30 s at 60ºC and 3 m at 72ºC, ending with 10 m elongation at 72ºC. Next, the PCR-mutagenesis products were incubated with *Dpn*I (Fermentas) for 2h at 37ºC in order to digest the methylated DNA template. *Escherichia coli* DH5α electrocompetent cells were transformed with 2µl of these reactions products and plated on LB agar supplied with 100 µg/mL ampicillin. Bacterial colonies were inoculated in 8 mL LB liquid medium supplied with 100 µg/mL ampicillin and grown for 16 h in an orbital shaker (37°C, 225 rpm). Plasmid preparations were done using Promega kit following user manual. Incorporation of mutation was confirmed by sequencing a ca. 800 bp fragment circumventing the mutagenized nucleotide. The plasmid DNA was *Bgl*II linearized and i*n vitro* transcribed using mMESSAGE mMACHINE® SP6 Kit (Ambion) as described in Carrasco *et al.* (2007) in order to obtain infectious RNA of each virus genotype.

Plants inoculation experiments

Since *A. thaliana* cannot be infected by viral RNA, but instead, with virions, we used *N. tabacum* for growing virus particles. Batches of eight-week old *N. tabacum* plants were inoculated with 5 µg of RNA of each viral genotype by abrasion of the third true leaf. Ten days post-inoculation (dpi), the whole infected plants were collected and pooled for each virus genotype. Next, plant tissue was frozen by liquid nitrogen, homogenized using mortar and pestle and aliquoted in 1.5 ml tubes. Saps were prepared by adding 1 ml of 50 mM potassium phosphate buffer (pH 8.0) per gram of N_2-frozen plant tissue. Next, the homogenate was centrifugated at 4ºC and 10 000g for 10 min and the upper liquid phase with 10% carborundum served as sap inocula.

A. thaliana L*er*-0 plants plants were grown in a BSL-2 greenhouse at 25ºC and 16 h light period. Plants were inoculated at growth stage 3.5 regarding the scale of Boyes *et al* 2001. Six plants per virus genotype per block were inoculated with extracts of virus genotypes. The inoculations were done in three independent blocks. Infection was determined by one step RT-PCR as described previously (Lalić *et al.*

2010). Infected whole plants were collected at 21 dpi. Total RNA was extracted and virus accumulation was quantified as described in (Lalić *et al.* 2010).

Statistics

First, we calculated Malthusian growth rate per day, for each genotype, according to the expression: $m = \frac{1}{t} log_{10} Q$ where Q is the number of pg of TEV RNA per 100 ng of total plant RNA quantified at t = 21 dpi (Lalić *et al.* 2011). Then, the relative fitness of each genotype (or just fitness, W), used in all analyses, was calculated by dividing the absolute fitness of a genotype (e^m) (Crow & Kimura, 1970) by the mean absolute fitness of the wildtype TEV ($e^{\overline{m}_{wt}}$): $W = e^{m-\overline{m}_{wt}}$.

Statistical tests were performed using SPSS version 20.

Lethal genotypes

Genotypes 11000, 10100, 10010, 10001, 11100, 11010, 10101, 11101 and 11011 (Figure 1) were not viable in *N. tabacum*. The following section describes our intents to obtain virions of these genotypes since their lethality in tobacco represents artificial holes in the fitness landscape because we know nothing about their viability in *A. thaliana* Ler-0. In first instance, we suspected that the apparent lethality might be due to infidelity of the RNA polymerase supplied with the kit used for *in vitro* transcription that might have incorporated some other mutation(s). Thus, we repeated the *in vitro* transcription twice, independently, for each genotype. The phenotypic effect of all repeated genotypes' transcripts on tobacco was as before: lethal.

In second instance, we suspected that genotypes lethal in tobacco may not really be lethal, but instead, may have lost their ability to infect systemically tobacco plants. Thus, we inoculated 12 *A. thaliana* plants of per virus genotype with sap prepared from tobacco leaf inoculated with these genotypes (necrotic lesions). Herewith, it is important to notice that we did not observe visually the typical necrotic spots, but instead, we performed one-step RT-PCR detection for virus presence in the inoculated leaf using primers that specifically amplify 334 nucleotides of a conserved region from the virus NIb gene as described (Lalić *et al.* 2010). The infection result was negative: none of the plants were infected with either of the virus genotypes. Herewith, false positives might have occurred due to the presence of *in vitro* transcript RNA that served as inoculum and not due to local infection of tobacco leaf.

Thirdly, since we have previously shown the occurrence of antagonistic pleiotropy (Lalić *et al.* 2011), we decided to test the viability of given genotypes on other hosts: *N. benthamiana* and *D. stramonium*. Inoculation of batches of each plant species with 5 µg of RNA *in vitro* transcripts within two independent blocks resulted in all genotypes remaining lethal phenotype in both hosts that are most proximate to tobacco.

RESULTS

Epistasis

Interaction among mutations or epistasis can be calculated using two different approaches: firstly, as a deviation from the expected fitness value of an *n*-tuple mutant and secondly, as the dependence of the effect of a mutation on the genetic background. The latter refers to the sum of squares for the interactions measuring the combined effects of levels of factors; in our case presence or absence of mutations on each of the five loci. Both approaches tend to magnify the interaction sum of squares (Sokal & Rohlf 1995).

As a first approach, the epistasis among the M set of mutations (ε_M) was calculated as $\varepsilon_M = W_M - \prod_{i \in M} W_i$ (Da Silva *et al.* 2010), where W_M is the fitness of the genotype comprising the whole set of mutations and W_i is the relative fitness of a single-mutant genotype comprising the mutation i from M set. Under the null hypothesis of multiplicative (*i.e.*, non-epistatic effects so $\varepsilon_M = 0$), the observed fitness of the genotype M (W_M) equals its expected value given by the product of the fitnesses estimated for each single mutation of which has the M genotype. If $\varepsilon_M > 0$, the epistasis is positive (antagonistic), whereas $\varepsilon_M < 0$ indicates negative (synergistic) epistasis. One-sample t-tests were applied for inferring the significance of the deviation of observed

fitness values form the mean expected fitness for each genotype (Sokal & Rohlf 1995). Figure 2 shows the relationship between observed fitness values of seventeen constructed genotypes (all except single mutants) and their fitness values expected under null hypothesis of multiplicative fitness effects.

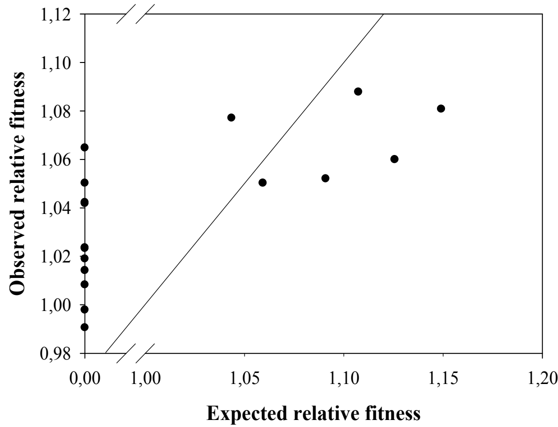


Figure 2. Comparison between the observed and expected multiplicative fitnesses values for seventeen genotypes. The straight line denotes the null hypothesis of multiplicative effects. Deviations from this line result from the epistatic fitness effects.

Four genotypes showed independent fitness effects. Only two genotypes (combinations of mutations at the loci 01010 and 01011) showed significant negative (synergistic) epistasis even when applying a stringent Bonferroni test for multiple comparisons ($P \leq 0.0217$). Eleven genotypes showed significant positive (antagonistic) epistasis ($P < 0.0001$). All these eleven genotypes comprised the compensatory mutation in the locus C3629U that was lethal in a single mutant genotype but viable in combination with any other mutation present in our dataset. Consequentially, zero expected fitness values were obtained using multiplicative fitness thus leading to possible overestimate of the presence of antagonistic epistasis. Therefore, we tested the epistasis using another approach as described in the following.

An alternative way to estimate epistasis is by computing the sum of squares for the interaction among the five mutations. So, in order to address which genotypes act epistatically, we fitted relative fitnesses of genotypes from the whole dataset in a sorted factorial general linear model using 'loci' as factors. The term 'locus' is binomial, stating for the presence or absence of a particular mutation in a particular genotype, so, it total, five factors were used in the model. The fitted model first tests for the single mutations (main effects), sequentially adding pairwise, 3-way, 4-way and 5-way interactions between 'loci', assuming normal distribution and identity linking function. Table 2 shows only the effects of interactions between mutations; the main effects and the intercept of the fitted model make no biological sense and therefore are not presented. Zero interaction coefficients indicate the lack of epistasis. Out of 17 combinations of mutations in our dataset, eleven cases corresponded to significant epistasis and six genotypes were not significantly epistatic (Table 2).

Table 2. Sorted factorial GLM estimates of epistasis testing for the interactions among mutations across genotypes.

Genotype	Epistasis (±S.E.)	χ^2	*P*
01100	0,1558±0,0331	20,6568	0,0000
01010	-0,2982±0,0310	71,6540	0,0000
01001	-0,1731±0,0302	29,6551	0,0000
00110	0,2620±0,0309	58,4557	0,0000
00101	0,1716±0,0310	27,8120	0,0000
00011	-0,2845±0,0310	66,4177	0,0000
11001	0,0087±0,0842	0,0107	0,9175
10110	0,0000		
10011	0,0000		
01110	-0,4842±0,0617	51,2870	0,0000
01101	-0,3394±0,0619	27,2613	0,0000
01011	0,4920±0,0619	52,3785	0,0000
00111	-0,5622±0,0617	65,5349	0,0000
11110	0,0000		
10111	0,0000		
01111	0,9060±0,1234	45,8106	0,0000
11111	0,0000		

Among eleven genotypes with significant epistasis, five cases corresponded to positive epistasis and six cases corresponded to negative epistasis (Figure 3). Among the six non-significant cases, five cases corresponded to multiplicative effects (*i.e.*, were not epistatic). All non-significant cases of epistasis were genotypes having the mutation at the

locus U357C, whereas all other genotypes that showed significant epistasis did not comprise this mutation. Still, the differences in relative fitness between these two groups of genotypes, epistatic and non-epistatic, were not significant (one tailed *t*-test, 15 d.f., $P = 0.3156$).

Two different tests applied for testing the presence of epistasis were incongruent in estimating the proportions of positive and negative epistasis. Indeed, there is a significant association between the type of the test and epistasis outcome (Fisher's exact test, $P = 0.0551$).

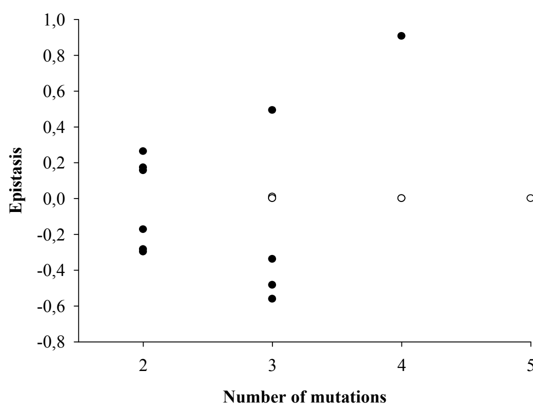


Figure 3. Relation between the epistasis resulting from GLM and the classes of genotypes represented by the number of mutations that they carry. Filled circles correspond to significant cases of epistasis; open circles correspond to non-significant cases of epistasis.

Figure 3 reveals that epistatic interactions increase in frequency in the first steps of adaptation (*i.e.*, fixation of the second mutations) and become less common and approach to zero epistasis with increasing number of mutations. In concrete: all double mutants (100%) showed significant epistatic interactions; 57.14% of triple mutants had significant epistasis; 33.33% of quadruple mutants were epistatic and quintuple mutant showed no epistasis (0%). Moreover, Figure 3 pictures one more interesting feature: the bimodal distribution of epistasis per class of genotypes. Here, bimodal distribution refers to unequal proportion of positive and negative epistatic effects among genotypes belonging to the same class. With increasing number of mutations, the epistasis becomes more negative and tends to zero as in the case of the evolved quintuple genotype. There was equal proportion of positive and negative epistasis cases among double mutants (three *vs*. three genotypes). Only one out of four triple mutants had positive epistasis; the rest had negative epistasis coefficient. One out of three quadruple mutants had strong positive epistasis while the rest lack epistasis. Quintuple mutant showed multiplicative fitness effects.

Next, following definitions given by Poelwijk *et al*. (2011), we calculated sign and reciprocal sign epistasis among constructed genotypes by pairwise comparison of fitnesses of single (W_i) and $M \mid i$ (W_M, $i \notin M$) genotypes. Out of 32 combinations of significant cases of epistasis, eleven genotypes corresponded to the cases of magnitude epistasis: all double mutants: 01100, 01010, 01001, 00110, 00101, and 00011, four triple mutants: 01110, 01101, 01011, 00111 and one quadruple mutant: 01111; and four genotypes corresponded to the sign epistasis: three triple mutants: 01110, 01011 and 00111, and one quadruple mutant: 01111. Out of these four cases of sign epistasis, two genotypes: 01110 and 01111 corresponded to reciprocal sign epistasis (Figure 5). Thus, magnitude epistasis was prevalent in the dataset (73.33%). The rest, 26.67% of significant cases of epistasis, corresponded to sign epistasis. To test this ratio, we counted the number of cases among $n = 32$ possible combinations that fulfilled the condition of either $n - 1$ single mutant having fitness smaller than observed for their corresponding n-tuple mutant. The expected probability of sign epistasis was 100%. Using one-tailed Binomial test, the probability of observing 4 or fewer cases of sign epistasis among 15 was $P < 0.0001$ thus confirming the expected ratio of magnitude and sign epistasis. The same test was applied for inferring the significance of reciprocal sign epistasis where we counted the number of cases that fulfilled the condition of both single mutants having fitness smaller than the observed for the corresponding n-tuple mutant. The expectation of reciprocal sign epistasis was 7/22= 0.3182 and the probability of observing 2 or less cases of reciprocal sign epistasis among 4 cases of sign epistasis was $P = 0.3805$, rejecting deviations from the expectation.

