VNIVERSITAT DE VALÈNCIA

Evaluating fundamental life-history traits for *Tobacco etch potyvirus*

Nicolas Tromas

Programa Oficial de Postgrado en Biotecnología

Director

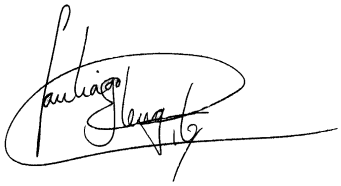
Prof. Santiago F. Elena Fito

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 D. Nicolas Tromas, Licenciado en Biología y Master en Microbiología por la Universidad de París VI, ha realizado bajo mi supervisión la tesis doctoral titulada "*Evaluating fundamental life-history traits for Tobacco etch potyvirus*".

Y para que así conste, firmo la presente en Valencia, a 22 de Febrero de 2013.



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

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Venir vivre dans un pays étranger, et associer ce changement à la réalisation d'une thèse a surement été, pour le moment, l'un des plus gros challenge de ma vie. La thèse est un apprentissage, aussi bien d'un point de vue professionnel que personnel. J'ai eu l'opportunité de travailler au cote de personnes avec qui j'ai pu intéragir, avoir des débats passionnés et aussi avec qui me confronter. J'ai eu la chance de travailler avec des experts dans leur domaine, qui ont su être patient et m'enseigner certaines bases fondamentales pour évoluer en tant que chercheur. Car la thèse est aussi un apprentissage du métier de chercheur, dans lequel nous devons apprendre à gérer aussi bien les projets, que leur succès ou leurs échecs. Au cours de ces années passées dans le laboratoire de Santiago, j'ai été en contact avec de nombreuses personnes qui m'ont aidé, m'ont guidé et m'ont soutenu. Toutes n'ont pas eut le même rôle mais j'ai appris de chacune d'elles. Je tiens donc a remercier sincèrement Alexandra, Ana, Àngels, Anouk, Clara, Guillermo, Jasna, José, José Antonio, Julia, Patricia, Paula

et Romain pour leurs conseils, leur amitié, et pour tout les moments que nous avons partagés. Paqui, je te remercie pour toute ton aide, pour nos discussions interminables, pour nos fous rires, pour tes conseils et surtout pour ton incroyable générosité. J'ai adoré discuter avec toi de politique, religion, science et autres sujets, tu es une personne brillante et j'ai hâte de lire ton premier livre pour enfant!

Mark, Stéphanie et Susana, je vous remercie pour vos nombreux conseils et votre patience a mon égard, pour ces discussions qui m'ont permis d'apprendre énormément, pour les bons moments que l'ont a passé et pour votre amitié. Guillaume, merci d'avoir partagé ces années avec moi, d'avoir supporté tant de fois mon mauvais caractère et d'être le seul a ce jour à avoir réussi à travailler en paillasse avec moi! Merci pour tous les bons moments qu'on a passé et qui ont rendu cette étape de ma vie beaucoup plus facile. Javi, que dire de plus que grâce a toi, j'ai passé des moments inoubliables sur Valencia, entre nos almuerzos, nos discussions enflammées sur l'avenir, tes conseils et ton aide quand j'en ai eu le plus besoin. Un énorme merci pour toi et a très vite sur Valencia ou en

Californie. Santiago, je ne sais pas encore comment tu as fais pour me supporter et être autant patient avec moi, je crois que ca restera une énigme pour moi. J'ai eu la chance d'apprendre les bases de l'Evolution à tes cotés, je sais qu'il me reste encore un long chemin mais grâce a toi j'ai appris que la créativité associée a l'organisation et à la gestion du temps donnaient de bien meilleurs résultats. Tu m'as donné l'opportunité de créer et gérer mes projets, tu m'as toujours soutenu dans mes choix même quand tu savais déjà qu'ils n'étaient pas forcement judicieux, tu m'as appris à mieux apprendre de mes erreurs. Pour tout cela et pour bien d'autres choses, je te remercie.

Un énorme merci a ma famille et a ma belle famille, qui m'ont toujours soutenu dans mes projets et qui m'ont toujours aidé. Vous m'avez maintes et maintes fois demandé si tout allait bien, si mes projets avançaient bien et vous avez toujours montre un intérêt et une préoccupation a mon égard. Sans vous, je sais très bien que le chemin aurait été beaucoup plus difficile et périlleux. Merci d'être la pour moi. Coralie, ou devrais-je dire future Mme Tromas, tu ne peux pas t'imaginer ce que tu représentes pour moi, ton importance dans ma vie,

ton implication dans mes choix, ton énergie dans ma motivation. Tu es ma moitié, et je te dédie la moitié de ce travail car sans toi, il n'aurait pas pu être accompli. J'ai une incroyable chance et je m'en rends compte chaque jour. Merci mon amour d'être la pour moi et d'être toi.

RESUMEN

Los virus de ARN son probablemente algunos de los parásitos más extendidos que se pueden encontrar en todas las formas de vida. Estos agentes infecciosos parecen particularmente propensos a causar enfermedades emergentes en plantas, seres humanos y otros animales. Sus habilidades para escapar al sistema inmune, evadir estrategias antivirales o para infectar a nuevas especies son una consecuencia directa de su enorme capacidad para evolucionar rápidamente. Comprender los procesos básicos del cambio evolutivo puede ser unos de los principales pasos necesarios en el diseño de nuevas estrategias de intervención.

Desde una perspectiva evolutiva, una de las principales características de los virus de ARN es su capacidad para mutar. El conocimiento de las tasas de mutación y el espectro molecular de mutaciones espontáneas son importantes para entender la evolución de la composición genética de las poblaciones virales. Estudios anteriores han demostrado que la tasa de mutaciones espontáneas de los virus de ARN varía

ampliamente entre 0.01 y 2 mutaciones por genoma y generación, ocupando los virus de plantas la parte inferior de esta escala de valores.

En este estudio, se propuso analizar el espectro mutacional y la tasa de mutación del *virus del grabado del tabaco* (*Tobacco etch virus*, TEV), como modelo para los virus de ARN de cadena positiva. Nuestro experimento minimiza la acción purificadora de la selección en el espectro mutacional, dando así una imagen exacta de qué tipo de mutación ha producido la replicasa viral. Hemos calculado la tasa de mutación espontánea de este virus, hallándose en el intervalo de valores entre 10^{-6} - 10^{-5} mutaciones por sitio y generación. Nuestras estimaciones se encuentran en el mismo rango de valores que los anteriormente descritos para otros virus ARN de plantas.

La recombinación de un virus de ARN es un parámetro evolutivo que contribuye significativamente a la diversidad y a la evolución de los virus. La tasa de recombinación depende de dos parámetros: de la frecuencia de intercambio genético entre genomas virales dentro de una célula infectada del huésped y de la frecuencia de las células doblemente-infectadas. A pesar

de la importancia del conocimiento de dichos factores, actualmente solo se ha estimado experimentalmente la tasa de recombinación del retrovirus *virus del mosaico de la coliflor* (*Cauliflower mosaic virus*, CaMV) (Froissart *et al*., 2005), mediante una cuantificación *in vivo* y directa de la tasa de recombinación de dicho virus durante la infección de un huésped. Dicha tasa se estimó en 4×10^{-5} eventos de recombinación por nucleótido y por ciclo de replicación.

En los potyvirus, observaciones *in vivo* han demostrado que aislados del mismo virus se segregan espacialmente durante infecciones mixtas. Esta segregación debería reducir al mínimo la posibilidad de dos genotipos de infectar las mismas células y así limitar eventos de recombinación durante la replicación del ARN viral. Para conciliar estas observaciones, hemos evaluado la tasa de recombinación y la multiplicidad de infección (*MOI*) del virus TEV, *in vivo*. La tasa de recombinación se estimó en 1.03×10^{-5} eventos de recombinación por nucleótido y generación. Este valor se encuentra en el mismo orden de magnitud que la tasa de mutación del TEV, lo que sugiere que la recombinación tiene una importancia

comparable a la mutación puntual en la creación de variabilidad. La multiplicidad de infección celular se define como el número de genomas virales que logran infectar de manera eficaz una célula. Dos estudios recientes han mostrado estimaciones *in vivo* de la *MOI* para el *virus del mosaico del tabaco* (*Tobacco mosaic virus*, TMV) y del CaMV, gracias a métodos sofisticados que miden la distribución de dos genotipos virales en las células del huésped.

Aquí presentamos un análisis detallado de la dinámica temporal y espacial de la *MOI* celular durante la colonización de una planta por el TEV. Observamos una baja frecuencia tanto del número de células infectadas por el virus (media ± SD: 0.100 ± 0.073), como de las infectadas por dos genotipos del TEV (media ± SD: 0.012 ± 0.023). Se usó un nuevo método basado en un modelo de selección para determinar la *MOI*. Los valores de *MOI* predichos fueron bajos, oscilando entre 1.0 (hoja tres, 3 días después de la inoculación (dpi)) a 1.6 (hoja cuatro, 7 dpi).

Por último, la alta diversidad genética de las poblaciones virales da lugar a una nube de variantes, todas

vinculadas a través de mutación, que interactúan y contribuyen colectivamente, conocido también por el término de cuasiespecies. Las poblaciones virales pueden adaptarse rápidamente a sus entornos dinámicos, y son notablemente inestables en determinadas condiciones como hemos demostrado en este trabajo durante la evolución del TEV en el caso de redundancia genética y funcional. Después de varios pasos a tiempos largos de infección en plantas transgénicas que expresan la replicasa viral NIb, se observaron grandes deleciones y múltiples partículas defectuosas. Este resultado demuestra la gran plasticidad del genoma de los virus ARN y su capacidad para eliminar cualquier carga genética innecesaria.

ABSTRACT

RNA viruses are probably some of the most pervasive parasites that can be found in all life forms. These infectious agents seem particularly prone to causing emerging diseases in plants, humans and other animals. Their abilities to escape the immune system, evade antiviral strategies or to jump to new species are a direct consequence of their enormous capacity to evolve quickly. Understanding basic processes of evolutionary change may be a necessary primary step in designing new intervention strategies. One of the key characteristics of RNA viruses, especially from an evolutionary perspective, is their capacity for mutation. Knowing mutation rates and the molecular spectrum of spontaneous mutations is important to understanding how the genetic composition of viral populations evolves. Previous studies have shown that the rate of spontaneous mutations for RNA viruses widely varies between 0.01 and 2 mutations per genome and generation, with plant RNA viruses always occupying the lower side of this range. Here we analyse the spontaneous mutational spectrum

and the mutation rate of *Tobacco etch potyvirus* (TEV), a model system of positive sense RNA viruses. Our experimental set up minimizes the action of purifying selection on the mutational spectrum, thus giving a picture of what types of mutations are produced by the viral replicase. We have estimated that the spontaneous mutation rate for this virus was in the range 10^{-6} - 10^{-5} mutations per site and generation. Our estimates are in the same biological ballpark as previous values reported for plant RNA viruses.

Recombination is a virus characteristic that also contributes significantly to the diversity and evolution of viruses. The recombination rate depends on two parameters: the frequency of genetic exchange between viral genomes within an infected host cell and the frequency of co-infected cells. Despite this importance, only one direct quantification of the *in vivo* recombination rate for an RNA virus during host infection has been reported: the *in vivo* recombination rate for *Cauliflower mosaic virus* (CaMV) (Froissart *et al.*, 2005) was reported to be 4×10^{-5} events per nucleotide site and per replication cycle. In plant potyviruses, *in vivo* observations have

shown that strains of the same virus segregate spatially during mixed infections. This segregation shall minimize the chances of two genotypes to co-infect the same cells and henceforth, precludes template switching and recombination events during genomic RNA replication. To reconcile these confronting observations, we have evaluated the *in vivo* TEV recombination rate and multiplicity of infection (*MOI*). TEV recombination rate was estimated to 1.03×10^{-5} recombination events per nucleotide site and generation. This value is in the same order of magnitude than TEV mutation rate, suggesting that recombination should be at least as important as point mutation in creating variability.

The multiplicity of cellular infection is defined as the number of viral genome infecting effectively a cell. Two recent studies have reported *in vivo MOI* estimates for *Tobacco mosaic virus* (TMV) and CaMV, using sophisticated approaches to measure the distribution of two virus genotypes over host cells. Here, we present a detailed analysis of spatial and temporal dynamics of the cellular *MOI* during colonization of a host plant by TEV. We observe a low frequency of virus-infected

cells (mean ± SD: 0.100 ± 0.073), and cells infected by both virus variants (mean ± SD: 0.012 ± 0.023). A new, model-selection-based method was used to determine the MOI, and the predicted MOIs values were low, ranging from 1.0 (leaf three, 3 days post inoculation (dpi)) to 1.6 (leaf four, 7 dpi).

Finally, high genetic diversity in viral population gives rise to a cloud of variants; linked through mutation, interacting and contributing collectively; also called quasi-species. Viral populations can rapidly adapt to dynamic environments, remaining remarkably unstable under certain conditions as observed during evolution of TEV in case of genetic and functional redundancy. Large deletions and multiple defective particles were observed after various passages of long time TEV infection in transgenics plants expressing the viral replicase NIb. This result demonstrates the great genome plasticity of RNA virus and their capacity to eliminate any useless genetic load.