Since the mutations at the loci U357C and C6906U were synonymous, we tested their effects on relative fitness and epistasis by fitting the data to sorted full factorial linear generalized model using normal distribution and identity linking function. We found no significant effect of synonymous mutations nor their interaction on any of the response variables ($\chi^2 = 0.8881$, 1 d.f., $P = 0.3460$; $\chi^2 = 0.0499$, 1 d.f., $P = 0.8232$, respectively). In comparison, we performed the same test using only the three non-synonymous mutations and found significant effect of mutations on relative fitness of genotypes (Table 3; whole model test: $\chi^2 = 61.8555$, 7 d.f., $P < 0.0001$), but not on epistasis (whole model test: $\chi^2 = 1.2124$, 7 d.f., $P < 0.9906$). Additionally, to check whether adaptive mutations arise from codon-bias adaptation, we compared the contingency in codon usage frequency for two synonymous mutations in TEV-*At*17 (Table 1) and found no significant tendency (Fisher's exact test, $P = 0.4405$).

Table 3. Non-synonymous mutations interactions resulting from a generalized linear model.

Response	Genotype	d.f.	χ^2	*P*
Relative fitness	01100	1	5.0269	0.0250
	01010	1	23.7154	< 0.0001
	00110	1	13.9560	0.0002
	01110	1	13.8587	0.0002
	Full model	7	61.8555	< 0.0001

Fitness of virus genotypes

Differences in relative fitness were significant among genotypes ($\chi^2 = 2551.6320$, 21 d.f., $P < 0.0001$). Figure 5 shows the differences in fitness among genotypes carrying the same number of mutations. It can be noted that fitness differences between single mutants are greater than of the double mutants, those being greater than of the triple mutants with quadruple mutants having almost indistinguishable fitness. Indeed, differences in fitness between these genotypic classes were significant ($\chi^2 = 3.6585$, 4 d.f., $P = 0.0073$). Tukey-Kramer comparisons of means for all pairs revealed that *n*-tuple mutants cluster in total in three distinct groups where single mutants form the first cluster of lowest relative fitness; the second cluster is formed by quadruple and quintuple mutants and the third cluster form double and triple mutants. Since the individually lethal mutation at the locus C3629U would not contribute to the adaptation, we reexamined the fitness ranking of genotypes with respect to the number of mutations they carry and observed the following trend in fitness decrease. The highest fitness was among triple mutants, followed by double, simple, quadruple and quintuple mutants (Figure 5). The wildtype virus had the lowest fitness.

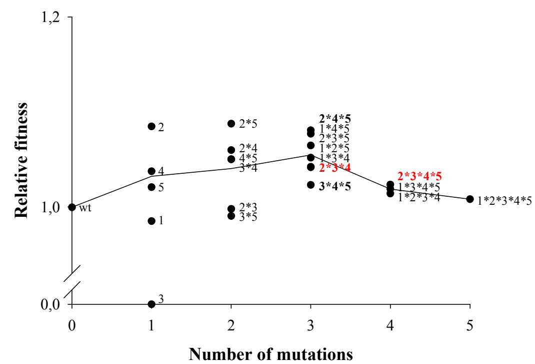


Figure 5. Relation between mean relative fitness of each genotype with respect to the number of mutations that it carries. Each point in the graph represents the mean fitness value of a particular genotype. Next to their fitness value, genotypes are denoted in correspondence to the locus where the particular mutation is present (Table 1). Higher order interaction between loci is denoted by a star (*). Genotypes marked in bold correspond to the sign epistasis. Genotypes marked in red are cases of reciprocal sign epistasis. The inserted line connects the group means between genotypes carrying *n*-tuple mutations.

Infectivity

Infectivity, as a fitness component, was estimated as the frequency of infected plants out of the total number of inoculated plants. Expected frequency with 95% confidence intervals (CI) was approximated as LaPlace's point estimator for the Binomial frequency parameter using http://www.measuringusability.com/wald.htm server. We found significant differences in infectivity among virus genotypes ($\chi^2 = 0.1025$, 22 d.f., $P < 0.0208$). Figure 6 represents the infectivity plotted for each genotype. Monotonic increase in infectivity indicates two most probable evolutionary trajectories. The first may be: 00000 → 01000 → 01001 → 01101 → 01111 → 11111. The second may be equally

probable: 00000 → 00001 → 00101 → 01101 → 01111 → 11111. Both paths consist of three steps that do not lead neither to increase nor decrease in infectivity and remaining two steps that do lead to the increase in infectivity (Figure 6). Thus, 60 % of the infectivity landscape consisted of neutral steps, while the rest lead to an increase in infectivity.

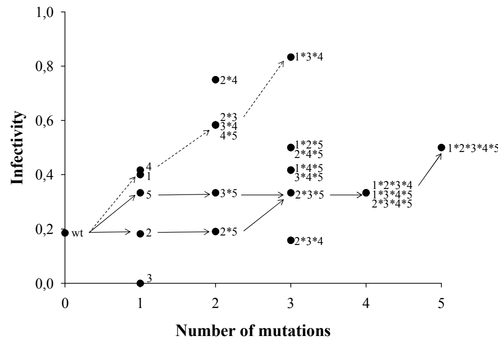


Figure 6. Relation between infectivity and classes of genotypes with corresponding number of mutations. Classes of genotypes are indicated next to their adjacent infectivity value represented by a dot. Genotypes are represented as loci; with respect to the presence of a particular mutation (Table 1). Higher order interaction between loci is denoted by star (*). Solid arrows indicate monotonic increase or stasis in infectivity, thus represent the most probable pathways in the infectivity landscape. Dashed arrows indicate continuous increase in infectivity leading to the dead-end pathway.

Discussion

Here we address empirically the partial fitness landscape of a positive sense plant RNA virus TEV-*At*17 adapted to the novel host *A. thaliana*. Our dataset consisted of 22 genotypes out of 32 possible combinations of the five mutations fixed by the evolved TEV-*At*17. We measured two fitness components: the infectivity and within host virus growth rate relative to the wildtype as an approximate to relative fitness.

Firstly, we focused on epistasis measured as the deviation from the expected null hypothesis of multiplicative mutational fitness effects. Epistasis was prevalent (76.47%) in our dataset and predominately positive (84.62%), meaning that the absolute effect of the second mutation is smaller than that of the first. Previously, we have found the significant prevalence of positive epistatic effects among deleterious mutations in TEV measured in its primary host *N. tabacum* (Lalić & Elena 2012). Thus, positive epistasis is common in TEV genome among both deleterious and beneficial mutations. Strong mutational effects give rise to positive epistasis (De Visser *et al*. 2003, 2011; Proulx & Phillips, 2005; Desai *et al*. 2007), thus indicating low genetic robustness of RNA viruses. In this case, at the extreme, a long series of such mutations in a population with sufficiently high fitness would eventually lead to an effectively neutral rate of mutation accumulation with no further measurable fitness changes (Loewe & Hill 2010). The fitness landscape data are in concordance to this (Figure 5). Moreover, our results show that the epistasis governs the initial steps of adaptation; during fixation of the second and third mutation. On the contrary, independent fitness effects took a commonplace in the final steps of adaptation (Figure 1).

Additional test for the interactions between mutations confirmed the prevalence of epistatic effects between mutations (herewith 64.71%). Still, in contrary to the former, this test showed almost equal proportion of negative (54.54%) and positive (45.45%) cases of significant epistasis. The slight dominance of negative over positive epistasis implies jumping over the low fitness valley that is very unlikely to occur. Still, since both tests applied are equally valid and accounting for lack of more than 30% of data in our dataset (*i.e.*, ten genotypes that could not be tested experimentally), we are not able to draw any further firm conclusions about the prevalence of particular type of epistasis. Analogous studies that sought to address the contribution of epistasis among beneficial mutations in evolving bacterial (Chou *et al.* 2011; Khan *et al.* 2011) and yeast (Kvitek & Sherlock 2011) found a predominance of negative epistasis that impeded the rate of adaptation. Negative epistasis has diminishing returns effect meaning that adaptive mutations of large individual effect result in a smaller fitness effect when they occur together. Diminishing returns fitness effects were firstly observed in the experiments of long-term evolution of *E. coli* (de Visser & Lenski 2002; Barrick *et al.* 2009) within which the initial fitness improvement was fast but it

rapidly decreased over time and remained low. However, out fitness data show no pattern of diminishing returns, but instead, the adaptation began with relatively small steps in fitness increase until the fixation of the third mutation, a part from which, the fitness (in quadruple and quintuple mutants) decreased (Figure 5).

Epistasis causes stochastic differences in the rank order of mutations, hence, directly influences the evolutionary trajectories of populations (Martínez *et al.* 2011). Recent works on the contribution of epistasis to the architecture of fitness of RNA viruses found significant portion of ruggedness of fitness landscapes. Lalić and Elena (2012) found prevalent contribution of epistasis, especially of reciprocal sign type, to the architecture of TEV fitness suggesting that the adaptive landscape of TEV in its primary host (tobacco) must be highly rugged. Simultaneously, Hinkley *et a*l. (2011) confirmed the commonality and strength of epistasis in HIV-1 protease and reverse transcriptase. In a concomitant study, Kouyos *et al.* 2012 analyzed fitness landscapes derived from *in vitro* fitness measurements of HIV-1 and reported ruggedness of the HIV-1 adaptive landscape. Because different peaks can also differ in magnitude, the adaptive landscape could impose an additional constraint on the evolvability of the organism. Rugged fitness landscapes have highly negative consequence on the evolvability of a population. Since a rugged fitness landscape lacks large constant-fitness plateaus, it is thus antirobust. In such a landscape, a population can become confined to a region of genotype space in which it must wait for the occurrence of the advantageous mutation(s). However, our data contrast former observations. Figures 5 and 6 reveal that fitness landscape of TEV was single-peaked, implying the lack sign epistasis (Poelwijk *et al.* 2007). Indeed, our results showed that majority of epistatic effects were attributed to magnitude epistasis (about 73%). In addition, the infectivity data revealed two possible pathways of adaptation, each carrying three out of five neutral steps (Figure 6), indicating the 60% neutrality of the infectivity landscape. Accounting for both measurements of fitness; the relative growth rate and the infectivity, we conclude that the fitness landscape of TEV adapted to *A. thaliana* in its new host is mostly neutral and smooth, bearing a single peak. Several other studies have reached the similar observations. Weinreich (2005) explored fitness landscape lacking sign epistasis and found a significant effect between landscape membership in fitness rank-ordering and genetic constraints of genotypes given by their fitness value. Later on, Weinreich *et al.* (2006) explored the fitness landscape of the *E. coli* β-lactamase conferring resistance to cefotaxime due to five mutations and found that majority of the trajectories contained fitness decreasing or neutral steps. Kvitek and Sherlock (2011) observed few deleterious mutations that hitchhiked along with one or more adaptive mutations in the evolved yeast clones.

Mutations are random with respect to their effect in improving the fitness. Evolution is driven either by natural selection of beneficial mutations or by stochastic fixation of selectively neutral or slightly deleterious mutations due to random genetic drift. Mutation at locus U357C showed independent fitness effects in all *n*-tuple mutant states implying that it was not important target of natural selection, but instead, results as a by-product of genetic drift. In favor to this observation goes the codon usage of *A. thaliana* that was lower for the mutated codon in comparison to codon of the wildtype virus (Table 1). The evidences we presented here about the commonplace of epistasis together with the genotypes' fitness results (Figure 5) suggest that the evolution of TEV within *A. thaliana* L*er*-0 have not occurred solely by natural selection of mutations that improved TEV fitness within a new host. A major role in this evolutionary event most probably played the genetic drift associated with serial passages. In nature, virus populations experience bottlenecks during transmission and cell-to-cell movement. Thus, genetic drift is an ever-present source of stochastic variation in allele frequencies among virus populations so it cannot be disregarded. In fact, genetic drift can lead to decrease in the mean fitness of an asexual population due to the process known as Muller's ratchet (Mueller, 1964) which proceeds if all

individuals with the minimum number of deleterious mutations are lost by chance. Thus, Muller's ratchet can lead a population into extinction (Lynch *et al.* 1993). In the case of independent mutational fitness effects, the rate of fitness decline is constant, but if there is a positive epistasis between deleterious mutations, as the ratchet advances, the frequency of the best available genotype will increase, making its loss less probable (Kondrashov 1994). Consequently, synergistic epistasis can arrest the action of Muller's ratchet and provide the survival of the population although with lower mean fitness (Kondrashov 1994).