PROLOGUE

During the last decades, we have witnessed the emergence of many new diseases with potentially devastating effects for people's health, livelistocks, agriculture, and environments. A large majority of these new diseases are caused by viruses, and in particular the RNA viruses, such as HIV-1/AIDS, SARS-coronavirus or new and rare reassortants of *Influenza A virus* (IAV). Despite advances in medicine and research, the amount of human suffering caused is unacceptable. For example, there are two million new HIV-1 cases per year and almost 30 million accumulated deaths since HIV-1 first discovered (UNAIDS, 2010). Moreover, the economic cost associated with viral infections is also enormous. Each year, the total cost of IAV infections in the USA alone is estimated to be $50 billion (World Health Organization, 2003). Emerging plant viruses are also common and known to induce enormous economic losses in crops (Navas-Castillo *et al*., 2011; Roossinck, 2008) estimated to $60 billion (Cann, 2005). Most of the studies on emerging viruses undertaken to date have

focused on the ecological processes that determine pathogen transmission dynamics (Antia *et al*., 2003; Woolhouse, 2002). Nevertheless, it is still unclear what evolutionary characteristics of virus determine their capacity to invade new host species (Parrish *et al*., 2008), and to escape from vaccines and antiviral strategies (Farci *et al*., 2000). As E.C Holmes has said "*To understand how RNA viruses are able to jump species boundaries and spread in new hosts it is essential to determine the basic processes of evolutionary change in these infectious agents*". RNA viruses evolution is largely determined by high rates of mutation and recombination, and taking into account their extremely small genomes and their enormous population size, evolution of these infectious agents looks imprevisible but not inexplicable.

Indeed, we can take advantage of their small and rapidly evolving genome to make viruses the ideal system to work on evolutionary dynamics. Others characteristics as the amount of sequenced genomes data in GenBank, their enormous population size, the knowledge of life cycle virus, viral protein structure and function make them a powerful tool

to study evolutionary process and define fundamental life-history traits.

INTRODUCTION

1. Mutation rate

1a. Sources of genetic variations

Mutation is the most important evolutionary source of genetic changes on which natural selection and random drift operate. A mutation event is defined by a genome modification such as the substitution, deletion or addition of a nucleotide, and it is generally caused by replication errors. Since the first study on bacteriophages in 1940s, replicase fidelity has been well studied. RNA-dependent RNA polymerases (RdRps) are known to be the most error-prone, followed by retrotranscriptases, RNA-dependent DNA polymerases and finally DNA polymerases (Flint *et al.*, 2004). Viral replicases differ biochemically between viruses, which suggests a plausible source of variation in mutation frequency.

The lack of a proofreading function in RdRps, for example, is the most important difference between RNA and DNA polymerases. Misincorporations during the amplification of RNA virus genomes by their RdRps cannot be corrected during replication, and result in an increase of the mutation

rate by one or two orders of magnitude, as compared to DNA replicases (García-Díaz and Bebeneck, 2007).

A second source of variation is viral sequence composition. Chemical base modifications such as oxidation, methylation or deamination can also affect the mutation rate (Caride *et al.*, 2002; Drake, 1993; Mangeat *et al.,* 2003). Base-repeated of AU rich regions can cause polymerase slippage during replication, causing a decrease of replicase fidelity and processivity (Nagy *et al.*, 1999). Finally, it was demonstrated that RNA secondary structures such as hairpins have a strong impact on variation in mutation rates, causing replicase pauses, which could lead to deletions and imprecise homologous recombination (Nagy and Bujarski, 1996; Pathak and Temin, 1992).

A third source of variation is the mode of viral replication of (+) and (-) strands, which can be exemplified by two extreme models: the linear model called "stamping machine" and the geometrical model (Loverdo *et al.*, 2012; Sardanyés *et al.*, 2009; Thébaud *et al.*, 2010). These models (Figure 1-1) impact differently the pattern of intra-cell mutation

accumulation in the riboviral genome. For the geometric model, RNA strands of each polarity are copied into their respective complementary strands with the same efficiency. In this case, mutations should accumulate geometrically: a mutated template would propagate the given error to all of its replicate copies, resulting in a relatively high mutation frequency (Luria and Delbruck, 1943). On the other hand, for the stamping machine model the RNA initially infecting the cell is the template for the asymmetric synthesis of genomic RNAs. Thus, given the same intrinsic rate of polymerase error, stamping-machine replication would result in lower mutation frequency than geometric genome replication (Drake, 1993; Drake and Holland 1999). Two studies on riboviral replication strategy suggest that these viruses replicate their genome mostly using the "stamping machine" model (Chao *et al.*, 2002; Martínez *et al.*, 2011).

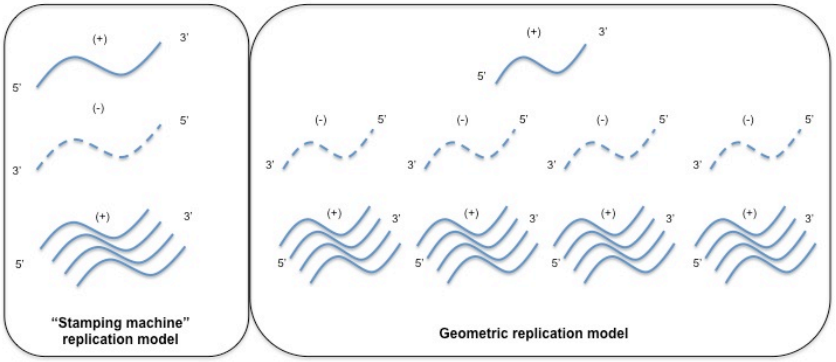


Figure 1-1. Scenarios describing the mode of the viral RNA amplification process.

1b. Mutation rate estimations

The rate of spontaneous mutation is a key parameter to understand the genetic structure of populations over time. Generally, RNA viruses have an extremely high mutation rate, higher than what is generally observed for DNA viruses (Figure 1-2-A) (Drake *et al*., 1998). Nevertheless, a broad range of mutation rates has been observed (Figure 1-2-B), from 1.1×10^{-3} mutations per nucleotide and per cell infection cycle (m/n/c) in the positive single-stranded (ss) RNA phage Qβ (Batschelet *et al*., 1976) to 1.7×10^{-6} m/n/c in the negative ssRNA *Influenza B virus* (Nobusawa and Sato, 2006). To achieve polymerase activity retroviruses use the RT, which could be less error prone

than the RdRp of RNA viruses (Drake *et al*., 1998). It is therefore somewhat surprising that retroviruses appear to have the same mutation rate as RNA viruses (Sanjuán *et al*., 2010). Furthermore, a significant correlation was observed between virus genome size and mutation rate of (+)ssRNA, (-)ssRNA and dsRNA viruses: the smaller the viral genome the higher the mutation rate (Sanjuán *et al*., 2010). It was suggested that RNA viruses evolve to genome sizes just below the upper limit supported by their mutation rate (Eigen, 1971; Swetina and Schuster, 1982), which implies that polymerase complexity, genome size and mutation rate are inextricably linked. For example, RNA viruses with larger genome sizes will accumulate more deleterious mutations than those with smaller genomes, *ceteris paribus*. The only way to increase size genome and genetic complexity would therefore be to increase replication fidelity. To replicate with greater fidelity, however, requires a polymerase with proofreading activity. This leads to a conundrum called Eigen's paradox: to obtain proofreading activity, an RNA virus must already have a higher genome size. (Smith and Szathmáry, 1995).

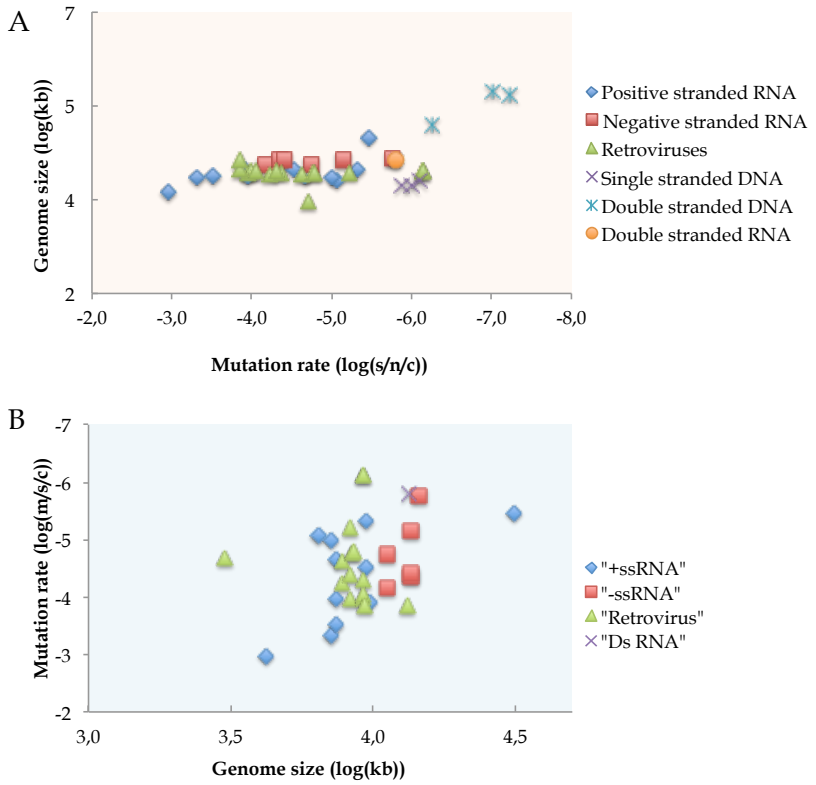


Figure 1-2. (A) Relationship between RNA/DNA virus genome size and mutation rates. **(B) Relationship between RNA virus genome size and mutation rates**. Data were taken from http://www.uv.es/rsanjuan/virmut.htm.

Finally, a number of misconceptions and biases concerning mutation rate estimation have been reported. First, a clear distinction must be made between the mutation frequency and rate (Drake, 1993; Sanjuán *et al.*, 2010). Mutation

frequency is the actual number of mutations observed during the experiment, whereas mutation rate is the estimated rate at which mutations occured during some time interval. This time interval can be one viral replication event, one round of replication, an infectious cycle, etc. The necessity to include time in the estimation of mutation rates leads to a further complication: the different units of time used (per replication, per generation or per infectious cycle). Sanjuan *et al*. (2010), reviewed various studies to facilitate comparative analyses using two time intervals: per strand copying event or per infected cell. These common time intervals finally allow for meaningful comparisons of the mutation rates of different RNA viruses. To correct the different mutation rates, the authors also proposed different strategies in function of the sampling method, taking into account two important biases: (i) errors induced by PCR amplification and (ii) purifying selection against or positive selection for mutants (Sanjuán *et al*., 2010). For example, bias induced by selection could be eliminated using a neutral target (Malpica *et al*., 2002; Tromas and Elena, 2010).

1c. High mutation rate selection

It has been argued that high mutation rates were selected to generate diversity in virus populations, and hereby allow them to confront the host immune system, rapid environment change (Giraud *et al*., 2001) or vaccines and antiviral drugs (Gerrish and García-Lerma, 2003). Although, low fidelity polymerases generate more variation, the majority comprises deleterious mutations that lead to a population generally less fit for a stable set of conditions. However, these variations may endow a population with the capacity to dominate if a sudden change in environment occurs. Conversely, a homogeneous viral population generated by conservative replication might be at a disadvantage in a variable host environment. Recent work with poliovirus (PV) by two groups provides experimental support for this idea. Thanks to a specific mutation, the authors observed a significative increase of the poliovirus polymerase fidelity. Both this high-fidelity virus as well as the WT poliovirus were able to replicate with a lower fitness, a lower virulence and with a restriction of tissue tropism (Pfeiffer and Kirkegaard, 2005;

Vignuzzi *et al*., 2006). Nevertheless, there is still no experimental proof for the proposed selective advantage of high genomic variation: no evidence has been found to support a positive correlation between mutation and adaptation rates, as one would expect according to this hypothesis (Clune *et al*., 2008).

An alternative explanation to the adaptive value of the high mutation rates is the existence of a trade-off between replication accuracy and efficiency, such that high mutation rates are only a consequence of selection for rapid replication. Two studies support this hypothesis, identifying the fitness costs of higher fidelity, such as a lower replication rate (Furió *et al*., 2005; 2007). Furthermore, recent works have demonstrated that more robust viral populations can be superior competitors, despite having lower replication rates (Codoñer *et al*., 2006; Sanjuán *et al*, 2007). This phenomenon (Figure 1-3) is called the "survival of the flattest" and occurs when robust genotypes produce equally fit phenotypes (*i.e*., mutations are neutral) while non-robust genotypes suffer from mutational fitness effects. In this case, RNA virus evolve as quasi-especies (Eigen,

1996); *i.e.*, a cloud of closely related genomes through mutations. One prediction of quasispecies theory is that at high mutation rates a slow-replicating quasi-species can outcompete a faster-replicating one if the first is more robust to deleterious mutational effects (Van Nimwegen *et al*., 1999; Wilke, 2001).

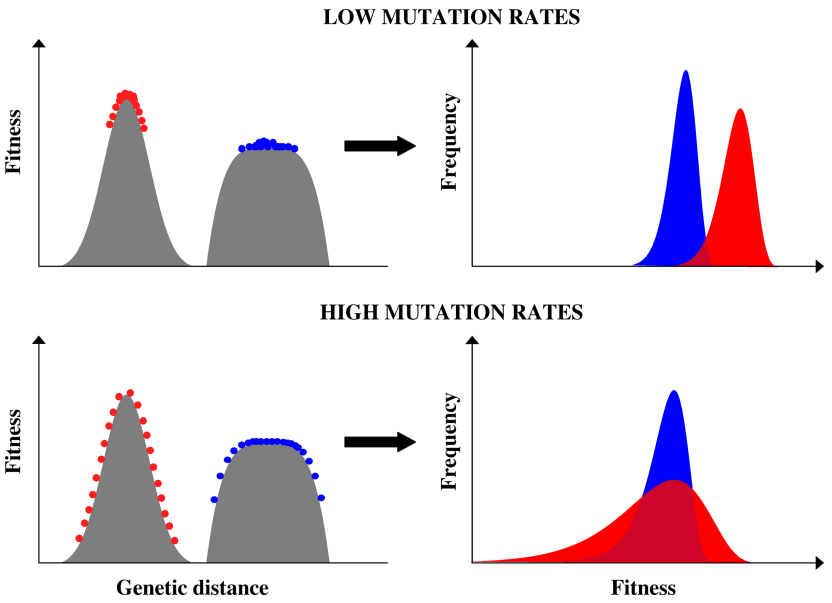


Figure 1-3. Hypothetical populations located at regions of sequence space that differ in their neutrality. In red, a population with high replication rate but low mutational robustness. In blue, a "flat" population with lower replication rate but a higher robustness. Left panel: Dots depict individuals located on each peak at low and high mutation rates. Right panel: Qualitatively expected distribution of individual fitness

values for the two populations. Individuals located at the top of the high narrow peak benefit from the highest fitness, although the other population shows lower variance in fitness due to its higher mutational robustness. At low mutation rates, the fitter population will always outcompete the flatter population, but the situation can be reversed at high mutation rates (from Sanjuán *et al*., 2007).

It has indeed been demonstrated that selection could induce mutation rate variation under different fitness landscape profiles (smooth or rugged), and should be reduced on rugged landscapes but not on smooth ones to avoid harmful mutation production. Then it seems that under certain conditions (*i.e*., the geometry of fitness landscape), mutation rates could vary, moving away from optimal rates so as to obtain short-term benefits while limiting adaptation on the long run (Clune *et al*., 2008).

1d. High mutation rate consequences

The majority of mutations are deleterious as demonstrated in previous studies on *Vesicular stomatitis virus* (VSV), where up to 40% of random mutations were lethal and

29.2% were deleterious (Sanjuán *et al.*, 2004b), and on TEV, where 40.9% were lethal and 36.4% were deleterious (Carrasco *et al.*, 2007). The principal consequence of high mutation rate is the accumulation of deleterious mutations in viral population and the associated fitness loss in the absence of purifying selection (Chao, 1990; De la Iglesia and Elena, 2007; Duarte *et al.*, 1992; Escarmís *et al.*, 1996; Yuste *et al.*, 1999). This phenomenon has also been confirmed using mutagens such as ribavirin or 5-fluorouracil. Different authors have shown that artificially increased mutation rates induce higher mutational loads and major decreases of fitness, probably to the point that the viral population eventually goes extinct (Crotty *et al.*, 2000; Loeb *et al.*, 1999; Sierra *et al.*, 2000). How can RNA viruses therefore maintain their high mutation rates?

Mutational robustness, *i.e.* phenotypic constancy in the face of mutational changes to the genome (Montville *et al.*, 2005), could be a characteristic approach for RNA viruses to deal with the accumulation of deleterious mutations. As viral RNA genomes do not contain any redundancy mechanisms, such as gene duplications or overlapping gene functions, the

impact of each mutation will be strong. In large populations, if the majority of mutations are deleterious, as expected for an anti-redundancy strategy, a new kind of robustness might be arising through population-level strategies that facilitate the preservation of the non-mutated genotype (Krakauer and Plotkin, 2002).

Previous studies confirmed the existence of genetic robustness in RNA viruses (Bonhoeffer *et al.*, 2004; Burch and Chao, 2004; Lalić and Elena, 2012; Sanjuán *et al.*, 2004b) and suggested that antagonistic epistasis (the combined effect of mutations being weaker than expected under a multiplicative model) is a characteristic trait of RNA viruses to minimize deleterious mutational effects. Antagonistic or positive epistasis can be explained by the multiple-hit-effect in a compact genome with widespread RNA secondary structures, essential multifunction proteins and overlapping genes without redundancy systems, thus meaning that a single mutation in one viral gene could cause a major negative impact on fitness. After this initial loss, further mutations that occur in this viral gene may have a comparatively lesser impact. Interestingly,

antagonistic epistasis and robustness appear to be inversely correlated such that antagonistic epistasis is correlated with weak robustness (Elena *et al*., 2006; Sanjuán and Elena, 2006). Finally, complementation of defective or less-fit genotypes by viruses with higher fitness during co-infection can be viewed as a form of robustness (Froissart *et al.*, 2004). Thus, defective or less-fit genotypes are maintained in the viral population, leading to selection acting with a lower efficiency to purge harmful mutations, and ultimately to a higher mutational load in the viral population. Indeed, complementation may be advantageous in the short term, because the cost of deleterious mutations could be partially or fully masked. Conversely, in the long term, complementation could slow the rate at which deleterious alleles are eliminated from the virus population. To counterbalance this, two different mechanisms were suggested: (i) genetic bottlenecks during cell-to-cell movement that allow a rapid selection to efficiently separate fit from defective genomes (Miyashita and Kishino, 2010) and (ii) recombination events that occur in co-infected cells, which could help viral

population to recover the non-mutated genotype (Carpenter and Simons, 1996; Lai, 1992; Otto and Lenormand, 2002).

2. Recombination

2a. Definition and mechanisms

Until its first observation in PV by Hirst (1962) and Ledinko (1963), the role of recombination in RNA viruses was underestimated and it was thought to be restricted to DNA molecules. In 1982, recombination was finally demonstrated experimentally for the first time by biochemical and genetic analyses of *Foot-and-mouth disease virus* (King *et al.*, 1982). RNA-RNA recombination is now well accepted as a common phenomenon in RNA viruses and has been observed in an increasing number of animal (Bergmann *et al.*, 1992; Lai *et al.*, 1985; Zhuang *et al.*, 2006), plant (Beck and Dawson, 1990; Bujarski and Kaesberg, 1986; Gal-On *et al.*, 1998) and bacterial riboviruses (Munishkin *et al.*, 1988; Onodera *et al.*, 1993). Mechanistically, RNA-RNA recombination in viruses can be

defined as an exchange of genetic material between at least two different viral genomes subjected to a replicase-driven-template-switching action. To recombine, at least three kinds of RNAs are necessary: the donor RNA strand, the nascent strand synthesized from the donor RNA template, and the acceptor strand. Generally, template switching occurs when the RdRp stops nascent RNA synthesis due to the RNA donor sequence or RNA secondary structures (Nagy and Bujarski, 1995; 1996). Then the RdRp must be able to bind the RNA acceptor and reinitiate RNA synthesis using the 3' end of the nascent RNA as a primer. Three classes of recombination (Figure 2-1) have been proposed to define precisely the mechanism of template switching (Nagy and Simon, 1997): the first class is base-pairing dependent, requires short (15 – 60 nucleotides) similar sequence (Nagy and Bujarski, 1995) and a perfect alignment between the donor and acceptor RNA. In the absence of a strict alignment of the parental RNAs, recombination crossovers may occur and lead to sequence insertions or deletions. The second recombination class does not require similarity between sequences, but other features such as RdRp binding sequences,

RNA secondary structure, heteroduplex formation between parental RNAs or *cis*-acting replication elements. Indeed, a previous study showed that for this kind of mechanism, RNA specific structures such as a heteroduplex between parental RNAs could drive the template switching (Nagy and Bujarski, 1993; Nagy *et al.*, 1995). The last recombination class combines the characteristics of the two first classes. Sequence similarity between parental RNAs influences sites and frequency of recombination but RNA secondary structure (*e.g.*, a hairpin) in the acceptor RNA region is also necessary to recruit the RdRp (Nagy *et al.*, 1998).

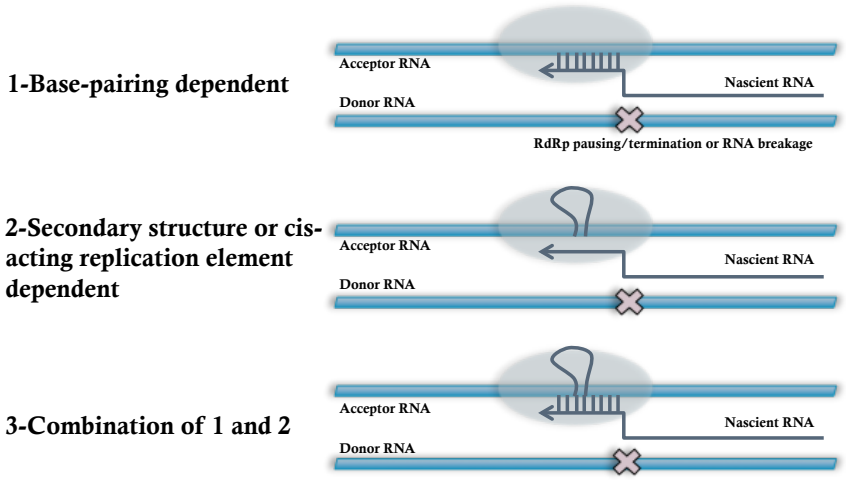


Figure 2-1. Recombination mechanisms (Inspired by Nagy and Simon, 1997).

2b. Recombination rates estimation

Due to the difficulties with detecting recombination events in some RNA viruses and the lack of standardized estimation methods, different strategies have been taken to estimate RNA recombination rates. These strategies include phylogenetic methods (Chare and Holmes, 2005; Codoñer and Elena, 2008; Martín *et al*., 2009; Ohshima *et al*., 2007; Revers *et al*., 1996), cell culture (Kirkegaard and Baltimore, 1986; Levy *et al*., 2004; Reiter *et al*., 2011), or *in vivo* (Bruyere *et al*., 2000; Froissart *et al*., 2005; Urbanowicz *et al*., 2005). A direct consequence of the use of various approaches to measure recombination rates is that highly variable frequencies of recombination have been observed in RNA viruses, although these variable frequencies probably also reflect the variability within RNA viruses to some extent. For example, recombination appears to occur frequently in some retroviruses such as the *Human immunodeficiency virus* type 1 (HIV-1), which has an estimated recombination rate of between 3×10^{-4} and 1×10^{-5} events per site per generation (Batorsky *et al*., 2011; Jetzt *et al*., 2000), *Spleen necrosis virus* (SNV), with 4×10^{-5} events per

site per generation (Anderson *et al*., 1998) and *Moloney's Murine Leukemia virus* (MMuLV) with an estimation rate estimated between 4.6×10^{-4} and 4.7×10^{-5} events per site per generation (Anderson *et al*., 1998; Zhuang *et al*., 2006).

For plant RNA viruses, evidences of recombination were reported in different viral families such as the *Bromoviridae* (Bruyere *et al*., 2000; Urbanowicz *et al*., 2005) and *Potyviridae* (Gibbs and Ohshima, 2010). In 2005, the first estimate of the *in vivo* recombination rate for a plant pararetrovirus was reported, for CaMV (Froissart *et al*., 2005). It was found that recombination was frequent and estimated its rate to be 4×10^{-5} events per nucleotide site and per replication cycle.

However, recombination appears to be far less frequent in other families of (+)ssRNA viruses, including the *Flaviviridae*, which is the case of *Hepatitis C virus* (HCV), with only 4×10^{-8} events per site per generation (Reiter *et al*., 2011). Nevertheless, as is the case with HCV or other members of viral families like *Flaviviridae* (Carney *et al.* 2012); technical limitations appear to be the reason for such low recombination rate estimates

(González-Candelas *et al*., 2011). Indeed, yet today, recombination has not been detected in a number of (+)ssRNA viruses, including members of the families *Barnaviridae*, *Leviviridae*, and *Narnaviridae*. This lack of detection may reflect the limited amount of data available. Surprisingly, current data indicate that recombination is even less frequent in (−)ssRNA viruses (Chare *et al*., 2003). For example, in *Rhabdoviridae* family, viral RNA is condensed by the nucleoprotein into a helical nucleocapsid also called ribonucleoprotein complex (RNP). This complex serves as a template for replication should not allow template switching and apparently never disassembles (Simon-Loriere and Holmes, 2011). Another example revealed only sporadic evidences for homologous recombination, after a large analysis of *Influenza A virus* genomic sequences from virus isolates (Boni *et al*., 2010).