Additionally, it is important to note that the missing data influenced the results interpretation. For instance, we cannot argue that all tested genotypes form part of the same continuous network, so that mutations arose in a successive order. It is possible that independent mutation events in polymorphic viral population contributed jointly (or by recombination) to fixation of the observed TEV-*At*17 consensus sequence. In the former case, a two or more distinct networks may exist, with multiple starting points.

CONCLUDING REMARKS

Here, we aimed to reconstructed partially the first empirical adaptive landscape of a ss(+)RNA virus. This is the first experimental fitness landscape for a whole organism. We found prevalent epistatic effects between mutations, both positive and negative, that led to different fitness ranking of genotypes. Furthermore, we characterized the landscape topography approximated by the number of the plants infected by a particular virus genotype and found that more than a half of possible evolutionary paths were neutral and the landscape was single peaked. The lack of ruggedness of TEV's landscape in *A. thaliana* suggests that virus easily adapted to its new host. This is an important result from the perspective of emerging viral diseases and in concordance to common observation that majority of emerging viruses are indeed RNA viruses.

ACKNOWLEDGEMENTS

We thank Francisca de la Iglesia for excellent technical support and José A. Daròs for helpful technical advices on cloning. This work was supported by grant BFU2009-06993 from Spanish Ministerio de Ciencia e Innovación to S.F.E. J.L. was supported by a JAE-pre contract from CSIC.

REFERENCES

Agudelo-Romero, P., P. Carbonell, M.A. Perez-Amador, and S.F. Elena (2008) Virus adaptation by manipulation of host's gene expression. *PLoS ONE* 3: e2397.

Barrick, J.E., D.S.Yu, S.H. Yoon, H. Jeong, T.K. Oh, D. Schneider, R.E. Lenski, and J.F. Kim (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243–1247.

Bedoya, L.C., and J.A. Daròs (2010) Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149: 234-240.

Boyes, D.C., A.M. Zayed, R. Ascenzi, A.J. McCaskill, N.E. Hoffman, K.R. Davis, and J. Görlac (2001) Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* 13: 1499–1510.

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

Chou, H.H., H.C. Chiu, N.F. Delaney, D. Segrè, and C.J. Marx (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332: 1190–1192.

Crow, J.F., and Kimura, M. (1970). *An Introduction to Population Genetics Theory* Harper and Row: New York.

Da Silva, J., M. Coetzer, R. Nedellec, C. Pastore, and D.E. Mosier (2010). Fitness epistasis and constraints on adaptation in a *Human immunodeficiency virus* type 1 protein region. *Genetics* 185: 293-303.

Dawid, A., D.J. Kiviet, M. Kogenaru, M. de Vos, and S.J. Tans (2010) Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape. *Chaos* 20: 026105.

De Visser, J.A.G.M., T.F. Cooper, and S.F. Elena (2011). The causes of epistasis. *Proc. R. Soc. B* 10: 3617–3624.

De Visser, J.A.G.M., J. Hermisson, G.P. Wagner, L. Ancel Meyers, H. Bagheri-Chaichian, J.L. Blanchard, L. Chao, J.M. Cheverud, S.F. Elena, W. Fontana, G. Gibson, T.F. Hansen, D. Krakauer, R.C. Lewontin, C. Ofria, S.H. Rice, G. von Dassow, A. Wagner and M.C. Whitlock (2003). Evolution and detection of genetic robustness (Perspective). *Evolution* 57(9): 1959–1972.

de Visser, J.A.G.M., and R.E. Lenski (2002) Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as

cause of declining rate of adaptation. *BMC Evol. Biol*. 2: 19.

de Visser, J.A.G.M., S.C. Park, and J. Krug (2009) Exploring the effect of sex on empirical fitness landscapes. *Am. Nat.* 174: S15–S30.

Desai, M.M., D. Weissman, and M.W. Feldman (2007) Evolution can favor antagonistic epistasis. *Genetics* 177: 1001-1010.

Franke, J., A. Klözer, J.A.G.M. de Visser, and J. Krug (2011) Evolutionary accessibility of mutational pathways. *PLoS Comp. Biol.* 7: e1002134.

Hinkley, T., J. Martins, C. Chappey, M. Haddad, E. Stawiski, J.M. Whitcomb, C.J Petropoulos, and S. Bonhoeffer (2011) A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat. Genet*. 43: 487–489.

Khan, A.I., D.M. Dinh, D. Schneider, R.E. Lenski, and T.F. Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332:1193–1196.

Kvitek, D.J., and G. Sherlock (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* 7: e1002056.

Kondrashov, A.S. (1994) Muller's ratchet under epistatic selection. *Genetics* 136: 1469–1473.

Kouyos, R.D., G.E. Leventhal, T. Hinkley, M. Haddad, J.M. Whitcomb, C.J. Petropoulos, and S. Bonhoeffer (2012) Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genet.* 8: e1002551.

Lalić, J., P. Agudelo-Romero, P, Carrasco, and S.F. Elena (2010) Adaptation of *Tobacco etch potyvirus* to a susceptible ecotype of *Arabidopsis thaliana* capacitates it for systemic infection of resistant ecotypes. *Phil. Trans. R. Soc. B* 65: 1997-2008.

Lalić, J., J.M. Cuevas, and S.F. Elena (2011) Effect of host species on the distribution of mutational effects for an RNA virus. *PLoS Genet*. 7: e1002378.

Lalić, J., and S.F. Elena (2012). Epistatic effects between mutations are host-dependent for a plant RNA virus. *Biol. Lett.* doi:10.1098/rsbl.2012.0396 1744-957X

Loewe, L., and W.G. Hill (2010) The population genetics of mutations: good, bad and indifferent. *Phil. Trans. R. Soc. B* 365: 1153–1167.

Lozovsky, E.R., T. Chookajorn, K.M. Brown, M. Imwong, P.J. Shaw, S. KamchonwongpaisaD.E. Neafsey, D.M. Weinreich and D.L. Hartl (2009) Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Nat. Acad. Sci. USA* 106: 12025–12030.

Lunzer, M., S.P. Miller, R. Felsheim, and A.M. Dean (2005) The biochemical architecture of an ancient adaptive landscape. *Science* 310: 4899–501.

Lynch, M., R. Bürger, D. Butcher and W. Gabriel (1993) The mutational meltdown in asexual popualtions. *J. Heredity* 84: 339–344.

Martínez, J.P., G. Bocharov, A. Ignatovich, J. Reiter, M.T. Dittmar, S. Wain-Hobson, and A. Meyerhans (2011) Fitness ranking of individual mutants drives patterns of epistatic interactions in HIV-1. *PLoS ONE* 6: e18375.

Muller, H.J. 1964. The relation of recombination to mutational advance. *Mut. Res*. 1: 2–9.

Poelwijk, F.J., D.J. Kiviet, D.M. Weinreich, and S.J. Tans (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445: 383–386.

Poelwijk, F.J., S. Tanase-Nicola, D.J. Kiviet, and S.J. Tans (2011) Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* 272: 141–144.

Proulx, S.R., and P.C. Phillips (2005) The opportunity for canalization and the evolution of genetic networks. *Am. Nat.* 165: 147–162.

Sokal, R.R., and F.J. Rohlf (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*. Freeman, New York.

Weinreich, D.M. (2005) The rank ordering of genotypic fitness values predicts genetic constraints on natural selection on landscapes lacking sign epistasis. *Genetics* 171: 1397–1405.

Weinreich, D.M., N.F. Delaney, M.A. DePristo, and D.L. Hartl (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312: 111-114.

Weinreich, D.M., R.A. Watson, and L. Chao (2005) Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59: 1165-1174.

Welch, J.J., and D. Waxman (2005) The *nk* model and population genetics. *J. Theor. Biol.* 234: 329–340.

Withlock, M.C., P.C. Phillips, F.B.G. Moore, and S.J. Tonsor (1995) Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Evol. Syst.* 26: 601-629.

Wright, S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Intl. Congress Genetics* 1: 356-366.

GENERAL SUMMARY

The relationship between genotype and fitness is one of the most important functions in determining the evolutionary dynamics of a population. In this Thesis, multidimensional genotype-fitness maps have been explored. In Chapter III, the fitness of genotypes was mapped across different environments. In Chapter IV, the fitness of genotypes was mapped with respect to a particular genetic background. In Chapter V, the fitness of genotypes was mapped with respect to both environment and genetic background. The final chapter aimed to map the fitness of a partial set of genotypes to the new environment where they had evolved. In all cases, except of part II, a significant non-linear relation for viral fitness was found indicating that fitness is a unique consequence of its genotype in interaction with the genetic background and the environment. Still, it is important to note that Population Genetics does not assume linear effects being a rule in Nature; actually, they are not. Instead; since the response to selection depends on the fraction of variance in fitness that can be attributed to statistically linear component of mutational effects, the linear component of genotype regression for fitness is of particular importance. Hence, the adaptation depends on these additive effects. Herewith, it is worth to recall that during evolution not only genotypes change in response to an environment, but environments are also rarely constant. The heterogeneity of environments is due to both biotic and abiotic causes. In this sense, the evolutionary theory remains to be a powerful tool for inferring parasite-host interactions because we still lack much of knowledge in understanding the environmental and genetic bases of interactions, particularly in plant-virus system. It is worth keeping in mind that evolution is a process and not an event, so, evolutionary theory provides also a powerful tool for interpretation of host-parasite dynamics and alerts about the necessary design and application of disease control programmes.

Biology is a science of complex systems. The constituents of a complex system interact in many different ways that lead to dynamic and emergent features that cannot be predicted satisfactorily by linear mathematical models that disregard interaction and non-additive effects. Concordantly, the essential message of the work presented here is that the dynamic interplay between viral and host populations is necessarily a non-linear one, and as such has many properties that cannot be predicted simply by analyzing the structure of their components.

CHAPTER III – *G*×*E* INTERACTIONS

New emerging and re-emerging viruses represent an increasing threat to human and animal health and to agronomy (Holmes 2009). However, the ability to predict when a virus may become emerging is limited by the lack of knowledge about the genetic basis of emergence. Particularly relevant in this sense is to know how the effect of mutations on viral fitness changes across the reservoir and potential new hosts. Here we provide a first description of the distribution of mutational fitness effects for an RNA virus across a panel of hosts of increasing phylogenetic distance from the reservoir species.

A key pre-requisite for a virus to jump the species barrier from its reservoir to a new one is the existence of genetic variants that may have a significant fitness in the new host (Turner & Elena 2000). These host-range mutations, usually, are deleterious in the reservoir host and their frequency is thus low. Therefore, to predict the likelihood of a virus to infect new hosts, it is necessary to characterize the distribution of mutational fitness effects (DMFE) on its reservoir host as well as on potential new hosts. DMFE have been characterized for a handful of viruses in their reservoir hosts (Sanjuán 2010), but whether and how these distributions change across potential hosts has never been experimentally addressed. To cover this gap, here we sought to characterize the DMFE for the plant RNA virus TEV. To do so, we generated 20 single-nucleotide substitution mutants of TEV. Both, the nucleotide site chosen and the replacement made were both randomly chosen. Then, we measured the absolute fitness of each genotype across a panel of eight susceptible hosts, including the natural one, *N. tabacum*. The alternative hosts spanned a wide range of genetic distances from the reservoir. Fitness on each host was measured as a growth rate by quantifying the accumulation of viral genomic RNA using RT-qPCR.

We first characterized the properties of the DMFE across hosts. We found that the distributions were almost undistinguishable for the natural host and its close relative *N. benthamiana*. However, as the genetic relatedness to tobacco decreased, the distributions moved towards lower fitness values, thus reflecting the overall lack of adaptation of TEV to these hosts. Interestingly, the shape of the distribution also changed as genetic distance among hosts increased, switching from being negatively skewed to positively skewed. This switch indicates that in the reservoir and close relatives TEV mutant genotypes are, in general, deleterious and have lower fitness than the non-mutated genotype. By contrast, in unrelated hosts the fraction of mutant

genotypes with fitness values larger than the wild-type dominates. This result shows that under stressful conditions the proportion of beneficial mutations increases, in agreement with certain theoretical predictions.

Next, we sought to quantify to which extent the phenotype of TEV (*i.e.,* its fitness) was determined by the interaction between its genotype and the environment (*i.e*., its host). Understanding how genotype and environment interact to determine the phenotype and fitness has been a central aim of ecology, genetics, and evolution and, for instance, the fate of genetic variation in populations depends on the form of the $G\times E$ interactions. For any organism, environmental components are prevalent, thus environment cannot be dissected from the phenotype of a genotype. Because phenotypically plastic adaptations are more likely to evolve in variable environments than fixed adaptations, and species interactions are intrinsically variable in space and time, the (co)evolution of species interactions has certainly resulted in $G\times E$ (Agrawal 2001). Since natural selection acts on phenotypes but evolution occurs only through genetic change in populations, $G\times E$ reduces the global efficiency of natural selection and can even result in the maintenance of polymorphism. Thus, the existence of genotype-by-environment interactions calls into question the existence of globally adaptive gene combinations (Wright, 1977). We found that the fitness of a given TEV genotype depended on the particular host wherein it was measured, which translates in a highly significant $G\times E$ interaction (67% of the observed variability in TEV fitness could be explained by this interaction). This is not the first observation the response of a particular genotype is conditional upon the environment and as such, cannot be predicted. The work presented here hints upon the form and the extent of the presence of $G\times E$ even in the simple organism such as a RNA virus. Lewontin (1974) pointed out upon the unpredictability of individual phenotypic responses once the $G\times E$ interaction over a wide range of environments was considered. Here we show that there is neither an overall effect of genotype nor environment, but both can appear in the given set of virus mutants in a particular environment, as discussed above.