2c. Evolutionary advantages and constraints of RNA recombination

It is still not clear why high RNA virus recombination rates have been selected, but an important number of studies

have described roles for recombination that may explain its evolutionary success. On the one hand, recombination may be a means to create novel genotypes, mitigating the effects of clonal interference and accelerating the rate at which multiple beneficial mutations can accumulate within the same genome. In this sense, recombination may play an important role in the development of drug resistance (Nora *et al.*, 2007; Pamilo and Li, 1987; Yusa *et al.*, 1997) and would act as an accelerator of variability. On the other hand, one could imagine that recombination would also accelerate the combination of deleterious mutation and hereby help to increase the mutational load. Indeed, it was found that sex can be advantageous for counteracting the accumulation of deleterious mutations when those mutations are negatively epistatic or more harmful together than would be expected from their separate effects (Eshel and Feldman, 1970). Nevertheless, previous studies suggest that even if epistatic interactions are a common phenomenon, they tend to be positive (Bonhoeffer *et al.*, 2004; Lalić and Elena, 2012; Sanjuán *et al.*, 2004a; Shapiro *et al.*, 2006).

Furthermore, recombination has been shown to increase the efficiency with which deleterious mutations are purged (Chao *et al*., 1992; Simon-Loriere and Holmes, 2011). Indeed, high mutation rates combined with small population sizes during viral transmission result in Muller's ratchet dominating the evolutionary dynamics (Chao, 1990; De la Iglesia and Elena, 2007; Muller, 1964). In its simplest form, this theory states that accumulation of deleterious mutations in small asexual populations leads to a progressive decrease in fitness, and that genetic random drift accelerates the loss of mutation-free individuals. Then, recombination would help to halt Muller's ratchet in viral population.

In addition to the evidence favouring a role for genetic exchange in fixing beneficial alleles or eliminating deleterious ones, recombination could provide viruses with genetic new variation by borrowing host genetic material. One surprising example of this is *Bovine viral diarrhoea virus* (BVDV), a pestivirus that recombines with host cellular protein-coding RNAs. Cytopathogenic pestivirus strains originated from noncytopathogenic viruses by nonhomologous RNA

recombination and caused host death (Meyers *et al*., 1989a). Therefore, by allowing the generation of new combinations of genes, recombination permits the population to explore regions of the genotypic space that were not represented in the previous generations and thus, under selection pressure, generate new viral genome variants with a different genomic organization (Crow, 1992).

However, constraints also exist, as there are costs associated with sex in RNA viruses. These costs include an increasing degree of intra-host competition under those conditions that foster high recombination rates (Turner and Chao, 1998). Indeed, because recombination between standing genetic variation can occur only during co-infection, the different genotypes must be in direct competition within the host. Then, social behaviours trait as cheating could be selected for under certain conditions (Turner and Chao, 2003).

Finally it has been suggested that recombination was selected as a repair mechanism driven by genetic damage, instead of a mechanism for sex: various experimental studies showed that weak or even non- replicative mutant strains could

recombine to form viable ones. White and Morris (1994) described the formation of functional genomes thanks to recombination between nonreplicating RNAs and DI RNAs of tombusviruses. Furthermore, examples of infectious recombinants generated by different combinations of mutationally altered *Sindbis virus* RNAs were also described (Raju *et al*., 1995; Weiss and Schlesinger, 1991). Plant viruses have also been observed to repair their genomes by recombining with host transgene transcripts (Borja *et al*., 1999; Gal-On *et al*., 1998; Greene and Allison, 1994; Rubio *et al*., 1999). Similarly, in one experiment with a deletion mutant of *Mouse hepatitis virus* (MHV) transfected with a synthetic RNA that contained the deleted region (Koetzner *et al*., 1992) recombination successfully repaired defective genes. It was eventually proposed that recombination could only be a by-product of the processivity of RNA polymerases (Simon-Loriere and Holmes, 2011). This theory suggests that genome organization and gene expression control are negatively correlated with recombination rate, based on the fact that (−)ssRNA viruses have more options than (+)ssRNA viruses for

the control of gene expression and that recombination rates for (−)ssRNA are very low (Chare *et al*., 2003).

3. Estimation of the multiplicity of infection (*MOI*) during RNA virus dynamic infection in a multicellular host

3a. Definition

The cellular multiplicity of infection (*MOI*), that is, the mean number of virions effectively infecting each cell, is a key parameter for understanding the dynamics and evolution of virus populations. This number is highly relevant for virus evolution because: (i) *MOI* is a determinant of the amount of genetic drift at the cellular level and the distribution of different viral genotypes over cells, (ii) complementation, recombination or reassortment between different genotypes can only occur in mixed-genotype infected cells (Figure 3-1), whilst mixed-genotype infections can only occur if the $MOI > 1$, (iii) *MOI* will be a determinant of the respective importance of different levels of selection in viral evolution, and (iv) for many viruses,

defective interfering particles can be generated and maintained for substantial periods of time if *MOI* is high. Competition between virus genotypes occurs at the between-host, within-host, within-tissue and within-cell levels, and the relative importance of different levels of selection is modulated by *MOI*. For example, low *MOI* levels (*MOI* ≈ 1) relax selection at the within-cell level and increase selection at higher levels (Froissart *et al*., 2004; Miyashita and Kishino, 2010; Taylor *et al*., 2002; Turner and Chao, 1999, 2003; Zwart *et al*., 2008). Cellular *MOI* is therefore not only relevant to mechanisms at the cellular level, but is of great relevance to understanding viral evolution.

3b. MOI estimation in a multicellular host

At first glance the concept of *MOI* is straightforward: it is the mean number of virions successfully infecting a population of cells. There are, however, two possible ways to define *MOI*: (i) the average number of infecting virions over the total number of cells, or (ii) the average number of infecting virions in infected cells. Both definitions are valid and are likely to be used in different contexts. The first definition is

particularly useful for describing manipulable units in an experiment (*e.g.*, virion dose and the number of cells for infections of cultured cells). The second definition, however, gives a more readily interpretable value for understanding the population genetics and evolution of a virus population. In this case we are only interested in infected cells because no viral replication or interactions between genotypes occur in uninfected cells. However, in both cases there are problems when applying these concepts to a complex multi-cellular host. In particular, two important assumptions are being made: (i) the population of cells is homogenous, with each cell being equally susceptible to viral infection, and (ii) there is free mixing of virions and cells, such that each cell is equally accessible to virions. Moreover, if two virus variants are used to estimate *MOI*, then there must also be free mixing of the two virus variants. Although these assumptions may be largely met for a monolayer of cultured cells, they will probably not be met for a multi-cellular organism, with its complex spatial organization of differentiated cells with varying susceptibilities (Morra and Petty, 2000; Silva *et al.*, 2002). A key question is

therefore to what extent the assumptions of current *MOI* models are met, and whether this has important implications for making meaningful *in vivo MOI* estimates.

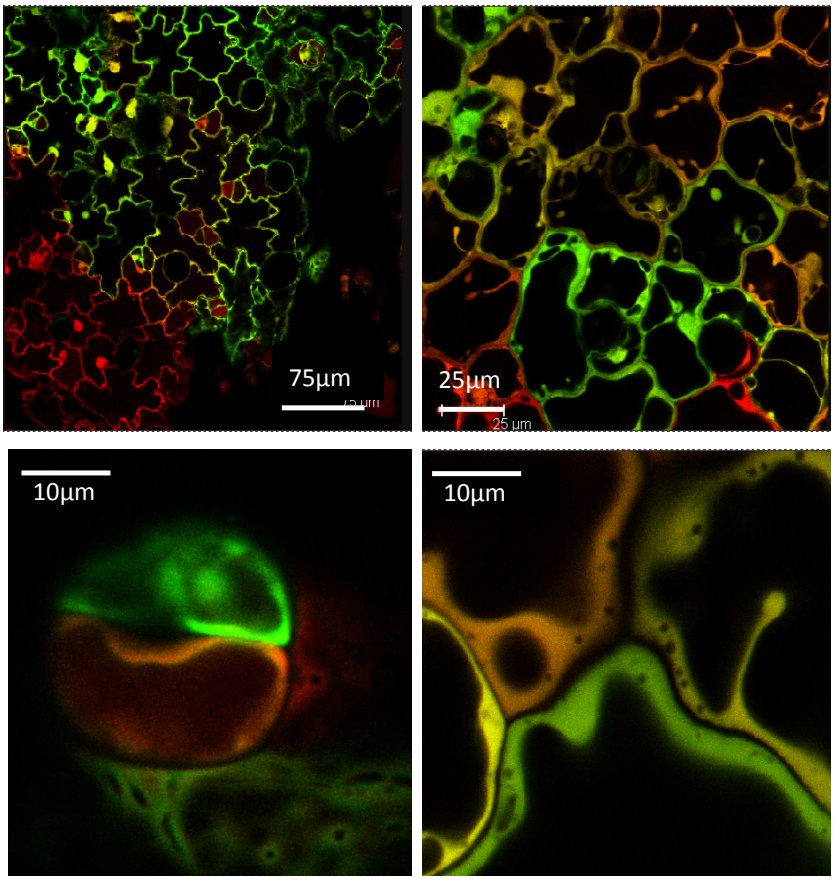


Figure 3-1. Distribution of two engineered TEV genotypes (TEV-GFP and TEV-mCherry). Confocal imaging of mixed infections in systemically infected *N. tabacum* cv. Xanthi tissues.

3c. MOI previous estimations

Despite of its importance, only few estimates of *MOI* in the natural hosts of viruses have been undertaken: values of two to three were reported for an RNA bacteriophage (Olkkonen and Bamford, 1989; Turner *et al*., 1999), and values of four to five were reported for the final round of *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) replication in its insect host (Bull *et al.,* 2001). Recently, three studies reported, for the first time, estimates of the *in vivo MOI* during host colonization for plant viruses (González-Jara *et al*., 2009; Gutiérrez *et al.,* 2010; Miyashita and Kishino, 2010). For *Tobacco mosaic virus* (TMV), *MOI* was measured by infecting a host with virus variants marked with different markers, determining which cells were infected with one or both variants, and then using a simple mathematical model to estimate *MOI* (González-Jara *et al*., 2009). This approach was used to show that the *MOI* during TMV infection of *Nicotiana benthamiana* ranges from four during initial infection to two during early systemic infection. Similar experiments were performed for CaMV infection of *Brassica rapa*, although here *MOI* rose from two to 13 as the

virus expanded, and dropped to two again as the infection progressed even further. Finally, a similar experimental approach was also used to estimate *MOI* during primary infection of *Soil-borne wheat mosaic virus* (SbWMV), although a model considering the separation of genotypes at the cellular level was used to estimate *MOI*. This analysis rendered an estimated *MOI* of six and five for the first and second rounds of cellular replication, respectively (Miyashita and Kishino, 2010).

4. Organization of RNA viruses

4a. Evolution of genome size

Most RNA virus genomes have two evolutionary characteristics: a high variability and an extremely small size. Nevertheless, a large range of genome sizes has been found for RNA viruses, with a large majority having a genome size of between 5000 and 15000 nucleotides. Two notable exceptions are members of *Narnaviridae* (< 4000 nucleotides) that have the smallest RNA genomes, and members of *Coronaviridae* (> 30000 nucleotides) that have the largest ones. So what could explain

this range of genome sizes in RNA viruses? In one hand, the features of small genomes lead to apparently lower enzymatic complexity, lower fidelity replication enzymes and so on higher variability (Smith and Szathmáry, 1995). Thus, a clear advantage of a smaller genome would be to replicate quickly; taking advantage of a high degree of optimization. In the other hand, a larger genome size allows genetic complexities such as greater replication fidelity, as described in *Coronaviruses* (Denison *et al*., 2011), and/or allows them to modulate host immune response (Shackelton and Holmes, 2004). Therefore, it appears that each genome size has its own advantages, and interestingly various RNA viruses with different genome sizes could infect and evolve into the same host (Holmes, 2009).

4b. Process leading to an increase of genome size

If RNA viruses are really limited in their genome size due to the biochemical properties of their RdRp, then we should not observe increases in the genome size of RNA viruses. Indeed, few examples of the integration of host genetic material into RNA genomes have been observed, and most of

the observed cases concern BVDV, where inserts of host ubiquitin genes were found (Meyers *et al*., 1989b). Those insertions had a direct impact on the cytopathogenicity, leading to positive selection for these BVDV variants. In a similar way, only few examples of viral gene duplications have been observed in RNA viruses, such as short duplications in the UTRs flanking conding sequences (Gritsun and Gould, 2006) or gene duplication involving intra-genic regions that result mainly in defective viruses (Nagai *et al*., 2003). Why those different examples are an exception? When a host gene is inserted into a viral genome, the size of the RNA genome increases, that could implicates an increase of replication time, a loss of replication accuracy and so on a reduced viral fitness (Holmes, 2009). For example, Dolja *et al*. (1993) described large deletions in a foreign gene, GUS, inserted into TEV genome.