A significant $G\times E$ can arise by two non-exclusive mechanisms. First, pleiotropic effects may change the rank order of mutations across hosts (*e.g*., a mutation beneficial in one host may not be so in an alternative one). Second, while still retaining the rank order of fitness effects, $G\times E$ can be generated by altering the genetic component of phenotypic variance. We evaluated these two hypotheses and found that we could not specifically reject any of them: pleiotropic effects were abundant (positive among hosts

of the tobacco family but negative for the outsider hosts) and the genetic component of variance for fitness decreased in comparison to what was observed in the reservoir. The quantity of genetic variation is not a fixed property of a population, but may vary according to the environment in which the population is living. Therefore, the purpose of this study was to predict the fate of virus genetic variation in response to selection represented by different hosts that mutually differ in their phylogenetic proximity. It has been showed that genetic variation can be maintained by mutation in polygenic characters even under strong selection; when selection is weak, a small increase in the strength of selection can produce a major decrease in the heritability (Lande 1976). Our results confirm these predictions; we showed that the direct action of natural selection in the form of the environment (*i.e.*, host) is proportional to the contribution of genetic variance for fitness. Indeed, when comparing *Solanaceae vs.* non-*Solanaceae*, or, in other words, "local" *vs.* "foreign" environments, the heritability drops from 0.965 to 0.239 (Table 5 in Chapter III) respectively, and the genetic variance drops from 0.083 to 0.002 respectively. Keeping in mind the high amount of genetic variability that RNA virus populations posses, this result indicates that it can be maintained even in the new hosts to which the virus is not adapted nor have shared a common coevolutionary history.

The major objection to this work can be attributed to the experimental setup. Within our results interpretation we assume genetic homogeneity of the virus genotype population quantified at 10 dpi. During this time period, some additional mutations might have appeared within the viral populations and might have contributed to the observed fitness of genotypes approximated as within-host growth rate. That is why we have collected plants and measured viral fitness at the minimal possible time point for all hosts. In addition, the inexactitude of our methods of measurement can be reflected in the significant variance arising between plants; *i.e.,* biological replicates.

In conclusion, we have shown that the location and shape of the DMFE for an RNA virus depends on the host wherein it is evaluated. Therefore, virus genotype and the host species interact in a non-linear manner to determine viral fitness. Pleiotropic effects and reductions in genetic variance contribute to generate this genotype-by-host interaction. The implications of these observations for understanding the emergence of new viruses are multiple, and hint on the unpredictability of the process; in the light of information collected on the reservoir host one can not anticipate which viral genotypes will be more likely to emerge.

CHAPTER IV – *G*×*G*

In this work we addressed the multi-dimensional epistasis within TEV genome by measuring the growth rate as a proxy for fitness of 53 double mutants constructed by randomly combining 20 single-nucleotide substitution mutants whose deleterious fitness effects were also measured in its primary host, *N. tabacum*. The epistasis was prevalent and predominately positive indicating strong mutational fitness effects, and consequently, low genetic robustness of the virus. Sign epistasis dominated over the magnitude epistasis. The major contribution of reciprocal sign epistasis among the significant cases of sign epistasis implies ruggedness of TEV fitness landscape.

Further on, some possible objections to our work are discussed. Apparently, low power in our measurements is reflected in the small gap between the lowest detected fitness (0.21) and the fitness of the lethals (0). Indeed, all of the single- and double-mutant fitnesses are within a relatively narrow range (0.21-0.34) of fitnesses and this gap between viable and lethal genotypes is actually falsely too small. This is due to definition of epistasis used here; as the difference between observed fitness and the product of fitnesses of single mutants. Initially, we used Malthusian parameter as in Chapter III in order to calculate epistasis. Epistasis among pair of mutations x and y, ε_{xy}, was calculated as: $\varepsilon_{xy} = m_{00} + m_{xy} - m_{x0} - m_{0y}$ (Fisher, 1918; Phillips, 2008), where m_{00}, m_{xy}, m_{x0}, m_{0y} correspond to the Malthusians of the wildtype, the double mutant and of each single mutant, respectively. In the light of null hypothesis of additive mutational effects, the results were following. Forty epistasis values significantly departed from the null hypothesis of additive fitness effects (t-tests, in all cases $P \leq 0.0469$). Comparing between additive and multiplicative epistasis, the results remained qualitatively identical although, obviously, differ quantitatively. Mainly, the number of significant cases of epistasis is reduced to 20 when using multiplicative epistasis. Still, justification of the use of multiplicative epistasis is noted in section 1.5.

CHAPTER V – *G*×*G*×*E*

Here we studied the variability of interactions between pairs of random point mutations in the genome of TEV across four different hosts. In nature, TEV infects two of these hosts (*N. tabacum* and *D. stramonium*), that belong to the same plant family, the

Solanacea. The other two species are not TEV natural hosts, although they are experimentally susceptible to systemic infection. They belong to the *Asteraceae* (*H. annuus*) and the *Amaranthaceae* (*S. oleracea*) families, respectively. We quantified fitness of all single and double mutants on each of the hosts and calculated the multiplicative epistasis coefficient. We found significant epistatic effects for nine double mutants in the two TEV natural hosts (*i.e.,* permissive environments) among which four corresponded to cases of sign epistasis and the rest corresponded to magnitude epistasis. For non-natural hosts, the number of significant epistatic interactions dramatically dropped to one. Furthermore, when we tested the effect of host on the magnitude of epistasis, we found that 40% of cases showed significant variation among hosts. The intensity of epistasis decreased as the genetic relatedness between the primary host and alternative ones increased.

To conclude, we have shown that magnitude of epistasis between mutations declines as the host becomes more distantly related to the typical host. This suggests that epistasis is small when conditions are stressful. Further, there is evidence for both synthetic lethality and suppression of lethality, both of which are extreme forms of negative and positive epistasis for fitness, respectively. Such epistasis makes predicting host shifts difficult, which is an important result in the aspect of viral emergence. Understanding the variability of sign and magnitude epistasis in the context of virus host range expansion will give us new and very much necessary insights into the evolutionary ecology of infectious diseases.

CHAPTER VI - Empirical fitness landscape

> *The world as we perceive it is three dimensional. Physicists currently believe one needs on the order of a dozen dimensions to explain the physical world. However, biological evolution occurs in a space with millions of dimensions* (Gavrilets 2003).

The dynamics of adaptation depend on the underlying fitness landscape, thus it is critical to know the proportion of peakedness *vs.* flatness of the adaptive landscapes. This is especially important in the context of emerging viruses in order to predict the tempo and mode of virus adaptation to a new host and how this adaptation may be constrained. In this work we addressed the partial, empirical fitness landscape of TEV

adapted to *A. thaliana*. The adaptation to a new host consisted in acquiring five point mutations. We have constructed approximately all intermediate genotypes of adaptation and quantified the infectivity (as a proxy to transmissibility) and within host virus growth rate relative to the wildtype (as an approximate to relative fitness) within *A. thaliana* for a large fraction of them. Epistasis among the adaptive mutations was strong and, in major proportion, corresponded to magnitude epistasis. We found cases of both positive and negative deviations from the expected multiplicative fitness effects. Epistasis was more prevalent among the genotypes bearing less than four mutations, while quadruple and quintuple mutants showed independent mutational fitness effects. Infectivity data showed that the potential evolutionary pathways consist of neutral steps in major proportion (60%). The landscape consisted of a single peak. The smoothness of TEV landscape in *A. thaliana* suggests that the virus apparently had no constraints in adapting to this new host, even if the new host belongs to different family of plants. Recall that TEV typically infects members of the family *Solanaceae*, while *A. thaliana* belongs to *Brassicaceae*. This observation has strong implications for the emerging virus diseases because RNA viruses comprise great evolutionary potential that enables them to rapidly adapt to the new hosts.

Additionally, here I argue about two experimental procedures that might have influenced our observations. First is the dosage effect arising from the inoculation of *A. thaliana* plants with unknown quantity of virus particles of a particular genotype. The quantity of virus particles of a particular genotype within sap that served as inocula was not measured. Thus, the inocula doses were, most probably, not equal for all virus genotypes that might have influenced our observations. Since the plants were very small when inoculated, it was physically impossible to do the inoculation with more than 4 μl of sap. Inoculation with lees quantity might lead to underestimation of the infectivity and fitness due to imprecision inherently biased to the inoculation procedure. Previous works have shown that the effective number of viruses colonizing the host is low (González-Jara *et al* 2009), but can depend on the virus doses (Zwart *et al.* 2012). Secondly, but not less important, leaving virus genotypes to grow for ten days in tobacco plants might have led to appearance of some other mutations so that polymorphic populations. Still, previous work (Torres-Barceló, *pers.commun.*) showed that TEV populations in *N .tabacum* is relatively stable meaning that the virus does not fix new mutations even after more than 20 serial passages.

SUMMARY

One of the major threats to human and animal health as well as to agronomy is the emergence of new infectious diseases, most of which are caused by RNA viruses. It is a complex, multilevel problem that consists in acquisition of genetic variation by mutation or recombination within a virus population in the reservoir host that would enable the host-switch. RNA viruses show a remarkable evolvability owed to their large population sizes, short generation times and high mutation and recombination rates. Understanding the underlying evolutionary mechanisms by which a virus may become an emergent one is pivotal for the rational design of control strategies and antiviral therapies.

Thus, can virus emergence and jump to a new host species be predicted by knowing its phenotype, i.e., fitness, in its natural host? If so, then the architecture of virus fitness would be determined only by its genotype (the *G* component, *i.e.*, mutation) and environment (the *E* component, *i.e.*, host). Still, interactions between these components may exist and compromise the predictability of virus phenotype in an alternative host. How mutations affect the fitness of viral populations is essential to understanding viral emergence and adaptation to a new host.

The widespread observation that the majority of mutations are deleterious coincides with the theoretical prediction that an organism will be well adapted to its particular environment, so that any genomic change would represent a move-away from the optimal phenotype. In reality, this view is overly simplified; mutational fitness effects constitute a continuum and are conditional upon the environment and genetic background; effects commonly referred to as genotype-by-environment ($G \times E$) genetic ($G \times G$ or epistasis) and $G \times G \times E$ (epistasis-by-host) interactions. In this Thesis the contribution of these components to the architecture of viral fitness was addressed using a plant positive sense RNA virus, TEV. RNA viruses form the largest group of plant viruses and cause many economically important diseases.

First, to infer $G \times E$, we characterized the distribution of mutational fitness effects (DMFE) for a collection of twenty single-nucleotide substitution mutants of TEV across a set of eight environments represented by different hosts. Five of these host species were naturally infected by TEV, all belonging to family *Solanaceae*, whereas 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, thus containing a significant proportion of beneficial mutations. Within *Solanaceae* the tail of the distribution was drawn out more to the left side, thus comprising an excess of deleterious mutations. 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.

Secondly, to infer epistasis, we generated 53 TEV genotypes carrying pairs of single nucleotide substitutions and measured their separated and combined deleterious fitness effects in its primary host: *Nicotiana tabacum*. We found up to 38% of pairs had significant epistasis for fitness, including both positive and negative deviations from the null hypothesis of multiplicative effects. Moreover, we found the predominance of cases of reciprocal sign epistasis, indicating that adaptive landscapes for RNA viruses maybe highly rugged. Finally, we found that the magnitude of epistasis correlated negatively with the average effect of mutations. Negative correlation observed between epistasis and deleterious fitness effects indicates low genetic robustness of the compact RNA virus genomes. These observations are bad news regarding predictability of which viral genotypes may be more prone to emerge.

Thirdly, in order to characterize the degree to which epistatic effects vary across hosts, we quantified the fitness of ten TEV genotypes carrying pairs of single nucleotide substitutions across four hosts that differ from the primary host in taxonomic proximity. We found that epistasis among a particular pair of mutations is host-dependent and positive, on average, in natural host and weaker in more distant ones. The existence of epistasis and its variation across hosts makes the effect of individual mutations unpredictable.

Finally, we empirically characterized the fitness landscape of TEV adapted to *A. thaliana* by reconstructing the major part of intermediate and final genotypes of adaptation and measuring their fitness components in the new host. We found prevalent magnitude epistatic effects among the beneficial mutations, especially in the first steps of adaptation. Epistasis is small in the more adapted genotypes. Epistatic interactions included both positive and negative deviations from the expected values.

The landscape topography is predominately neutral and single-peaked. The smoothness of TEV landscape in *A. thaliana* suggests that virus easily adapted to its new host. This is an important result from the perspective of emerging viral diseases and in concordance to common observation that majority of emerging viruses are indeed RNA viruses.

RESUMEN EN CASTELLANO

Una de las mayores amenazas tanto para la salud humana y animal, como para la agronomía es la emergencia de nuevas enfermedades infecciosas, la mayoría de las cuales están causadas por los virus de RNA. La emergencia viral es un problema complejo que consista en la adquisición de la variación genética, por mutación o recombinación, dentro de la población viral en el huésped reservorio la cual podría facilitar la capacidad de infectar de manera eficiente nuevos huéspedes. Los virus de RNA presentan a una evolucionabilidad extraordinaria por sus grandes tamaños poblacionales, cortos tiempos de generación y altas tasas de mutación y recombinación. Comprender los mecanismos evolutivos que podrían dar lugar a un virus emergente es imprescindible para hacer un diseño racional de las estrategias del control y las terapias antivirales.

¿Podríamos predecir la emergencia de un virus y su salto a un huésped nuevo sabiendo el fenotipo del virus, es decir, su eficacia biológica, en su huésped natural? Si es así, la arquitectura de la eficacia biológica del virus estaría determinada solamente por su genotipo (la componente *G*, o en otras palabras, la mutación) y el ambiente (la componente *E* o, en otras palabras, el huésped). Sin embargo, podrían existir interacciones entre estos componentes y comprometer la predictibilidad del fenotipo del virus en un huésped alternativo. Saber cómo las mutaciones afectan la eficacia biológica de las poblaciones virales es esencial para llegar a entender la emergencia viral y la posterior adaptación del nuevo virus a su nuevo hospedador.

La observación común de que la mayoría de las mutaciones son deletéreas coincide con la predicción teorética de que un organismo ya está adaptado a su ambiente en particular, por lo cual, cualquier cambio genético supondría apartarse del fenotipo optimo. En realidad, este punto de vista es demasiado simplificado; los efectos mutacionales sobre la eficacia biológica constituyen un continuo y están condicionados tanto por el ambiente como por el fondo genético. Estos efectos se nombran comúnmente como interacción genotipo-ambiente ($G \times E$), interacción genética ($G \times G$ o epistasia) e interacción $G \times G \times E$ (epistasia-huésped). En esta Tesis, la contribución de estos componentes a la arquitectura de la eficacia biológica de un virus de RNA de la cadena positiva que infecta plantas, TEV. Los virus de RNA forman el grupo más grande de los virus de las plantas y causan muchas enfermedades económicamente importantes.

En primer lugar, para inferir $G \times E$, hemos caracterizado la distribución de los efectos mutacionales sobre la eficacia biológica (DMFE) de TEV utilizando una colección de veinte mutantes simples que tenían una única sustitución nucleotídica. Los efectos mutacionales se midieron en un conjunto de ambientes representados por ocho huéspedes distintos. Cinco de estos huéspedes pertenecían a la familia *Solanaceae* y se infectaban con TEV de la forma natural. Otros tres huéspedes pertenecían a otras tres familias de las plantas y eran parcialmente susceptibles a la infección con TEV. Primero, hemos encontrado una interacción significativa entre el genotipo del virus y la especie del huésped, generada por diferencias en la varianza genética de la eficacia biológica y los efectos pleiotrópicos de las mutaciones entre huéspedes. Segundo, las DMFE eran profundamente distintos entre los huéspedes que pertenecían a *Solanaceae* y los que no. La exposición de los genotipos de TEV a las no-solanáceas resultó en una gran reducción en la eficacia biológica, mientras que la varianza permanecía constante y la asimetría de la distribución era positiva. Una asimetría positiva implica que hay más valores mayores que la media, es decir, la cola derecha de la distribución contiene una proporción significativa de mutaciones beneficiosas. Entre solanáceas, la cola izquierda de la distribución tenía mayor peso, indicando un exceso de las mutaciones deletéreas. Conjuntamente, este resultado muestra que TEV puede fácilmente extender su rango de huéspedes y mejorar su eficacia biológica en los nuevos huéspedes, y que conocer la eficacia biológica de un mutante en un huésped no nos permite extrapolar que su eficacia se mantenga en otro(s) huésped(es).

En segundo lugar, para inferir la epistasia, hemos generado 53 genotipos de TEV los cuales llevaban pares de únicas sustituciones nucleotídicas y hemos medido sus efectos deletéreos sobre la eficacia biológica por separado y en combinación en el huésped primario: *Nicotiana tabacum*. Hemos encontrado que el 38% de los pares mostraban epistasia significativa sobre la eficacia biológica. Las desviaciones de la hipótesis nula de efectos multiplicativos eran tanto positivas como negativas, aunque predominaban los casos de epistasia de signo reciproco, lo cual indica de que el paisaje adaptativo de TEV debe de ser muy rugoso. Por último, hemos observado que la epistasia de magnitud correlaciona negativamente con el efecto promedio de las mutaciones, lo cual revela baja robustez genética de los genomas compactos de los virus de RNA. Estas observaciones son malas noticias con respeto a la predictibilidad de cuales genotipos virales podrían ser más propensos a emerger.

En tercer lugar, con el objetivo de caracterizar el grado de variación de los efectos epistaticos con los huéspedes en los que estos se miden, hemos cuantificado la eficacia de diez genotipos de TEV que llevaban pares de únicas sustituciones nucleotídicas a través de cuatro huéspedes que variaban en su proximidad taxonómica con el huésped primario. Hemos observado que la epistasia entre pares de mutaciones era dependiente del huésped y, en promedio, positiva en el huésped natural, haciéndose más débil a medida que el parentesco genético de estos huéspedes disminuía respecto al primario. La existencia de la epistasia y su variación entre huéspedes hace que los efectos de las mutaciones individuales sean aun más imprevisibles.

Por último, hemos caracterizado empíricamente el paisaje de la eficacia del TEV adaptado a *A. thaliana* reconstruyendo la mayor parte de los genotipos intermediarios posibles durante el proceso de adaptación y hemos medido dos componentes de la eficacia biológica en el nuevo huésped. En nuestro conjunto de datos predominaba la epistasia de magnitud entre las mutaciones beneficiosas, especialmente en los primeros pasos de la adaptación. La epistasia era pequeña en los genotipos más adaptados. Las interacciones epistaticas consistían en desviaciones de los valores esperados que eran ambos positivos y negativos. La topografía del paisaje era predominantemente neutral y consistía de un único pico. La suavidad del paisaje de TEV en *A. thaliana* sugiere que el virus se haya adaptado fácilmente a su nuevo huésped. Este resultado es importante desde la perspectiva de las enfermedades virales emergentes y en concordancia con la observación común de que la mayoría de los virus emergentes son efectivamente los virus de RNA.

BIBLIOGRAPHY

Adams, M.J., J.F. Antoniw, and F. Beaudoin (2005) Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Mol. Plant Pathol.* 6: 471–487.

Ahlquist, P., A.O. Noueiry, W-M. Lee, D.B. Kushner, and B.T. Dye (2003). Host factors in positive-strand RNA virus genome replication *J. Virol.* 77: 8181–8186.

Agrawal, A.A. (2001) Phenotypic plasticity in the interactions and evolution of species. *Science* 294: 321–326.

Agrawal, A.F., and M.C. Withlock (2010) Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? *Trends Ecol. Evol*. 25: 450–458.

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

Anderson, P.K., A.A. Cunningham, N.G. Patel, F.J. Morales, P.R. Epstein, and P. Daszak (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19: 535–544.

Auld, J.R., A.A. Agrawal, R.A. Relyea (2010) Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proc. R. Soc. B.* 277: 503-511.

Ayme, V., S. Souche, C. Caranta, M. Jacquemond, J. Chadoeuf, A. Palloix, and B. Moury (2006) Different mutations in the genome-linked VPg of *Potato virus Y* confers virulence on the *pvr2(3)* resistance in pepper. *Mol. Plant-Microbe Interact*. 19: 557-563.

Bagheri, H.C., and G.P. Wagner (2004). Evolution of dominance in metabolic pathways. *Genetics* 168: 1716-1735.

Baranowski, E., C.M. Ruíz-Jarabo, and E. Domingo (2001) Evolution of cell recognition by viruses. *Science* 292: 1102–1105.

Barr, J.N., and Fearns R. (2010) How RNA viruses maintain their genome integrity. *J. Gen. Virol.* 91: 1373-1387.

Barrell, B.G., G.M. Air, and C.A. Hutchison (1976) Overlapping genes in bacteriophage Φ174. *Nature* 264: 34–41.

Bateson, W. (1909) *Mendel's Principles of Heredity*. Cambridge University Press, Cambridge.

Baulcombe D. (2004) RNA silencing in plants. *Nature* 431: 356–363.

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

Bedoya, L.C., and J.A. Daròs (2010) Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149: 234-240.

Belshaw, R., O.G. Pybus, and A. Rambaut (2007) The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res*. 17: 1496–1504.

Belshaw, R., A. Gardner, A. Rambaut, and O.G. Pybus (2008) Pacing a small cage: mutation and RNA viruses. *Trends Ecol. Evol.* 23: 188–193.

Bershtein, S., M. Segal, R. Bekerman, N. Tokuriki, and D.S. Tawfik (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444: 929–932.

Betancourt, A.J. (2010). Lack of evidence for sign epistasis between beneficial mutations in an RNA bacteriophage. *J. Mol. Evol.* 71: 437–443.

Boevink, P., K.J. Oparka (2005) Virus-host interactions during movement processes. *Plant Physiol.* 138: 1815–1821.

Bonhoeffer, S., C. Chappey, N.T. Parkin, J.M. Whitcomb, and C.J. Petropoulos (2004) Evidence for positive epistasis in HIV-1. *Science* 306: 1547-1550.

Burch C.L., and L. Chao (1999) Evolution by small steps and rugged landscapes in the RNA virus Φ6. *Genetics* 151: 921–927.

Burch, C.L., and L. Chao (2000) Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* 406: 625–628.

Burch, C. L., and L. Chao (2004) Epistasis and its relationship to canalization in the RNA virus Φ6. *Genetics* 167: 559-567.

Burch, C.L., C. Guyader, D. Samarov, and H. Shen (2007). Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus Φ6. *Genetics* 176: 467–476.

Burgyan J., and Z. Havelda (2011) Viral suppressors of RNA silencing. *Trends Plant Sci*. 16: 265–272.

Callaway, A., D. Giesman-Cookmeyer, E.T. Gillock, T.L. Sit, S.A. Lommel (2001). The multifunctional capsid proteins of plant RNA viruses. *Annu. Rev. Phytopathol.* 39: 419–60.

Cameron, C.E., M. Gotte and K.D. Raney (2009) *Viral Genome Replication*. Springer, New York, USA.

Cann, A.J. (2005) *Principles of Molecular Virology*, Elsevier Academic Press, San Diego, California, USA

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

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

Carrington, J.C., and D.D. Freed (1990) Cap-independent enhancement of translation by a plant potyvirus 5′ nontranslated region. *J. Virol.* 64:1590–1597.

Carrington, J.C., and K.L. Herndon, (1992). Characterization of the potyviral HC-Pro autoproteolytic cleavage site. *Virology* 187: 308–315.

Chao, L. (1990) Fitness of RNA virus decreased by Muller's ratchet. *Nature* 348: 454–455.

Chao, L., C.U. Rang, and L.E. Wong (2002) Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage Φ6. *J. Virol.* 76: 3276–3281.

Charron, C., M. Nicolaï, J.L. Gallois, C. Robaglia, B. Moury, A. Palloix, and C. Caranta (2008) Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. *Plant J.* 54: 56-68.

Chirico, N., A. Vianelli, and R. Belshaw (2010) Why genes overlap in viruses? *Proc. R. Soc. B* 277: 3809–3817.

Chou, H.H., H.C. Chiu, N.F. Delaney, D. Segre, and C.J. Marx (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332: 1190–1192.

Chung, B.Y.W., W.A. Miller, J.F. Atkins, and A.E. Firth (2008) An overlapping essential gene in the *Potyviridae*. *Proc. Natl. Acad. Sci. USA* 105: 5897–5902.

Cleaveland, S.C., M.K. Laurenson, and L.H. Taylor (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos. Trans. R. Soc. B* 356: 991–999.

Cleaveland, S.C, D.T. Haydon, and L.H. Taylor (2007) Overviews of pathogen emergence: which pathogens emerge, when and why? *Curr. Top. Microbiol. Immunol.* 315: 85-111.

Codoñer, F.M., J.A. Darós, R.V. Solé, and S.F. Elena (2006) The fittest versus the flattest: Experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog* 2: e136.

Coffey, L.L., N. Vasilakis, A.C. Brault, A.M. Powers, F. Tripet, and S.C. Weaver (2008) Arbovirus evolution in vivo is constrained by host alternation. *Proc. Natl. Acad. Sci. USA*. 105: 6970–6975.

Coleman, J.R., D. Papamichail, S. Skiena, B. Futcher, E. Wimmer, and S. Mueller (2008) Virus attenuation by genome-scale changes in codon pair bias. *Science* 320: 1784–1787.

Cong, M., W. Heneine, and J.G. García-Lerma (2007) The fitness cost of mutations associated with *Human immunodeficiency virus* type 1 drug resistance is modulated by mutational interactions. *J. Virol.* 81: 3037-3041.

Coyne, J.A. (1992) Genetics and speciation. *Nature* 355: 511-515.

Crill, W. D., H. A. Wichman, and J. J. Bull (2000) Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154:27–37.

Crow, J.F., and Kimura, M. (1970) *An Introduction to Population Genetics Theory* Harper and Row: New York.

Cooper, L.A., and T.W. Scott (2001) Differential evolution of *Eastern equine encephalitis virus* populations in response to host cell type. *Genetics* 157:1403–1412.

Cuevas, J. M., S.F. Elena, and A. Moya (2002) Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162: 533-542.

Cuevas, J.M., P. Domingo-Calap, and R. Sanjuán (2011). The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol. Biol. Evol*. 29, 17–20.

Da Silva, J., M. Coetzer, R. Nedellec, C. Pastore, and D.E. Mosier (2010) Fitness epistasis and constraints on adaptation in a *Human immunodeficiency virus* type 1 protein region. *Genetics* 185: 293-303.