4c. Non-retroviral integrated RNA viruses

On the other hand, there have been recent reports of the integration of large amounts of non-retroviral (without DNA stage) RNA virus genes into eukaryotic genomes. These

integred sequences have been named as non-retroviral integrated RNA viruses (NIRV) and have been found in mammals (Belyi *et al*., 2010; Geuking *et al*., 2009; Horie *et al*., 2010; Katzourakis and Gifford, 2010) plants (Chiba *et al*., 2011) and fungi (Taylor and Bruenn, 2009) (Table 4-1). In this last category, NIRVs are widespread and become functional genes (Koonin, 2010). Interestingly, it seems that those insertions could confer a resistance against the original virus (Bertsch *et al*., 2009; Flegel, 2009). Therefore, a particularly interesting situation appears when the same virus that gave rise to the NIRV infects a host carrying it. How would a virus population evolve if faced with a situation in which functional redundancy, provided by the NIRV, has arisen?

Host	Virus	Gene(s)	Transcription
Fungi			
Candida parapsilosis	Totivirus (dsRNA) related to *Saccharomyces cerevisiae* L-A (L1) virus	Capsid protein (CP)	Not tested
Debaryomyces hansenii	Totivirus (dsRNA) related to *S. cerevisiae* L-A (L1) virus; M2 killer virus (unclassified dsRNA virus)	CP, RdRp; apparently complete viral genome with overlapping CP and RdRp genes as in L1 virus; an extra copy of the CP gene; 4 copies of K2 toxin gene	RdRp but not Cp transcripts detected by reverse transcriptase polymerase chain reaction (RT-PCR)
Penicillium marneffei	Totivirus (dsRNA) related to *S. cerevisiae* L-A (L1) virus	CP, RdRp; apparently complete viral genome with overlapping CP and RdRp genes as in L1 virus	Not tested
Pichia stipitus	Totivirus (dsRNA) related to *S. cerevisiae* L-A (L1) virus	CP, RdRp; apparently complete viral genome with fused CP and RdRp genes; 3 extra copies of the CP genes	RdRp but not CP transcripts detected by RT-PCR; transcripts of both genes detected in EST database
Uromyces appendiculatus	Totivirus (dsRNA) related to *S. cerevisiae* L-A (L1) virus	CP (from EST database)	CP gene transcript detected in EST database

Candida tropicalis	M2 killer virus (unclassified dsRNA virus)	K2 toxin gene	Not tested
Kluyveromyces lactis	M2 killer virus (unclassified dsRNA virus)	K2 toxin gene (3 copies)	Not tested
Vanderwaltozyma polyspora	*Penicillium stoloniferum* virus F	CP	Not tested
Animals			
Mouse (*Mus musculus*)	Lymphocytic choriomeningitis virus (negative-strand ssRNA, Arenaviridae)	Glycoprotein and nucleoprotein genes (small genomic segment)	Not tested
Mosquitoes (*Aedes albopictus* and *Aedes aegypti*)	Flavivirus (positive- strand ssRNA) related to Cell Fusing Agent and Kamiti River virus	Non-structural protein genes (3'-terminal part of the genome)	Expressed

Table 4-1. Non-retroviral integrated RNA viruses in eukaryotic genomes (from Koonin, 2010). CP, capsid protein; Ds, double-stranded; ss, single-stranded; RdRp, RNA-dependent RNA polymerase.

5. Viral model: *Tobacco etch virus*

5a. History and generalities

A new group of viruses with filamentous flexuous virions was discovered by Brandes and Wetter (1959), and named *Potyvirus* (from *Potato virus Y*) by Harrison *et al*. (1971). Nowadays, *Potyvirus* is the largest of six genera within the family *Potyviridae* (Adams *et al*., 2005b; Shukla *et al*., 1994) and responsible for the largest losses of crop plants worldwide (Agrios, 1997). *Potyviridae* genomes belong to the "Picornavirus supergroup" of (+)ssRNA viruses (group IV in Baltimore's classification).

Virions of this family are generally described as non-enveloped, flexuous filaments of 11 to 15 nm in diameter, with monopartite genomes of length 650-900 nm, or bipartite genomes of 500 - 600 and 200 - 300 nm (Shukla *et al*., 1998). One of the characteristics of the *Potyviridae* is to induce the formation of three-dimensional crystalline cytoplasmic inclusions (CI) within infected cells. Some members of the *Potyvirus* genus can also induce the formation of crystalline

nuclear inclusions (NI) that consist of two proteins, NIa and NIb (Shukla *et al*., 1998).

Figure 5-1. An aphid (*Myzus persicae*) feeding on a leaf (from *Cornell plant pathology herbarium photograh collection*).

All members of *Potyvirus* are transmitted in a non-persistent, or stylet-borne, manner by aphid insects (Figure 5-1) feeding on leaves and stems. Aphids acquire the virus by feeding on infected plants and then, simply by means of passively carrying virions and in the absence of replication, transmitting it when feeding on uninfected plants thanks to the ingestion and salivation process (Gray and Banerjee, 1999).

TEV is a well-characterized member of the genus *Potyvirus*. It is mainly distributed in North and South of America, and transmitted by aphid insects or by mechanical

inoculation. Its principal hosts are *Solanaceae* species where TEV induces mottling, necrotic etching and leaf distortion (Johnson, 1930; http://www.dpvweb.net/dpv/showdpv.php?dpvno=258).

TEV has a monopartite RNA genome of approximately 10 kilobases (kb) covalently linked to a virus encoded protein (VPg) at the 5′ end and tailed with a poly(A) tail at its 3′ end (Riechmann *et al.*, 1992). The TEV genome contains two open reading frames (ORF); ORF1 is translated into a large 346 kilodalton (KDa) polyprotein, which is autoprocessed by the proteases into ten mature viral proteins (Figure 5-2). ORF2 is embeeded within the P3 cistron of the polyprotein. ORF2 is translated in the +2 reading-frame, resulting in a 7 - 17 KDa protein (Chung *et al.*, 2008) named P3N-PIPO, a fusion of the N-terminus of P3 and, following a frameshift during translation, PIPO. Because the small genomes of plant viruses generally encode fewer than a dozen proteins, the function of each viral protein must be well defined and optimized for host interaction. P1 is a trypsin-like serine peptidase corresponding to the 5′-terminal gene of the TEV genome, which catalyzes its

own cleavage from the polyprotein in its C terminus (Verchot *et al.*, 1991, 1992). Helper component proteinase (HC-Pro) is a papain-like cysteine multidomain peptidase, also containing a C-terminal autocleavage domain (Carrington *et al.*, 1989). HC-Pro is also implicated in aphid transmission, systemic movement and suppression of gene silencing. P3 could be implied in plant pathogenecity (Riechmann *et al.*, 1995) and recently, data suggested that P3N-PIPO complex acts as movement protein and coordinates the utilization of plasmodesmata, essential for systemic infection (Wei *et al.*, 2010). Cytoplasmic inclusion (CI) is a multidomain protein that is involved in cell-to-cell movement and has a helicase motif, suggesting it has a role in TEV replication (Carrington *et al.*, 1998). 6K2 is a small protein responsible for anchoring the replication complex to the endoplasmic reticulum via a hydrophobic domain (Restrepo-Hartwig and Carrington, 1994). The N-terminal domain of nuclear inclusion protein a (NIa-VPg) and the nuclear inclusion protein b (NIb), which is the RdRp, are the viral replication machinery (Schaad *et al.*, 1997a, 1997b). The C-terminus of NIa is also a trypsin-like serine

proteinase (NIa-Pro) involved in the maturation of replication-associated and capsid proteins (Dougherty and Semler, 1993). Capsid protein (CP) functions are essential in aphid transmission, cell-to-cell and systemic movement (López-Moya and Pirone, 1998) and virus assembly (Jagadish *et al.*, 1993).

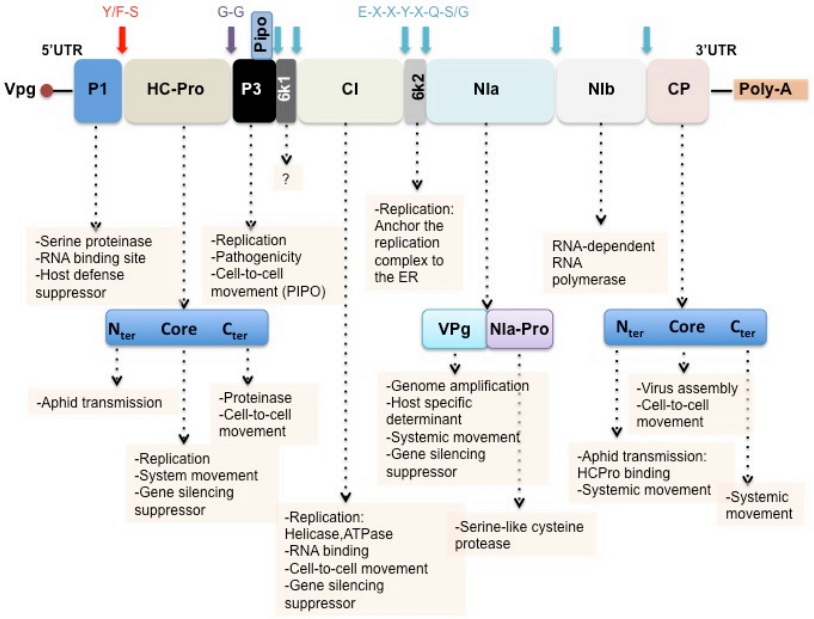


Figure 5-2. TEV genome organization. P1, P1 protein; HC-Pro, helper component-protease protein; P3, third protein; VPg, viral protein genome–linked; NIa-Pro, major protease of small nuclear inclusion protein –NIa; CP, coat protein; UTR, untranslated region. The functions of the genes are also

indicated. The cleavage site of P1 serine protease is indicated in red, those of HC-Pro in dark grey and those of NIa-Pro in blue.

5b. TEV replication

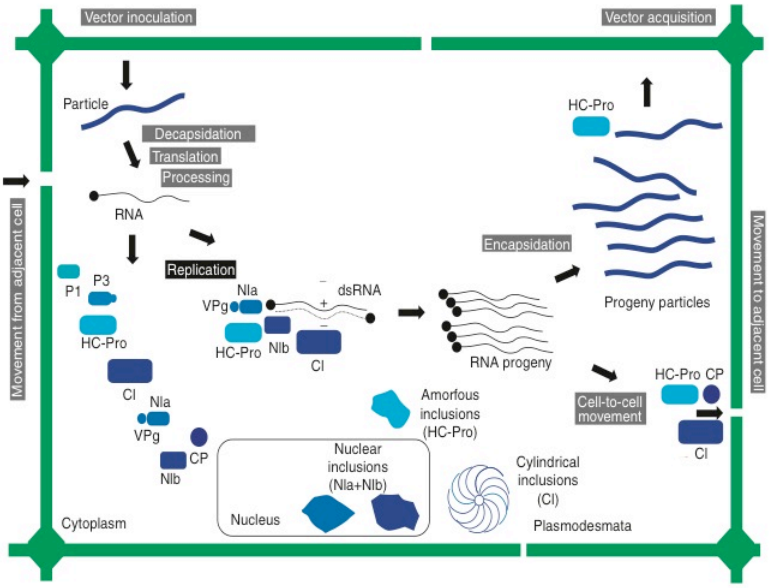


Figure 5-3. Schematic representation of different events during *Potyvirus* infection of a plant cell (from Encyclopedia of Virology, 3rd edition).

The mechanisms of TEV replication are still unknown. It is nevertheless possible to infer with a reasonable degree of certainty the basis for the replication process. TEV replication (Figure 5-3) takes place in the cytoplasm of infected cells, and is

specifically localized on heaps of tubules in the endoplasmic reticulum (Schaad *et al*., 1997a), probably to avoid non-viral RNA replication. It is useful to differentiate between two different types of cellular infection: (i) the infection of the first cell by aphid action, and (ii) the infection of neighbouring or distant cells by local or long-range movement. In the case of local movement, both viral RNA and virions are transferred into the next cell via plasmodesmata (Santa Cruz *et al*., 1998). Then, replication occurs directly on the viral RNA, with an uncoating process if necessary, by the action of a viral RdRp. The NIb was characterized as the TEV RdRp (Hong and Hunt, 1996) and could form a replication complex with other viral protein as CI (helicase motif), NIa-VPg (replication initiator) and 6K2 (endoplasmic reticulum anchoring factor). Indeed strong protein-protein interactions were observed between NIb and NIa (Li *et al*., 1997) and NIb and VPg (Fellers *et al*., 1998). Probably host proteins join the viral replication complex and finally the genomic (+)ssRNA will be used as template to generate a complementary (-) chain through a dsRNA

intermediate, and proceed with the asymmetric synthesis of numerous genomic RNAs (Martínez *et al*., 2011).

5c. TEV cell-to-cell and systemic movement

Viral infection begins with the infection of a relatively small number of cells in the inoculated or exposed leaf (Zwart *et al*., 2011). However, many hurdles must be surpassed before systemic infection of the plant is established. Indeed several steps are required, starting with expansion from a single infected epidermal cell to adjacent tissue such as mesophyll, bundle sheath or phloem parenchyma/companion cells (Carrington *et al*., 1996) by means of relatively slow cell-to-cell movement (Gillespie *et al*., 2002). The cell-to-cell movement of virions or infectious RNA requires modification of the plasmodesma, which are brought about by three viral proteins: HC-Pro, CP and a complex CI/P3N-PIPO that coordinates the formation of plasmodesmata-associated structures (Dolja *et al*. 1994; Maia and Bernardi, 1996; Rojas *et al*., 1997; Wei *et al*., 2010). However, plasmodesma structure differs according to the plant tissue and it has been shown that *Potyvirus* have a low

propensity to infect companion cells (Ding *et al*., 1998). Infection of companion cells to sieve elements (Ding *et al*., 1996) offers access to long distance transport within the plant, allowing for rapid expansion to distant tissues (Leisner *et al*., 1993). The infection of companion cells appears to be a key process in plant colonization; where HC-Pro could play a major role, facilitating this passage by both loading and unloading virus from vascular cells (Cronin *et al*., 1995). Only by establishing systemic infection can a plant virus readily have access to the tissue in which vectors will feed, and hereby ensure its transmission to new hosts. Virions loaded into phloem are responsible for long-distance movement in TEV.

During mixed-genotype infection; plants infected by PVX/potyvirus combinations showed synergistic symptoms and a large numbers of coinfected cells (Dietrich and Maiss, 2003). In contrast, spatial segregation of closely related genotypes was observed during virus colonization of the host (Dietrich and Maiss, 2003). Indeed, in general clusters of each genotype are formed in leaf space; limiting superinfection by blocking the access for a new infection. This potyvirus

characteristic could be a direct consequence of the modification of plasmodesma during cell-to-cell movement. However, yet today the exclusion process, also called superinfection exclusion, is poorly understood and the underlying mechanism unknown. Superinfection exclusion protects an already infected cell from a second infection by a related virus. Besides elimination of competition for host resources, superinfection exclusion could help to maintain the stability of viral sequences by, for example, limiting opportunites for recombination (Folimonova, 2012).

MATERIALS AND METHODS

1. The rate and spectrum of spontaneous mutations in a plant RNA virus

1a. Virus and plants

The pTEV7DA (GenBank DQ986288) infectious clone (Dolja *et al.*, 1992) was used as source for TEV. A TEV genotype was produced that lacked the full replicase gene (Δ*NIb*) by inverse PCR using *Pfu* turbo DNA polymerase (Stratagene) and primers conserving the proteolytic NIa-NIb (Δ*NIb*-F) and NIb-CP (Δ*NIb*-R) sites (Table 1-1). The resulting clone was named pTEV7DA-Δ*NIb*.

Two different genotypes of *N. tabacum* were used in these experiments, the wild-type cv. Xanthi and the transgenic *Nt*::NIb line derived from cv. Samsun by Li and Carrington (1995). These transgenic plants express TEV NIb protein in a stable and functional manner. Prior to starting our experiments, the presence of the transgene was confirmed by PCR using *Taq* polymerase (Roche) and the primers F90-95 (identical to bases 7767 - 7786 of TEV NIb) and R86-91 (complementary to bases 8084 - 8102 of TEV NIb) (Table 1-1). The expression of the gene

also was confirmed by RT-PCR. MMuLV reverse transcriptase (Fermentas) was used to obtain cDNA from plants RNA extracts using primer R92-96 (complementary to bases 8761 - 8779 of CP gene). Then this cDNA was amplified using *Taq* and primers F90-95 and R86-91 (Table 1-1). Finally, the biological activity of the NIb protein encoded by the transgene was confirmed by inoculating batches of *Nt*::NIb plants with infectious RNAs from both viruses. All *Nt*::NIb plants inoculated with either TEV (n = 20) or TEV-Δ*NIb* (n = 10) developed a systemic infection after 6 - 7 dpi. By contrast, none of the wild-type plants inoculated with TEV-Δ*NIb* (n = 5) became infected, while all plants inoculated with TEV (n = 5) were so. Furthermore, these results confirm that no putative RNA secondary folding structure within the NIb coding sequence is necessary for completing the infectious cycle of the virus.

1b. Experimental procedure

Infectious plasmid pTEV7DA was linearized with *Bgl*II (Takara) and transcribed into 5'-capped RNAs using the SP6

mMESSAGE mMACHINE kit (Ambion Inc). Transcripts were precipitated (1.5 volumes of DEPC-treated water, 1.5 volumes of 7.5 M LiCl, 50 mM EDTA), collected, and resuspended in DEPC-treated water (Carrasco *et al*., 2007). RNA integrity was assessed by gel electrophoresis and its concentration spectrophotometrically determined using a Biophotometer (Eppendorf). Twenty four-weeks old *Nt*::NIb plants were inoculated mechanically on the third true leaf with TEV transcripts (4 - 7 μg) and 10% of inoculation buffer (100 mg/mL carborundum, 0.5 M K_2HPO_4, 3% polyethylene glycol 8000, pH = 7). In all cases, first symptoms appeared 6 - 7 dpi.

Total RNA was extracted using RNeasy Plant Mini Kit (Quiagen) from symptomatic leaves of 3 *Nt*::NIb plants at 5, 10, 15, 20, 25, and 60 dpi. One of the plants at 20 dpi was not sampled because it dried out. The full NIb gene was reverse-transcribed MMuLV and primer R92-96 and PCR-amplified using the high fidelity PrimeSTAR HS DNA polymerase (Takara Bio Inc) and primers F73-80 (identical to bases 6377 - 6396 of TEV NIa gene) and R92-96 (Table 1-1). By using this pair of primers we ensure that only NIb sequences from viral

genomes and not the mRNA from the transgene will be amplified. PCR products of 2403 pb were gel purified with Zymoclean (Zymo Research), cloned into the plasmid pUC19/*Sma*I (Fermentas) and used to transform *Escherichia coli* DH5α electrocompetent cells. At least 25 clones per plant were purified and sent out for sequencing by GenoScreen (www.genoscreen.fr) using BIGDYE 3.1 and a 96-capillars ABI3730XL sequencing system (Applied Biosystems). The following five internal primers were used for fully sequencing NIb with overlapping readouts: F1, F2, F3, F4, F5 (Table 1-1). Contigs were assembled using GENEIOUS version 4.7 (www.geneious.com). The number of clones that rendered useful sequences was 472 (instead of the 500 submitted for sequencing). The number of sequenced clones per plant ranged between 12 and 34, with a median value of 24.

1c. Mutation rate estimations

Two different approaches have been used to estimate TEV mutation rate. In the first approach, we proceeded as follows. For a given plant the number of clones sequenced that

contained zero, one, two, …, k mutations was fitted to a Poisson distribution with parameter $\lambda = \mu lT$, where λ is the expected number of mutations per clone, T the number of generations of viral replication, l = 1536 the length of the amplicon, and μ the mutation rate per base and per generation (m/b/g). Defining generations *in vivo* for a plant virus is troublesome, given that a viral population colonizing a plant is not replicating synchronously but with overlapping generations. A good approximation is to define viral generations as the number of cycles of cell infections (Malpica *et al*., 2002). For this definition to be operative, it is necessary first to have an estimate of the average number of viruses produced per infected cell. By performing one-step accumulation curves in tobacco protoplasts, Martínez *et al*. (2011) have estimated that, on average, an infected cell yields 1555 genomes (quantified by real-time quantitative RT-PCR). To estimate the number of generations experienced by TEV at the time points where the samples were taken, we revisited previously published data on the kinetics of TEV accumulation (Carrasco *et al*., 2007). Reanalyzing these data, we found that the model that better

describes TEV accumulation within an infected plant was a 4-parameters Gompertz growth equation (R^2 = 0.975) (Campbell and Madden, 1990). From the parameters of the model and using the above estimate of virus yield per cell, it is possible to calculate that during the exponential growth phase, the viral population experienced 3.2 generations per day, but this number reduces as growth rate flats off and the carrying capacity of the system is reached. After estimating the number of generations corresponding to each sampling day, it is then possible to transform the above per clone mutation rate values into the biologically meaningful scale of mutations per base and per generation using the simple expression $\mu = \lambda/lT$. Each plant has been treated as an independent replicate, rendering 19 estimates of μ.

For the second approach, we focused only on putatively lethal mutations, that is, mutations generating frameshifts or stop codons. Readers need to recall that ORF1 encoded by TEV genome is translated into a single polyprotein. Our method is based on the fact that amino acid substitutions affecting NIb would in turn be neutral because the *trans* complementation

provided by the host (and the best evidence of such active *trans* complementation is the ability of TEV-Δ*NIb* to infect *Nt*::NIb plants). However, frameshift mutations and stop codons affecting the NIb sequence would be lethal because they will produce a virus deficient not only in NIb but also in CP, the gene downstream from NIb, which is not complemented by the host. In haploid populations at the mutation-selection balance, the frequency of deleterious mutations, p, is given by $p = \mu/s$, where s is the selection coefficient. For lethal mutations, however, $s = 1$, then $\mu_L = p$ and the equilibrium is reached instantaneously because all lethal mutations have been generated in the previous generation (Crow and Kimura, 1970). In other words, this method provides an estimate of mutation rate per replication event (m/b/r) rather than by generation, as in the first method. Following Cuevas *et al*. (2009), it is possible to calculate a mutation rate for the *i*th amplicon using the expression

$$\mu_{NSTMT,j} = \frac{1}{n}\sum_{i=1}^{K} W_i , \qquad (1)$$

where n is the total number of nonsense mutational targets

(NSMT: sites that can generate a stop codon after a single nucleotide substitution) in an amplicon, W_i a weighting factor for the two types of nonsense mutations (W_i = 3 if only one of the three possible mutations in a NSMT produces a stop codon and W_i = 1.5 if two out of three possible produce a stop codon), and K is the total number of observed nonsense mutations in the amplicon. According to the standard genetic code, there are 18 NSMT-containing codons and 19 different NSMTs (the UGG codon contains two). In our experiments, we have 472 independent estimates of μ_{NSMT}. If the frequency of insertions and deletions is μ_{indel} (it can be computed using the Poisson distribution, as described above), then μ_L can be estimated as

$$\mu_L = \mu_{indel} + \frac{1}{472}\sum_{j=1}^{472}\mu_{NSMT,j}\,. \qquad (2)$$

where $\mu_{NSMT,j}$ is estimated using Equation 1.

Hereafter, we will use the notation μ_L when referring to the estimated based on the frequency of lethals (units of m/b/r) and reserve the notation μ for the Poisson estimate (units of m/b/g).

1d. Statistical analyses

All statistical tests were performed using IBM SPSS version 16. All molecular evolutionary analyses were done using MEGA4 (Tamura *et al.,* 2007).

Primers	TEV genome position (5′)	Sequence (5′ to 3′)
ΔNIb-F	6940	TTGCGAGTACACCAATTCACTCATGA GTTGAGTCGCTTCCTT
ΔNIb-R	8518	AGTGGCACTGTGGGTGCTGGTGTTGA CGCTGGTAAGAAGAAA
73-80-F	6377	TCATTACAAACAAGCACTTG
90-95-F	7768	GCTGTATTGAAAGTGCGAC
86-91-R	8084	AGGCCCAACTCTCCGAAAG
92-96-R	8762	GCAAACTGCTCATGTGTGG
F1	6911	GCAAACCTGAAGAGCCTTTTCAG
F2	7265	GCATGCTCATCACAAAGCTCAAG
F3	7591	GTGGATGATTTCAACAATCAATTTTA TGAT
F4	8531	ACCAGCGTCAACACCAGCAC
F5	8215	GATCTGTCCCATTCCAAAATAGAAAC

Table 1-1. List of primers used to amplify or sequence TEV genome.

2. Estimation of the *in vivo* recombination rate for a plant RNA virus

2*a. Generation of restriction sites as genetic markers*

We used the pTEV7DA infectious clone (Dolja *et al*., 1992) as a source for TEV. To analyze TEV recombination rate, we introduced five genetic neutral markers, in the form of artificial restriction sites, along TEV genome. New *Asc*I, *Pme*I and *Sgf*I restriction sites were created at positions 402, 3735 and 9463, respectively, whereas natural restrictions sites *Eco*47III and *Sal*I were removed from positions 4969 and 7166, respectively (Figure 2-2A). All mutations necessary to create or remove restriction sites were introduced by PCR-directed mutagenesis using the Quickchange® II XL kit (Stratagene) and following the indications given by the manufacturer. In all cases, mutations rendered synonymous substitutions. Table 2-1 shows the pairs of primers used for mutagenesis, that were also designed following Stratagene's recommendations. To minimize unwanted errors during the mutagenesis process, the kit incorporates the *Pfu*Ultra™ high fidelity DNA polymerase

(Stratagene). The amplification conditions were 1 min at 95ºC (initial denaturation), followed by 18 cycles consisting of 30 s at 95 ºC, 45 s at 65 ºC and 18 min at 68 ºC, and a final extension step of 28 min at 68ºC. PCR products were digested with *Dpn*I (New England Biolabs) to remove the parental methylated strands and transformed into electrocompetent *E. coli* DH5α. At least 15 clones were sequenced to confirm the successful incorporation of desired mutations.

2b. Neutrality of restriction markers

Sequence-validated plasmids containing the corresponding restriction site marker were linearized with *Bgl*II (Takara) and transcribed into 5'-capped RNAs using the SP6 mMESSAGE mMACHINE kit (Ambion Inc). Transcripts were precipitated, collected and resuspended as described previously (Carrasco *et al*., 2007).