De la Iglesia, F., and S.F. Elena (2007) Fitness declines in *Tobacco etch virus* upon serial bottleneck transfers. *J. Virol.* 81: 4941-4947.

De la Peña, M., S.F. Elena, and A. Moya (2000) Effect of deleterious mutation-accumulation on the fitness of RNA phage MS2. *Evolution* 54: 686–691.

De Visser, J.A.G.M., and S.F. Elena (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. *Nat. Rev. Genet.* 8: 139-149.

De Visser, J.A.G.M., J. Hermisson, G.P. Wagner, L. Ancel Meyers, H. Bagheri-Chaichian, J.L. Blanchard, L. Chao, J.M. Cheverud, S.F. Elena, W. Fontana, G. Gibson, T.F. Hansen, D. Krakauer, R.C. Lewontin, C. Ofria, S.H. Rice, G. von Dassow, A. Wagner, and M.C. Whitlock (2003). Evolution and detection of genetic robustness. *Evolution* 57: 1959–1972.

De Visser JAGM, Cooper TF, and Elena SF (2011). The causes of epistasis. *Proc. R. Soc. B* 10: 3617-3624.

Deardorff, E.R., K.A. Fitzpatrick, G.V.S. Jerzak, P.Y. Shi, L.D. Kramer, and G.D. Ebel (2011). *West Nile virus* experimental evolution in vivo and the trade-off hypothesis. *PLoS Pathog.* 7:e1002335.

DeFilippis, V.R., and L.P. Villarreal, (2000). An introduction to evolutionary ecology of viruses. In C.J. Hurst, ed., *Viral Ecology*, pp. 126–208. Academic Press, New York.

Denhardt D.T., and R.B. Silver (1966) An analysis of the clone size distribution of φX174 mutants and recombinants. *Virology* 30: 10–19.

Dennehy J.J., Friedenberg N.A., Holt R.D., and P.E. Turner (2006) Viral ecology and the maintenance of novel host use. *Am. Nat.* 167: 429–439.

Draghi, J., T. Parsons, G. Wagner, and J. Plotkin (2010). Mutational robustness can facilitate adaptation. *Nature* 436: 353–355.

Desai, M.M., D. Weissman, and M.W. Feldman (2007) Evolution can favor antagonistic epistasis. *Genetics* 177: 1001–1010.

Ding, S.W., and O. Voinnet (2007) Antiviral immunity directed by small RNAs. *Cell* 130: 413–426.

Domingo, E., and J.J. Holland (1997) RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51: 151–178.

Domingo-Calap, P., J.M. Cuevas, and R. Sanjuán (2009). The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* 5: e1000742.

Drake, J.W., and J.J. Holland (1999) Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* 96:13910–13.

Duarte, E.A., D.K. Clarke, A. Moya, E. Domingo, and J.J. Holland (1992) Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl Acad. Sci. USA* 89: 6015–6019

Duffy, S., P.E. Turner, and C.L. Burch (2006) Pleiotropic costs of niche expansion in the RNA bacteriophage Φ6. *Genetics* 172: 751–757.

Duffy, S., C.L. Burch, and P.E. Turner (2007) Evolution of host specificity drives reproductive isolation among RNA viruses. *Evolution* 61: 2614–2622.

Duffy, S., Shackelton, L.A., and E.C. Holmes (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–276.

Edlund, J.A., and C. Adami (2004). Evolution of robustness in digital organisms. *Artif. Life* 10: 167-179.

Elena, S.F. (1999) Little evidence for synergism among deleterious mutations in a nonsegmented RNA virus. *J. Mol. Evol.* 49: 703-707.

Elena, S.F. (2002) Restrictions to RNA virus adaptation: An experimental approach. *Antonie van Leeuwenhoek*, 81:135–142.

Elena, S.F., P. Agudelo-Romero, P. Carrasco, F.M. Codoñer, S. Martín, C. Torres-Barceló, and R. Sanjuán (2008) Experimental evolution of plant RNA viruses. *Heredity* 100: 478–483.

Elena, S.F., P. Agudelo-Romero, and J. Lalić (2009) The evolution of viruses in multi-host fitness landscapes. *Open Virol. J.* 3: 1–6.

Elena, S.F., S. Bedhomme, P. Carrasco, J.M. Cuevas, F. de la Iglesia, G. Lafforgue, J. Lalić, À. Pròsper, N. Tromas, and M.P. Zwart (2011). The evolutionary genetics of emerging plant RNA viruses. *Mol. Plant-Microb. Interact*. 24: 287–293.

Elena, S.F., P. Carrasco, J.A. Daròs, and R. Sanjuán (2006) Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* 7: 168–173.

Elena, S.F., V.S. Cooper, and R.E. Lenski (1996a) Punctuated evolution caused by selection of rare beneficial mutations. *Science* 272: 1802–1804.

Elena S.F., M Dávila, I.S. Novella, J.J. Holland, E. Domingo, and A. Moya (1998) Evolutionary dynamics of fitness recovery from the debilitating effects of Muller's ratchet. *Evolution* 52: 309–314.

Elena, S.F., and R. Froissart (2010) New experimental and theoretical approaches towards the understanding of the emergence of viral infections. *Phil. Trans. R. Soc. B.* 365: 1867-1869.

Elena S.F., F. González-Candelas, I.S. Novella, E.A. Duarte, D.K. Clarke, E. Domingo, J.J. Holland, and A. Moya (1996b) Evolution of fitness in experimental populations of *Vesicular stomatitis virus*. *Genetics* 142: 673–679.

Elena S.F., and A. Moya (1999) Rate of deleterious mutation and the distribution of its effects on fitness in *Vesicular stomatitis virus*. *J. Evol. Biol.* 12:1078–1088.

Elena S.F., R. Sanjuán (2005) Adaptive value of high mutation rates of RNA viruses: separating causes from consequences *J. Virol.* 79: 11555–11558.

Elena, S.F., and R. Sanjuán (2008) The effect of genetic robustness on evolvability in digital organisms. *BMC Evol. Biol*. 8: 284.

Elena, S.F., R.V. Solé, and J. Sardanyés (2010) Simple genomes, complex interactions: epistasis in RNA virus. *Chaos* 20: 026106.

Eyre-Walker, A., and P.D. Keightley (2007) The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8: 610–618.

Edwards, R.A., and F. Rohwer (2005) Viral Metagenomics. *Nat. Rev. Microbiol.* 3: 504–510.

Falconer, D.S., (1989) *Introduction to Quantitative Genetics*, Ed. 3. Longman, New York.

Félix, M., and A. Wagner (2008) Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity* 100: 132–140.

Ferris, M.T., P. Joyce, and C.L. Burch (2007) High frequency of mutations that expand the host range of an RNA virus. *Genetics* 176: 1013-1022.

Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edin.* 52: 399-433.

Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*. Oxford, Clarendon Press.

Fisher, R.A. (1934) Adaptation and mutations. *School Sci. Rev.* 15: 294-301.

Franke, J., A. Klözer, J.A.G.M. de Visser, and J. Krug (2011) Evolutionary accessibility of mutational pathways. *PLoS Comp. Biol.* 7: e1002134.

French, R., and D.C. Stenger (2003) Evolution of *Wheat streak mosaic virus*: dynamics of population growth within plants may explain limited variation. *Annu. Rev. Phytopathol.* 41: 199–214.

Fry, J.D. (1996) The evolution of host specialization: are tradeoffs overrated? *Am. Nat.* 148: S84–S107.

Fry, J.D., S.L. Heinsohn, and T.F.C. Mackay (1996). The contribution of new mutations to genotype-environment interaction for fitness in *Drosophila melanogaster*. *Evolution* 50:2316–27

Futuyma, D.J. and G. Moreno (1988) The evolution of ecological specialization. *Annu. Rev. Ecol. Syst*. 19: 207–233.

Gabrenaite-Verkhovskaya, R., I.A. Andrew, N.O. Kalinina, L. Torrance, M.E. Taliansky, and K Mäkine (2008) Cylindrical inclusion protein of *Potato virus A* is associated with a subpopulation of particles isolated from infected plants. *J. Gen. Virol.* 89: 829–838.

Gago, S., S.F. Elena, R. Flores, and R. Sanjuán (2009) Extremely high mutation rate of a hammerhead viroid. *Science* 323:1308.

Gallie, D.R. (2001) Cap-independent translation conferred by the 5′ leader of *Tobacco etch virus* is eukaryotic initiation factor 4G dependent. *J. Virol*. 75: 12141–12152.

Gallie, D.R., and K.S. Browning (2001) eIF4G functionally differs from eIFiso4G in promoting internal initiation, cap-independent translation, and translation of structured mRNAs. *J. Biol. Chem.* 276: 36951–36960.

Gandon, S. (2004) Evolution of multihost parasites. *Evolution* 58: 455–469.

Gao, H., and M.W. Feldman (2009) Complementation and epistasis in viral coinfection dynamics. *Genetics* 182: 251–263.

García-Arenal, F., A. Fraile, J.M. Malpica (2001) Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* 39: 157–186.

Gavrilets, S. (2003) Evolution and speciation in a hyperspace: the roles of neutrality, selection, mutation, and random drift. In J.P. Crutchfield and P. Schuster (eds.), *Evolutionary Dynamics*, pp.135-162, Oxford University Press, Inc.

Gavrilets, S. (2010) High-dimensional fitness landscapes and the origins of biodiversity. In M. Pigliucci and G. Muller (eds), *Toward an Extended Evolutionary Synthesis*. MIT Press, Cambridge, MA.

Gerrish, P.J.R., and R.E. Lenski (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* 102/103: 127–144.

Ghedin, E., N.A. Sengamalay, M. Shumway, J. Zaborsky, T. Feldblyum, V. Subbu, D.J. Spiro, J. Sitz, H. Koo, P. Bolotov, D. Dernovoy, T. Tatusova, Y. Bao, K. St George, J. Taylor, D.J. Lipman, C.M. Fraser, J.K. Taubenberger, and S.L. Salzberg (2005) Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437: 1162–1166.

Gibbs, A. J., C. H. Calisher and F. Garcia-Arenal (1995) *Molecular Basis of Virus Evolution*. Cambridge University Press, Cambridge, United Kingdom.

Gillespie, J.H., and M. Turelli (1989) Genotype-environment interactions and the maintenance of polygenic variation. *Genetics* 121: 129–138.

González-Jara, P., A. Fraile, T. Canto, and F. García-Arenal (2009) The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J. Virol.* 83: 7487–7494.

Gorbalenya, A.E., L. Enjuanes, J. Ziebuhr, and E.J. Snijder (2006) Nidovirales: Evolving the largest RNA virus genome. *Virus Res*. 117: 17–37.

Govier, D.A., B. Kassanis, and T.P. Pirone (1977) Partial purification and characterization of the *Potato virus Y* helper component. *Virology* 78: 306–314.

Greene, I.P., E.Y. Wang, E.R. Deardorff, R. Milleron, E. Domingo, and S.C. Weaver (2005) Effect of alternating passage on adaptation of Sindbis virus to vertebrate and invertebrate cells. *J Virol.* 79:14253–14260.

Gupta, A. P. and R. C. Lewontin (1982) A study of reaction norms in natural populations of *Drosophila pseudoobscura*. *Evolution* 36: 934–948.

Harries, P., and B. Ding (2011) Cellular factors in plant virus movement: at the leading edge of macromolecular trafficking in plants. *Virology* 411: 237–243.

Hodgins-Davies, A., and J.P. Townsend (2010) Evolving gene expression: from *G* to *E* to *G*×*E*. *Trends Ecol. Evol.* 24: 649-658.

Holmes, E.C. (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol*. 11: 543–546.

Holmes, E.C. (2009) The evolutionary genetics of emerging viruses. *Annu. Rev. Ecol. Evol. Syst.* 40: 353–72.

Holmes, E.C. (2009a). *The Evolution and Emergence of RNA Viruses*. Oxford University Press, Oxford, USA

Holmes, E.C., and A.J. Drummond (2007) The evolutionary genetics of viral emergence. *Curr. Top. Microbiol. Immunol.* 315: 51–66.

Holmes, E.C., and A. Rambaut (2004) Viral evolution and the emergence of SARS coronavirus. *Philos. Trans. R. Soc. Lond. B* 359: 1059–1065.

Hull, R. (2002) *Plant Virology* (4th edition). Academic Press, London, UK.

Johnson, K. P., Malenke, J. R. and D.H. Clayton (2009) Competition promotes the evolution of host generalists in obligate parasites. *Proc. R. Soc. B* 276: 3921–3926.

Jones, R.A.C. (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.

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

Kasschau, K.D., and J.C. Carrington (1998) A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. *Cell* 95: 461–470.

Kassen, R., (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* 15: 173–190.

Kawecki, T.J. (1994). Accumulation of deleterious mutations and the evolutionary cost of being a generalist. *Am. Nat.* 144: 833–838.

Kawecki, T.J. (1998) Red queen meets Santa Rosalia: arms races and the evolution of host specialization in organisms with parasitic lifestyles. *Am. Nat.* 152: 635–651

Kawecki, T.J., and D. Ebert (2004) Conceptual issues in local adaptation. *Ecol. Lett.* 7: 1225–1241.

Khan, M.A., H. Miyoshi, D.R. Gallie, and D.J. Goss (2008) Potyvirus genome-linked protein, VPg, directly affects wheat germ in vitro translation: interactions with translation initiation factors eIF4F and eIFiso4F. *J. Biol. Chem.* 283: 1340–1349.

Kimura, M. (1968). Genetic variability maintained in a finite population due to the mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* 11, 247–269.

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press

King, A.M.Q., Adams, M.J., Carstens, E.B. and E.J. Lefkowitz (2012) Virus taxonomy: classification and nomenclature of viruses. In: *Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier, San Diego, USA

Kirschner, and M. J. Gerhart (1998) Evolvability *Proc. Natl. Acad. Sci. USA* 95: 8420–8427.

Kondrashov, A.S. (1994) Muller's ratchet under epistatic selection. *Genetics* 136: 1469–1473.

Kondrashov, A.S., and J.F. Crow (1991) Haploidy or diploidy: which is better. *Nature* 351: 314-315.

Kondrashov, A.S., and D. Houle (1994) Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proc. R. Soc. B*. 258: 221–227.

Kondrashov, F.A., and A.S. Kondrashov (2001) Multidimensional epistasis and the disadvantage of sex. *Proc. Natl Acad. Sci. USA* 98: 12089–12092.

Koonin, E.V. (2012) *The Logic of Chance: The Nature and Origin of Biological Evolution*, FT Press Science, NJ, USA

Korona, R. (1999) Genetic load of the yeast *Saccharomyces cerevisiae* under diverse environmental conditions. *Evolution* 53: 1966–1971.

Kouyos, R.D., O.K. Silander, and S. Bonhoeffer (2007). Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol. Evol.* 6: 308–315.

Krakauer, D., and J.B. Plotkin (2002). Redundancy, anti-redundancy, and the robustness of genomes. *Proc. Nac. Acad. Sci.* 99: 1405–1409.

Kvitek, D.J., and G. Sherlock (2011) Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* 7: e1002056.

Lajeunesse M.J., and M.R. Forbes (2001) Host range and local parasite adaptation. *Proc. R. Soc. B* 269: 703–710.

Lalić, J., P. Agudelo-Romero, P. Carrasco, and S.F. Elena (2010) Adaptation of *Tobacco etch potyvirus* to a susceptible ecotype of *Arabidopsis thaliana* capacitates it for systemic infection of resistant ecotypes. *Phil. Trans. R. Soc. B* 65: 1997–2008.

Lalić, J., J.M. Cuevas, and S.F. Elena (2011) Effect of host species on the distribution of mutational effects for an RNA virus. *PLoS Genet*. 7, e1002378.

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

Lande, R. (1976) The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* 26: 221–235.

Lenski, R. E. (1991). Quantifying fitness and gene stability in microorganisms. Pp. 173-192 in L. R. Ginzburg (ed). *Assessing Ecological Risks of Biotechnology*. Butterworth-Heinemann, Boston, USA.

Lenski, R.E., J. Barrick, and C. Ofria (2006) Balancing robustness and evolvability. *PLoS Biol.* 4: e428.

Lenski, R.E., and M. Travisano (1994) Dynamics of adaptation and diversification: a 10,000 generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. USA* 91: 6808–6814.

Lenski, R.E., M.R. Rose, S.C. Simpson, and S.C. Tadler (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* 138: 1315–1341.

Lewontin, R.C. (1974) Annotation: The analysis of variance and the analysis of causes. *Amer. J. Hum. Gen.* 26: 400–411.

Llave, C., K.D. Kasschau, and J.C. Carrington (2000) Virus-encoded suppressor of posttranscriptional gene silencing targets a maintenance step in the silencing pathway. *Proc. Natl. Acad. Sci. USA* 97: 13401–13406.

Loewe, L., and W.G. Hill (2010) The population genetics of mutations: good, bad and indifferent. *Phil. Trans. R. Soc. B* 365: 1153–1167.

Longdon, B., J.D. Hadfield, C.L. Webster, D.J. Obbard, F.M. Jiggins (2011) Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog*. 7: e1002260.

Luria, S.E. (1951) The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. *Cold Spring Harbor Symp. Quant. Biol.* 16: 463–470.

Lynch, M., and B. Walsh, B (1998) *Genetics and Analysis of Quantitative Traits*, pp. 745–751. Sinauer Associates, Inc., Sunderland, MA.

Macía, J., R.V. Solé, and S.F. Elena (2012) The causes of epistasis in genetic networks. *Evolution* 66: 586–596.

Maclean, R.C. (2005) Adaptive radiation in microbial microcosms. *J. Evol. Biol.* 18: 1376–1386.

Maia, I.G., A.L. Haenni, and F. Bernardi (1996) Potyviral HC-Pro: a multifunctional protein. *J. Gen. Virol*. 77: 1335–1341.

Maisnier-Patin, S., O.G. Berg, L. Lijas, and D.I. Andersson (2002) Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol. Microbiol.* 46: 355-366.

Marsh, G.A., R. Rabadán, A.J. Levine, and P. Palese (2008) Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. *J. Virol.* 82: 2295–304.

Martin, G.B., A.J. Bogdanove and G. Sessa (2003) Understanding the functions of plant disease resistance proteins. *Annu. Rev. Plant Biol*. 54: 23–61.

Martin, G., and T. Lenormand (2006) The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution* 60: 2413-2427.

Martínez, F., J. Sardanyés, S.F. Elena, and J.A. Daròs. (2011) Dynamics of a plant RNA virus intracellular accumulation: stamping machine *vs*. Geometric replication. *Genetics* 188:637– 646.

Martínez, J.P., G. Bocharov, A. Ignatovich, J. Reiter, M.T. Dittmar, S. Wain-Hobson, and A. Meyerhans (2011) Fitness ranking of individual mutants drives patterns of epistatic interactions in HIV-1. *PLoS ONE* 6: e18375.

Martínez-Picado, J., and M.A. Martínez (2009) HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and *ex vivo*. *Virus Res.* 134: 104-123.

McBride, R.C., C.B. Ogbunugafor, and P.E. Turner (2008) Robustness promotes evolvability of thermotolerance in an RNA virus. *BMC Evol. Biol.* 8: 231

McDonald, B.A., and Linde, C. (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40: 349–379.

Melcher, U. (2012) Virus operation control centers. In: *Biocommunication of Plants, Signaling and Communication in Plants* 14, 231–254, Springer Berlin Heidelberg

Minskaia, E., T. Hertzig, A. E. Gorbalenya, V. Campanacci, C. Cambillau, B. Canard, and J. Ziebuhr (2006) Discovery of an RNA virus 3′→5′ exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. USA* 103: 5108–5113.

Miralles R., A. Moya, and S.F. Elena (1999) Effect of population patchiness and migration rates on the adaptation and divergence of *Vesicular stomatitis virus* quasispecies populations. *J. Gen. Virol.* 80: 2051–2059.

Miralles R., A. Moya, S.F. Elena (2000) Diminishing returns of population size in the rate of RNA virus adaptation. *J. Virol.* 74: 3566–3571.

Molla, A., M. Korneyeve, Q. Gao, S. Vasavanonda, P.J. Schipper, H.M. Mo, M. Markowitz, T. Chernyavskiy, P. Niu, N. Lyons, A. Hsu, G.R. Granneman, D.D. Ho, C.A. Boucher, J.M. Leonard, D.W. Norbeck, and D.J. Kempf (1996) Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat. Med.* 2: 760-766.

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

Montville, R., R. Froissart, S.K. Remold, O. Tenaillon, and P.E. Turner (2005) Evolution of mutational robustness in an RNA virus. *PLoS Biol* 3: 1939–1945.

Mukai, T. (1964) The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50: 1–19.

Muller, H. J. (1964) The relation of recombination to mutational advance. *Mut. Res.* 1: 2-9.

Navas-Castillo, J., E. Fiallo-Olive, and S. Sánchez-Campos (2011) Emerging virus diseases transmitted by whiteflies. *Annu. Rev. Phytopathol*. 49: 219–248.

Novella I.S., M. Cilnis, S.F. Elena, J. Kohn, A. Moya, E. Domingo E., and J.J. Holland (1996) Large-population passages of vesicular stomatitis virus in interferon-treated cells select variants of only limited resistance. *J. Virol.* 70: 6414–6417.

Novella I.S., E.A. Duarte, S.F. Elena, A. Moya, E. Domingo, and J.J. Holland (1995) Exponential increases of RNA virus fitness during large population transmissions. *Proc. Natl. Acad. Sci. USA* 92: 5841–5844.

Novella, I.S., C.L. Hershey, C. Escarmis, E. Domingo, and J.J. Holland (1999). Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J. Mol. Biol.* 287: 459–465.

Novella I.S., J. Quer, E. Domingo, and J.J. Holland (1999) Exponential fitness gains of RNA virus populations are limited by bottleneck effects. *J. Virol.* 73: 1668–1671.

Novella, I.S., S. Zárate, D. Metzgar, and B.E. Ebendick-Corpus (2004) Positive selection of synonymous mutations in *Vesicular stomatitis virus*. *J. Mol. Biol.* 342: 1415-1421.

Nowak, M.A., M.C. Boerlijst, J. Cooke, J. Maynard Smith (1997) Evolution of genetic redundancy. *Nature* 388: 167–171.

Ohta, T. (1992) The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol.* Syst. 23: 263–286.

Orr, H. A. (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52: 935–949.

Orr, H.A. (1999) The evolutionary genetics of adaptation: a simulation study. *Genet. Res.* 74: 207–214.

Orr, H.A. (2000) Adaptation and the cost of complexity. *Evolution* 54: 13–20.

Orr, H.A. (2002) The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 56: 1317–1330.

Oruetxebarria, L., D. Guo, A. Merits, K. Mäkinen, M. Saarma, and J.P.T. Valkonen (2001) Identification of the genome-linked protein in virion of *Potato virus A*, with comparison to other members in the genus *Potyvirus*. *Virus. Res*. 73: 101–112.

Parera, M., N. Pérez-Álvarez, B. Clotet, and M.A. Martínez (2009) Epistasis among deleterious mutations in the HIV-1 protease. *J Mol Biol* 392: 243–250.

Parrish, C.R., E.C. Holmes, D.M. Morens, E.C. Park, D.S. Burke, C.H. Calisher C.A. Laughlin, L.J. Saif, and P. Daszak (2008) Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev*. 72: 457–470.

Pathak, V.K., and H.M. Temin (1992) 5-Azacytidine and RNA secondary structure increase the retrovirus mutation rate. *J. Virol*. 66: 3093–3100.

Patwa, Z., and L. Wahl (2008) The fixation probability of beneficial mutations. *J. R. Soc. Interface* 5: 1279–1289.

Pepin, K.M., and H.A. Wichman (2007) Variable epistatic effects between mutations at host recognition sites in ϕX174. *Evolution* 67: 1710-1724.

Peris, J.B., P. Davis, J.M. Cuevas, M.R. Nebot, and R. Sanjuán (2010) Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics* 185: 603–609.

Pfaffl, M.V. (2004) Quantification strategies in real-time PCR. In *A-Z of Quantitative PCR* (ed. SA Bustin). International University Line: La Jolla, USA. Pp 87-112.

Phillips, P.C. (2008) Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9: 855–867.

Pirone, T.P., and S. Blanc (1996) Helper-dependent vector transmission of plant viruses. *Annu. Rev. Phytopath.* 34: 227–247.

Pita, J.S., J.R. de Miranda, W.L. Schneider, and M.J. Roossinck (2007) Environment determines fidelity for an RNA virus replicase. *J. Virol*. 81: 9072–9077.

Plotkin, J.B., and G. Kudla (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet*. 12: 32–42.

Poelwijk, F.J., D.J. Kiviet, D.M. Weinreich, and S.J. Tans (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445: 383-386.

Poelwijk, F.J., S. Tanase-Nicola, D.J. Kiviet, and S.J. Tans (2011) Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* 272: 141–144.

Pogue, G.P., C.C. Huntley, and T. C. Hall (1994) Common replication strategies emerging from the study of diverse groups of positive-strand RNA viruses. *Arch. Virol*. 9: S181–S194.

Poon, A.F.Y., and L. Chao (2006). Functional origins of fitness effect-sizes of compensatory mutations in the DNA bacteriophage ϕX174. *Evolution* 60: 2032–2043.

Proulx, S.R., and P.C. Phillips (2005) The opportunity for canalization and the evolution of genetic networtks. *Am. Nat.* 165: 147-162.

Pruss, G., X. Ge, X.M. Shi, J.M. Carrington, and V. Bowman (1997) Plant viral synergism: the potyviral genome encodes a broad range pathogenicity enhancer that transactivates replication of heterologous viruses. *Plant Cell* 9: 859–868.

Puustinen, P., M.L. Rajamäki, K.I. Ivanov, J.P.T. Valkonen, and K. Mäkinen (2002) Detection of the potyviral genome-linked protein VPg in virions and its phosphorylation by host kinases. *J. Virol*. 76: 12703–12711.

Rajamäki, M.L., T. Mäki-Valkama, K. Mäkinen, and J.P.T Valkonen (2004) Infection with potyviruses. In *Plant-Pathogen Interactions*: 68-91. Blackwell Publishing, Oxford, UK.

Rambaut, A., D. Posada, K.A. Crandall, E.C. Holmes (2004) The causes and consequences of HIV evolution. *Nat. Rev. Genet.* 5: 52–61.

Rancurel, C., M. Khosravi, K. Dunker, P. Romero, and D. Karlin (2009) Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J. Virol*. 83: 10719–10736.