To evaluate the neutrality of the five markers, we proceeded as follows. Three 24-week-old *N. tabacum* cv. Xhanti plants were inoculated by abrasion on the third true leaf with 7 μg of transcribed RNA from each individual marker as

described elsewhere (Carrasco *et al.*, 2007). Inoculated plants were placed in a BSL-2 greenhouse at 25 ^{o}C and 16 h light/8 h dark period. No differences in the time to produce symptoms were observed among marked viruses and, in all cases, symptoms appeared after 6 – 7 dpi.

2c. Virus quantification and markers neutrality test

The accumulation of TEV genomes was evaluated by quantitative RT-PCR 7 dpi for each engineered genotype as described elsewhere (Lalić *et al.*, 2011). RT-qPCRs were performed using One Step SYBR PrimeScript RT-PCR kit II (Takara) following the instructions provided by the manufacturer. The primers forward TEV-CP-qPCR-F 5′-TTGGTCTTGATGGCAACGTG-3′ and reverse TEV-CP-qPCR-R 5′- TGTGCCGTTCAGTGTCTTCCT-3′ amplify a 50 nucleotides 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).

No differences were observed between markers (Model II nested ANOVA, $F_{4,14} = 0.436$, $P = 0.781$) thus confirming their neutrality.

2d. Coinoculation experiments and restriction analysis

Twenty-four-week-old *N. tabacum* plants were mechanically inoculated on the third true leaf with 7 µg of an equimolecular mixture of RNA transcripts. The four combinations assayed were *Asc*I/*Pme*I, *Pme*I/*Eco*47III, *Eco*47III/*Sal*I, and *Sal*I/*Sgf*I (Figure 2-2). Each combination was inoculated on 10 plants. Total RNA was extracted using InviTrap® Spin Plant RNA Mini Kit (Invitek) from the symptomatic leaves for each of five plants 15 dpi and from the other five plants 45 dpi. Each plant was analyzed separately,

thus providing independent replicates of the recombination rate among pairs of markers. For each combination of restriction markers, the regions of interest was reverse transcribed using the following reaction mixture: 1× RT reaction buffer (Fermentas), 0.2 mM each dNTP, 0.25 μM forward primer (Table 2-2), 0.2 μL RNase inhibitor, 40 units of MMuLV (Fermentas), 100 ng total RNA, and DEPC-treated water to complete 20 μL reaction volume. PCR was then done using the following reaction mixture: 1× HF buffer (Finnzymes), 0.2 mM each dNTP, 0.25 μM each primer (Table 2-2), 0.5 units of ultra high-fidelity Phusion DNA polymerase (Finzymes), 1 μL DMSO, 3 μL from the reverse-transcription reaction, and DEPC-treated water to complete a reaction volume of 25 μL. The cycling conditions were: 1 min at 98 °C; followed by 30 cycles consisting of 8 s at 98 °C, 25 s at 57 °C and 25 s/kb at 72 °C; and then a final extension of 5 min at 72 °C. PCR products were gel purified with GeneJET™ Gel Extraction Kit (Fermentas), cloned into the plasmid pUC19/*Sma*I (Fermentas), and used to electro-transform *E. coli* DH5α. A

number of clones per infected plant were purified for subsequent restriction analyses (range 15 – 148, median 43).

Finally, after linearizing the recombinant plasmids with *Kas*I (Takara), recombinant clones were genotyped by digestion with the corresponding pair of restriction enzymes. The expected restriction profiles are shown in Figure 2-2B. For each pair of markers, four progeny genotypes are expected, the two parentals (Figure 2-2B, left column) and the two recombinants (Figure 2-2B, right column).

2e. Statistical analyses

The frequency of recombinant genotypes on each analyzed plant was estimated using the Laplace's point estimator for the Binomial frequency parameter (Chew, 1971)

$$f = \frac{Recombinants+1}{Recombinants+Parentals+2}.$$

This method provides more robust estimates than the commonly used maximum likelihood estimator when sample sizes are small. Recombination rates, *r*, between restriction markers were then computed according to Kosambi (1944) formula

$$r = \frac{1}{4L} \ln \frac{1+2f}{1-2f},$$

where *L* is the physical distance separating the two markers (in nucleotides). The units of the resulting estimate are recombination events per site (r/s). The Kosambi method was chosen to minimize the potential effect that multiple crossovers may have in the inference of *r*. The independent estimates for each pairs of markers were averaged and the corresponding SEM computed.

A genome-wide *r* was computed averaging all estimates obtained from different combinations of restriction markers and experimental replicates. Finally, since TEV produces an average of ca. three generations per day in *N. tabacum* (Tromas and Elena, 2010), recombination rate can be expressed as recombination events per site and generation (r/s/g), r_g.

All other statistical analyses reported have been performed using IBM SPSS version 19.

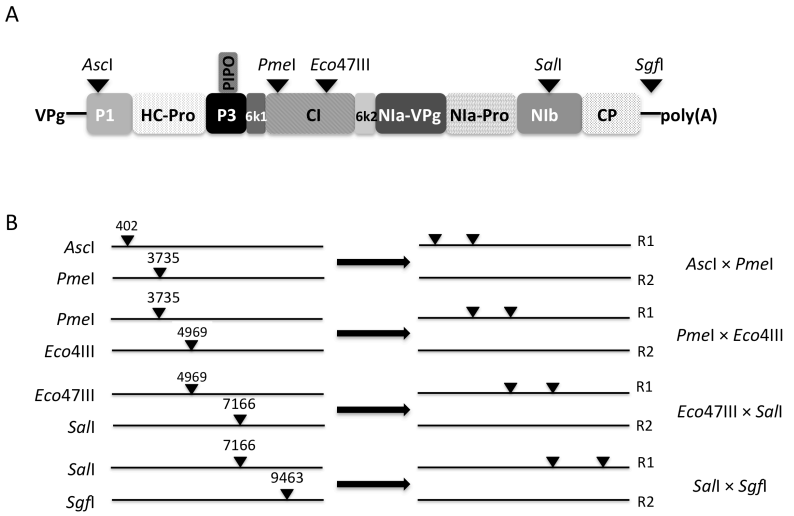


Figure 2-2. (A) Location of the different restriction site markers in TEV genome. (B) Expected restriction profile for each pair of markers. The left column shows the two parental genotypes and the right column the two recombinant ones

Restriction enzyme	Cistron	Genomic position for the cut	Mutagenesis primer (5′ to 3′)
*Asc*I	P1	402	TTATCTTGGTCG***G***CGCGC ***C***CTCACCCATGGC
*Pme*I	CI	3735	AGCCTTCCTGGAGTCAC ***G***TTTAA***A***CAATGGTGGA ACAACCA
*Eco*47III	CI	4969	AGTCATACATGACAAGC TGAA***A***CG***T***TTAAGCTAC ACACTTGTGAG
*Sal*I	NIb	7166	GATGGGAGCATATAAGC CAA***CC***CGACTTAATAGA GAGGCG
*Sgf*I	3′UTR	9463	CTTAGGCGAACGACAAA GTG***C***G***A***TC***G***CCTCGGTCT AATTCTCCTAT

Table 2-1. Artificial restriction sites engineered as genetic markers for this study. Restriction sites are underlined**.** Mutagenized sites are indicated with bold cursive type.

Combination	TEV genome position (5′)	Sequence (5′ to 3′)
*Asc*I/*Pme*I	46	GCAATCAAGCATTCTACTTC
	3894	ATCCAACAGCACCTCTCAC
*Pme*I/*Eco*47III	3541	TTGACGCTGAGCGGAGTGATGG
	5275	CTATTGATGCATGCTAGAGTC
*Eco*47III/*Sal*I	4972	TTAAGCTACACACTTGTGAGAC
	7394	TTCTTTCTTCTTGCCTTTG
*Sal*I/*Sgf*I	7058	CACAAAGCATGTGGTTAAAG
	9492	CGCACTACATAGGAGAATTAG

Table 2-2. Primers used to amplify the region containing the pair of restriction sites.

3. Estimation of the multiplicity of infection (*MOI*) during RNA virus dynamic infection in a multicellular host

3a. Construction of TEV-Venus and TEV-TagBFP

Infectious plasmid pMTEV (Bedoya and Daròs, 2010) containing TEV genome (GenBank accession DQ986288) was used to construct the TEV-Venus and TEV-TagBFP genotypes in which these two fluorescent marker genes were inserted between TEV P1 and HC-Pro cistrons. Venus or TagBFP cDNA (Bedoya *et al.*, 2012) was amplified by PCR using primers forward 5′-ATGGTGAGCAAGGGCGAGGAG-3′ and reverse 5′-TTGGAAGTACAAGTTTTCTCCGCCCTTGTACAGCTCGTCCATGC-3′ (Venus), or forward 5′-ATGAGCGAGCTGATTAAGGAG-3′ and reverse 5′-TTGGAAGTACAAGTTTTCTCCGCCATTAAGCTTGTGCCCCAGTTTG-3′ (TagBFP). The reverse primers inserted two glycines, as spacers, and a partial NIa-Pro proteolytic site

downstream of Venus or TagBFP sequences. This partial proteolytic site (ENLYFQ) is complemented by the first contiguous serine in HC-Pro cistron to mediate marker release from the viral polyprotein. The pTV1a vector, which contains the first 3221 nucleotides of the TEV genome including the complete P1 to HC-Pro cistrons, was amplified using the forward primer 5′-AGCGACAAATCAATCTCTGAGGC-3′ and reverse primer 5′-TTTGTCGCTATAATGTGTCATTGAG-3′. The PCR-amplified Venus and TagBFP sequences were then ligated into the amplified vector sequence, and transformed into electrocompetent *E. coli* DH5α. The identity of the resulting pTV1a-Venus and pTV1a-TagBFP plasmids was checked by restriction digests and sequencing. Finally, *Pau*I-*Aat*II restriction fragments from pTV1a-Venus and pTV1a-TagBFP were ligated into *Pau*I-*Aat*II digested pMTEV to construct pTEV-Venus and pTEV-TagBFP. All PCR reactions were performed with high fidelity Phusion DNA polymerase (Finnzymes).

3b. In vitro RNA transcription and Inoculation

TEV-Venus and TEV-TagBFP infectious plasmids were linearized with *Bgl*II (Takara) and transcribed into 5'-capped RNAs using the SP6 mMESSAGE mMACHINE kit (Ambion Inc). Transcripts were prepared as described previously. Four-weeks-old *N. tabacum* cv. Xanthi plants were mechanically inoculated on the third true leaf with RNA TEV transcripts mixes (10 µg). Plants were maintained in the green house at 25 °C and 16 h light for one week. Infected tissues were collected 7 dpi. Stocks of infectious virions obtained from freshly TEV-Venus and TEV-TagBFP infected *N. tabacum* cv. Xanthi were used as source of TEV inoculum for our experiments.

3c. Experimental design

Concentrated saps of TEV-Venus and TEV-TagBFP were obtained grinding 500 mg of infected tissue in a mortar with 800 µL of grinding buffer (50 mM potassium phosphate pH 7.0, 3% polyethylene glycol 6000). Viruses were inoculated separately, or by a 1:1 mixture of infectious saps on *N. tabacum*

cv. Xanthi plants four-weeks-old. Inoculation was performed by abrasion of the third true leaf with 15 µL of each infectious sap. To verify the equal proportion of both genotypes in the mix, fluorescence was observed after 3 dpi with a Leica MZ16F stereomicroscope, using a 1× objective lens with GFP2 filter (Leica) for TEV-Venus and Violet filter (Leica) for TEV-TagBFP to count foci of primary infection on the inoculated leaf. Plants showed equal levels of primary infection for both viruses. Protoplasts were then extracted from the inoculated leaf (number 1) and from the systemic infected leaves (numbers 3, 4 and 5 which is the youngest leaf) of five *N. tabacum* cv. Xanthi plants after 3, 5, 7, 10, and 13 dpi.

3d. Protoplasts extraction and fluorescence analysis by flow cytometry

Protoplasts extraction was performed following an adaptation of Sankara Rao and Prakash (1995) method. In short, sliced leaves were incubated overnight with enzymatic solution (0.04% cellulase and 0.015% of pectinase, from Sigma Aldrich;

MS 4.3 g/L, 0.6 M manitol, pH 5.8) in dark at 22 ± 2 °C. The incubated solution containing protoplast was filtered and centrifuged (4 min at 700 rpm). Protoplast were selected by sucrose gradient (21%sucrose, MS 4.3 g/L), washed (10 mM HEPES, 5 mM $CaCl_2$, 150 mM NaCl, 0.5 M manitol, pH 7) and conserved in a hormone solution (MS 4.3 g/L, 0.5 M manitol, pH 5.8, 1 mg/L NAA and 0.1 mg/L BAP). Analysis of the protoplast was done by flow cytometry (Gallios cytometer, Beckman Coulter, CA). This instrument is equipped with a Blue Solid State Diode, 488 nm/22 mW and a Violet Solid State Diode, 405 nm/40 mW, two detectors for light scattering (forward scatter *FS* and side scatter *SS*) and ten fluorescence detectors. In order to collect the whole *FS* signal from the protoplasts we have to use a neutral density filter to decrease 10-fold the *FS* signal. The *FS* signal provides a measure of cell size: large cells show a high *FS* signal. The *SS* signal provides a measure of cell granulometry: dead cells show low, close to null, *SS* signal. The selection for live protoplasts was performed using the chlorophyll fluorescence detected in the 670 nm channel (FL4). The more positive cells (live protoplasts) were

selected by size, granularity and chlorophyll content. Then, a selection of single cell using "time of flight" (FSPeaklin/FSTOFlin) was performed to separate aggregates from single protoplasts. Venus and TagBFP contents were measured, using the 525 nm channel (FL1) and the 450 nm channel (FL9), respectively, on each individual protoplast selected to quantify the number of TEV-Venus infected, TEV-TagBFP infected and mixed infected cells.