Ratcliff, F.G., B.D. Harrison, and D.C. Baulcombe (1997) A similarity between viral defense and gene silencing in plants. *Science* 276: 1558–1560.

Remold, S. K., and R. E. Lenski (2001) Contribution of individual random mutations to genotype-by-environment interactions in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 98: 11388–11393.

Remold, S.K., and R.E. Lenski (2004) Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. *Nat. Genet.* 36: 423–426.

Remold, S.K., A. Rambaut, and P.E. Turner (2008) Evolutionary genomics of host adaptation in *Vesicular stomatitis virus*. *Mol. Biol. Evol*. 25: 1138–1147.

Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution* 43: 223–225.

Rico, P., P. Ivars, S.F. Elena, and C. Hernández (2006) Insights into the selective pressures restricting *Pelargonium flower break virus* genome variability: evidence for host adaptation. *J. Virol*. 80: 8124–8132.

Riechmann, J.L., S. Laín, J.A. García (1992) Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol*. 73:1–16.

Rojas, M.R., F.M. Zerbini, R.F. Allison, R.L. Gilbertson, and W.J. Lucas (1997) Capsid protein and helper component-proteinase function as potyvirus cell-to-cell movement proteins. *Virology* 237: 283–295.

Rokyta, D.R., C.J. Beisel, and P. Joyce (2006) Properties of adaptive walks on uncorrelated landscapes under strong selection and weak mutation. *J. Theor. Biol.* 243: 114–120.

Rokyta, D.R., P. Joyce, B. Caudle, C. Miller, C.J. Beisel, and H.A. Wichman (2011) Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS Genet.* 7: e1002075.

Roossinck, M.J. (2008) Mutant clouds and bottleneck events in plant virus evolution. In *Origin and Evolution of Viruses* 2nd edition (Domingo, E., Parrish C.R and J.J. Holland, eds.), pp. 254, Elsevier Academic Press, Amsterdam.

Roossinck, M.J. (2008a) *Plant Virus Evolution*, Springer, Berlin, Germany.

Ruiz-Medrano, R., F. Kragler, and S. Wolf (2012) Signaling and phloem-mobile transcripts. In: *Advances in Plant Biology* **3**:151–177, Springer New York, USA.

Saenz, P., B. Salvador, C. Simon-Mateo, K.D. Kasschau, J.C. Carrington, and J.A. García (2002) Host-specific involvement of the HC protein in the long-distance movement of potyviruses. *J. Virol*. 76: 1922–1931.

Salverda, M.L.M., E. Dellus, F.A. Gorter, A.J.M. Debets, J. Van der Oost, R.F. Hoekstra RF, D.S. Tawfik, and J.A.G.M de Visser (2011) Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* 7: e1001321.

Sanjuán, R. (2006) Quantifying antagonistic epistasis in a multifunctional RNA secondary structure of the *Rous sarcoma virus*. *J Gen Virol* 87: 1595–1602.

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.

Sanjuán, R., J.M. Cuevas, V. Furió, E.C. Holmes, and A. Moya (2007). Selection for robustness in mutagenized RNA viruses. *PLoS Genet.* 3: e93.

Sanjuán, R., J.M. Cuevas, A. Moya, and S.F. Elena (2005) Epistasis and the adaptability of an RNA virus. *Genetics* 170: 1001–1008.

Sanjuán, R., and S.F. Elena (2006) Epistasis correlates to genomic complexity. *Proc. Natl. Acad. Sci. USA* 103: 14402–14405.

Sanjuán, R., J. Forment, and S.F. Elena (2006) *In silico* predicted robustness of viroids RNA secondary structure. II. Interaction between mutation pairs. *Mol. Biol. Evol.* 23: 2123–2130.

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

Sanjuán, R., A. Moya, and S.F. Elena (2004a) The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. USA* 101: 15376–15379.

Sanjuán, R., and M.R. Nebot (2008) A network model for the correlation between epistasis and genomic complexity. *PLoS ONE* 3: e2663.

Sanjuán, R., M. R. Nebot, N. Chirico, L.M. Mansky, and R. Belshaw (2010) Viral mutation rates. *J. Virol.* 84: 9733–9748.

Schrag, S.J., V. Perrot, and B.R. Levin (1997) Adaptation to the fitness cost of antibiotic resistance in *E. coli*. *Proc. R. Soc. B* 264: 1287–1291.

Scherbakov, D.V., and M.B. Garber (2000) Overlapping genes in bacterial and phage genomes. *Mol. Biol*. *Evol*. 34: 485–495.

Schneider, W.L., and M.J. Roossinck (2001) Genetic diversity in RNA viral quasispecies is controlled by host-virus interactions. *J. Virol*. 75: 6566–6571

Segrè, D., DeLuna, A., Church, G.M., & Kishony, R. 2005 Modular epistasis in yeast metabolism. *Nat. Genet*. 37, 77-83.

Shukla, D.D., C. W. Ward, and A.A. Brunt (1994) *The Potyviridae*. Wallingford: CAB International.

Shukla, D.D., and C.W. Ward (1989) Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. *Adv. Virus Res.* 36: 273–314.

Simmonds, P.,and D.B. Smith (1999) Structural constraints on RNA virus evolution. *J. Virol*. 73: 5787–5794.

Simmonds, P., A. Tuplin, and D.J. Evans (2004) Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA* 10: 1337–1351.

Smith, J.M. (1999) *Evolutionary Genetics*, 2nd edition, Oxford University Press, UK.

Sniegowski, P.D., P.J. Gerrish, T. Johnson, and A. Shaver (2000) The evolution of mutation rates: separating causes from consequences. *Bioessays* 22: 1057–1066.

Sniegowski, P.D., and P.J. Gerrish (2010) Beneficial mutations and the dynamics of adaptation in asexual populations. *Phil. Trans. R. Soc. B* 365: 1255–1263.

Soltis, D.E., and P.S. Soltis (2000) Contributions of plant molecular systematics to studies of molecular evolution. *Plant Mol. Biol.* 42: 45-75.

Steinhauer, D.A., and J.J. Holland (1987) Rapid evolution of RNA viruses. *Annu. Rev. Microbiol.* 41: 409–433.

Thomas, J. H. (1993). Thinking about genetic redundancy. *Trends Genet.* 9: 395–399.

Timms, R. and A.F. Read (1999) What makes a specialist special? *Trends Ecol. Evol.* 14: 333–334.

Turner, P.E. and S.F. Elena (2000). Cost of host radiation in an RNA virus. *Genetics* 156: 1465–1470.

Turner, P.E., N.M. Morales, B.W. Alto, and S.K. Remold (2010) Role of evolved host breadth in the initial emergence of an RNA virus. *Evolution* 64: 3273–3286.

Urcuqui-Inchima, S., A.L. Haenni, and F. Bernardi (2001) Potyvirus proteins: a wealth of functions. *Virus. Res*. 74: 157–175.

Van Opijnen, T., M.C. Boerlijst, and B. Berkhout (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.

Van Valen, L. (1973). A new evolutionary law. *Evol. Theory* 1: 1–30.

Via, S. (1984) The quantitative genetics of polyphagy in an insect herbivore. I. Genotype-environment interaction in larval performance on different host plant species. *Evolution* 38: 881–895.

Via, S. (1984b) The quantitative genetics of polyphagy in an insect herbivore. II. Genetic correlations in larval performance within and among host plants. *Evolution* 38: 896–905.

Via, S. and R. Lande (1985) Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution* 39:505–522.

Via, S., R. Gomulkiewicz, G. de Jong, S.M. Scheiner, C.D. Schlichting, and P.H. van Tienderen (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol. Evol.* 10: 212–217.

Vijayapalani, P., M. Maeshima, N. Nagasaki-Takekuchi, and W.A. Miller (2012) Interaction of the trans-frame potyvirus protein P3N-PIPO with host protein PCaP1 facilitates potyvirus movement. *PLoS Pathog* 8: e1002639.

Voinnet, O. (2005) Induction and suppression of RNA silencing: Insights from viral infections. *Nat. Rev. Genet*. 6: 206–220.

Wagner, A. (2005) Robustness, evolvability, and neutrality. *FEBS Lett*. 579: 1772–1778.

Wagner, A. (2008) Robustness and evolvability: a paradox resolved. *Proc. R. Soc. Lond. B* 275: 91–100.

Wagner, G.P., and L. Altenberg (1996) Perspective: Complex adaptations and the evolution of evolvability. *Evolution* 50: 967–976.

Walkey, D. (1991) *Applied Plant Virology*. 2nd edition. Chapman and Hall, London.

Wallis, C.M., A.L. Stone, D.J. Sherman, V.D. Damsteegt, F.E. Gildow, and W.L. Schneider (2007) Adaptation of *Plum pox virus* to a herbaceous host (*Pisum sativum*) following serial passages. *J. Gen. Virol.* 88: 2839–2845.

Walsh, D., and I. Mohr (2011) Viral subversion of the host protein synthesis machinery. *Nat. Rev. Microb.* 9: 860–875.

Weaver, S.C., A.C. Brault, W.L. Kang, and J.J. Holland (1999). Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. *J. Virol.* 73: 4316–4326.

Webster, R.G., W.J. Bean, O.T. Gorman, T.M. Chambers, and Y. Kawaoka (1992) Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56: 152–179.

Weinreich, D.M. (2005) The rank ordering of genotypic fitness values predicts genetic constraints on natural selection on landscapes lacking sign epistasis. *Genetics* 171: 1397–1405.

Weinreich, D.M., N.F. Delaney, M.A. DePristo, and D.L. Hartl (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312: 111–114.

Weinreich, D.M., R.A. Watson, and L. Chao (2005) Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59: 1165–1174.

Welch, J.J., and D. Waxman (2005) The *nk* model and population genetics. *J. Theor. Biol.* 234: 329–340.

Withlock, M.C., P.C. Phillips, F.B.G. Moore, and S.J. Tonsor (1995). Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Evol. Syst.* 26: 601–629.

Wichman H.A., M.R. Badgett, L.A. Scott, C.M. Boulianne, J.J. Bull (1999) Different trajectories of parallel evolution during viral adaptation. *Science* 285: 422–424.

Wilke, C.O., and C. Adami (2001) Interaction between directional epistasis and average mutational effects. *Proc. R. Soc. B* 298: 1469–1474.

Wilke, C.O., and C. Adami (2003) Evolution of mutational robustness. *Mut. Res.* 522: 3–11.

Wilke, C.O., R.E. Lenski, and C. Adami (2003) Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evol. Biol.* 3: 3.

Wilke, C.O., J.L. Wang, C. Ofria, R.E. Lenski, and C. Adami (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412: 331–333.

Withlock, M.C., P.C. Phillips, F.B.G. Moore, and S.J. Tonsor (1995) Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Evol. Syst.* 26: 601–629.

Whitlock, M.C. (1996) The Red Queen beats the jack-of-all-trades: the limitations on the evolution of phenotypic plasticity and niche breadth. *Am. Nat.* 148: S65–S77.

Withlock, M.C., P.C. Phillips, F.B.G. Moore, and S.J. Tonsor (1995) Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Evol. Syst.* 26: 601–629.

Witwer, C., S. Rauscher, I.L. Hofacker, and P.F. Stadler (2001) Conserved RNA secondary structures in picornaviridae genomes. *Nucl. Acids Res.* 29: 5079–5089.

Woelk, C.H., and E.C. Holmes (2002) Reduced positive selection in vector borne RNA viruses. *Mol. Biol. Evol.* 19: 2333–2336.

Wolf, J.B., E.D. Brodie, and M.J. Wade (2000) *Epistasis and the Evolutionary Process*. Oxford, UK: Oxford University Press

Woolhouse, M.E.J., L.H. Taylor, and D.T. Haydon (2001). Population biology of multihost pathogens. *Science* 292: 1109–1112.

Woolhouse, M.E.J., and C. Dye (2001). Population biology of emerging and reemerging pathogens. *Phil. Trans. R. Soc. Lond. B* 356: 981–982.

Woolhouse, M.E.J. (2002). Population biology of emerging and re-emerging pathogens. *Trends Microbiol*. 10: S3–S7.

Woolhouse, M.E.J., J.P. Webster, E. Domingo, B. Charlesworth, and B.R. Levin (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat. Genet.* 32: 569–577.

Woolhouse, M.E.J., D.T. Haydon, and R. Antia (2005) Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.* 20: 238-244.

Woolhouse, M.E.J., and S. Gowtage-Sequeria (2005) Host range and emerging and reemerging pathogens. *Emerging Infect. Dis*. 11: 1842–1847.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* 16: 97–159.

Wright S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Intl. Congress Genetics* 1: 356–366.

Wright, S. (1977) *Evolution and the Genetics of Populations. Vol. 3. Experimental Results and Evolutionary Deductions*. Chicago University Press, Chicago.

Wylie, C.S., and E.I. Shakhnovich (2011) A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl. Acad. Sci. USA* 108: 9916–9921.

You, L., and J. Yin (2002) Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage T7. *Genetics* 160: 1273–1281.

Zeenko, V., D.R. Gallie (2005) Cap-independent translation of *Tobacco etch virus* is conferred by an RNA pseudoknot in the 5'-leader. *J. Biol. Chem.* 29: 26813–26824.

Zwart, M.P., J.A. Daròs JA, and S.F. Elena (2012). Effects of potyvirus effective population size in inoculated leaves on viral accumulation and the onset of symptoms. *J. Virol.* doi:10.1128/JVI.00909-12.