3e. MOI model fitting and selection

For model selection, we use the models described in Zwart, Tromas and Elena (unpublished manuscript), and an additional model (Model 4). These models are all based on the Poisson model, but incorporate a number of mechanisms that can account for deviations between the data and model. For all models the following notation is used: m_I is the *MOI* in infected cells (González-Jara *et al.*, 2009), with a range [1, ∞); m_T is the *MOI* in all cells, including uninfected cells (Gutiérrez *et al.*, 2010), and has a range [0, ∞); p_A is the frequency of TEV-Venus

(A), whereas p_B is the frequency of TEV-BFP (B), which can be estimated as:

$$p_B = 1 - p_A =$$

$$[f(\bar{A} \cap B) + f(A \cap B)]/[f(\bar{A} \cap B) + f(A \cap \bar{B}) + 2f(A \cap B)] \quad ,$$

where $f(\cdot)$ represents the experimentally-observed frequencies of cells infected by none, one or both marked virus variants.

MOI model 1:

This model was developed in Gutiérrez *et al*. (2010), and assumes that *MOI* follows a Poisson distribution over cells such that $\Pr(K = k) = m_T^k\, e^{-m_T}/k!$. It follows that frequency of mixed-variant infections is then

$$\Pr(A \cap B) = (1 - e^{-m_T p_a})(1 - e^{-m_T p_b}) \quad (3),$$

where $\Pr(\cdot)$ is the expected probability that cells infected by none, one or both marked virus variants will be observed. There are no model parameters that need to be estimated. *MOI* is assumed to follow a Poisson distribution and therefore an inextricable relationship between the fraction of uninfected cells, $\Pr(A \cup B)$, and the mean of the distribution exists:

$$m_T = -ln\big(\Pr(A \cup B)\big) \quad (4).$$

MOI model 2:

Model 2 allows for spatial segregation of genotypes during virus expansion. The model assumes that as infection progresses some tissues are only invaded by one viral variant. The fraction of the infectious tissue in which both viral variants are present, and hence in which coinfection can occur, decreases constantly over time and hence:

$\Pr(A \cap B) = \left(e^{-t\phi}\right)(1 - e^{-m_T p_A})\left(1 - e^{-m_T(1-p_B)}\right)$,

where ϕ determines the rate of spatial segregation and t is time post inoculation in days. Equation 3 still holds for this model, and only one parameter needs to be estimated: ϕ.

MOI model 3:

This model allows for the possibility that there is not perfect mixing of infected and non-infected cells, and that virus-infected cells tend to be aggregated. The model introduces the infection aggregation parameter Ψ, assuming it to be constant for all leaves, and again incorporates the infection aggregation parameter Ψ:

(3) $m_T = -ln(1 - \Pr(A \cup B)/\psi)$.

Equation 3 still holds for this model. The only parameter to be estimated to fit the model is Ψ. Note that in model 2 the distribution of viral variants can be heterogeneous over space, whereas there is still perfect mixing of viral infection. In model 3 there is perfect mixing of viral variants, but there can be heterogeneities in viral infection over space.

MOI model 4:

This model is an extension of *MOI* Model 3, and assumes that each leaf has its own aggregation constant Ψ. Four parameters must be estimated to fit the model: Ψ_1, Ψ_3, Ψ_4 and Ψ_5.

MOI model 5:

This model relaxes the assumption of independent action by virions by allowing for the effects of super-infection exclusion: Therefore, the *MOI* is:

$m_T = -\omega e^{-t\mu} ln\big(\Pr(\bar{A} \cap \bar{B})\big)$,

where ω determines exclusion effects at at $t = 0$, whilst μ controls the rate of change. Equation 3 still holds for this model.

MOI models 6-9:

We then combined the different models, in order to ascertain whether a combination of mechanisms could explain the data best. Model 6 combines models 2 and 4, model 7 combines models 4 and 5, model 8 combines models 2 and 5, and model 9 combines models 2, 4 and 5. No combinations with model 3 were tested, because model 4 incorporates the same mechanism and is much better supported by the data.

Model fitting, model selection and *MOI* estimates:

In order to perform model fitting and selection, we exploit the fact that for each model there is a relationship between the fractions of uninfected and mixed-variant-infected cells. *I.e.*, Equation 3 is used to predict m_T, which substituted into Equation 4 allows for a prediction of the frequency of mixed-variant infection. The likelihood of a particular probability of coinfection is then:

$$L(\Pr(A \cap B)|A_k, D_k) = \binom{A_k}{D_k} A_k^{\Pr(A \cap B)}(1 - D_k)^{A_k - \Pr(A \cap B)} \quad (6)$$

where D_k is the number of mixed-variant infected cells observed and A_k is the total number of valid observations made by flow cytometry. Model selection was again performed using AIC. For the best fitting model (model 4), we predicted *MOI* (m_T) with Equation 5. To calculate m_I values from m_T, we used the relationship between the means of a zero-truncated and a non-truncated Poisson distribution (Olkin, 1994):

$m_I = m_T e^{m_T}/(e^{m_T} - 1)$.

4. Evolution in case of genetic and functional redundancy in an RNA virus

4a. Virus genotypes

The pTEV7DA infectious clone (Dolja *et al.*, 1992) was used as source for TEV. An infectious plasmid containing TEV genome lacking the full RNA-polymerase gene (Δ*NIb*) was produced as described in paragraph *1. The rate and spectrum of spontaneous mutations in a plant RNA virus*. Both TEV and TEV-

Δ*NIb* were used to inoculate the transgenic *Nt*::NIb line of *N. tabacum* cv. Samsung (Li and Carrington, 1995). These transgenic plants express TEV NIb protein in a stable and functional manner. The presence of the transgene was previously confirmed (Tromas and Elena, 2010).

4b. Experimental evolution protocol

Infectious plasmids pTEV7DA and pTEV7DA-Δ*NIb* were linearized with *Bgl*II (Takara) and transcribed into 5'-capped RNAs using the SP6 mMESSAGE mMACHINE kit (Ambion Inc). Transcripts were prepared as described previously. Two batches of five four-week-old *Nt*::NIb plants were inoculated mechanically on the third true leaf either with TEV transcripts (4 - 7 μg) in inoculation buffer and TEV-Δ*NIb* trancripts (20 - 25 μg) in the same inoculation buffer. In all case, the first symptoms appeared 6 - 7 dpi.

Two different experiment were performed (Figure 4-1):

Experiment 1: Thirty weekly passages of TEV and TEV-Δ*NIb* viruses were performed in *Nt*::NIb plants. Every 7 dpi, virus were extracted from symptomatic leaves of each of the

five lineages per genotypes. 500 mg of symptomatic tissue were ground in 500 µL of grinding buffer. 10 µL of the resulting sap was then mixed with inoculation buffer and used to infect a batch of five healthy transgenic plants by abrasion for each TEV genotype.

Experiment 2: Four 3-weeks-passages of TEV and TEV-Δ*NIb* virus were performed in *Nt*::NIb plants. Every 21 dpi, virus were extracted from symptomatic leaves of each of the five lineages per genotypes. Infectious sap was prepared as described above.

4c. TEV genomic RNA extraction and purification for sequence analysis

Total RNA was extracted using RNAeasy Plant Mini Kit (Quiagen) from symptomatic leaves of five *Nt*::NIb infected by TEV and five *Nt*::NIb infected by TEV-Δ*NIb*.

Experiment 1: The full genome was reverse transcribed using MMuLV RT (Fermentas) and primers 97-101-R, 67-77-R, and 40-45-R and PCR amplified using the high fidelity DNA polymerase Phusion (Finnzymes) and 3 pairs of primers 2-10-

F/40-45-R, 45-48-F/67-77-R, and 73-80-F/97-101-R (Table 4-3). By using these pairs we ensured that the mRNA from the transgene was not amplified and that we obtained only TEV sequences from viral genomes. For both genotypes, TEV and TEV-Δ*NIb*, the three amplicons were purified and sequenced by GenoScreen using BIGDYE 3.1 and a 96-capillars ABI3730XL sequencing system (Applied Biosystems) with overlapping readouts using the same set of primers as in Agudelo-Romero *et al*. (2008). Contigs were assembled using GENEIOUS version 4.8.

Experiment 2: To verify the presence of deletions, the full genome of each genotype was reverse transcribed using MMuLV RT (Fermentas) and primer 97-101-R, then PCR amplified using the ultra high fidelity DNA polymerase Phusion (Finnzymes) and primers 73-80-F and 97-101-R (Table 4-3). Deletions were confirmed by size analyses in 1% agarose gels.

4d. Virus quantification

We first performed RT-qPCRs assays to compare viral accumulation of TEV and TEV-Δ*NIb* genotypes. RT-qPCRs were performed as previously described in paragraph *2. Estimation of the in vivo recombination rate for a plant RNA virus*.

4e. High-throughput sequencing and data analysis

To analyse the diversity of deletions that could occur in NIb gene during Experiment 2, the full genome of each genotype was reverse transcribed using MMuLV RT (Fermentas) and primer R97-101, then PCR amplified using the ultra high fidelity DNA polymerase Phusion (Finnzymes) and primers F73-80 and R97-101. This amplification was repeated seven times to limit possible PCR artefacts impact on the high-throughput sequencing. Then, amplicons were sequenced (Illumina HiSeq 2000, Genoscreen) using a paired-end (2×100b) protocol. The reads obtained were already de-multiplexed, with adapters removed. Data were analysed with GSNAP (http://research-pub.gene.com/gmap/). GSNAP aligned both single-end and paired-end reads. It can detect short- and long-

distance splicing, where in our case deletions take place of exons. GSNAP was used with its default set of parameters: (i) we excluded reads that are mapped on 100 consecutive bases on TEV genome. (ii) On the rest of sequences, we searched sequences that mapped in two different virus genome regions with a minimum of eight bases in 5′ and 3′. (iii) We calculated the distance between those regions and to define deletion size and position. (iv) We measured the frequency of each found deletions. Finally, we obtained the number of reads for each deletion variant in R1 and R2 and we compared it to the total number of reads for each lineage.

4e. Analysis of mutants of TEV-ΔNIb obtained during Experiment 1

Each mutant was created "artificially" by site-directed mutagenesis using the Quickchange® II XL kit (Stratagene) and following the indications given by the manufacturer. Table 4-2 shows the mutation introduced in a TEV-Δ*NIb* genotype following Stratagene's recommendations. To minimize unwanted errors during the mutagenesis process, the kit

incorporates *Pfu* Ultra™ high fidelity Turbo DNA polymerase (Stratagene). The amplification conditions were 1 min at 95 °C (initial denaturation), followed by 18 cycles consisting of 30 s at 95 °C, 45 s at 65 °C and 18 min at 68 °C, and a final extension step of 28 min at 68 °C and a limited number of PCR reaction cycles to minimize the chance of appearance of undesired mutations.

Infectious plasmids of each mutant were linearized with *Bgl*II (Takara) and transcribed into 5′-capped RNAs. Transcripts were prepared as described previously. Two batches of five 4-week-old *Nt::*NIb plants were respectively inoculated mechanically on the third true leaf with TEV transcripts (20 - 30 µg) in inoculation buffer. In all case, the first symptoms appeared 6 - 7 dpi. Seven dpi, infectious saps were obtained grinding 500 mg of TEV mutants infected tissue in a mortar with 500 µL of grinding buffer; then three 4-week-old *Nt*::NIb plants for every TEV mutants were respectively inoculated mechanically on the third true leaf with 15 µL of the corresponding sap.

Finally, after 64 hours post inoculation (hpi), total RNA was extracted from the youngest leaves of each of the three plants per TEV mutant and then RNAs were purified. RT-qPCR was performed to compare viral accumulation of each mutant to the *TEV-ΔNIb* ancestor as described previously. For each genotype, a per day Malthusian growth rate was computedas $m = (1/t)\log(Q_t)$, where Q_t are the pg of TEV RNA per 100 ng of total plant RNA quantified at t = 2.6 dpi. Absolute fitness was then defined as $W = e^m$ (Crow and Kimura, 1970).

Position	Type of		Mutation	Localization
3618	G	→	T	6K1
3622	A	→	G	6K1
3629	C	→	T	6K1
6805	A	→	G	NIa-Pro
6805+3618				
6805+3622				
6805+3629				

Table 4-2. Selected mutants of one-week evolved TEV-Δ*NIb*.

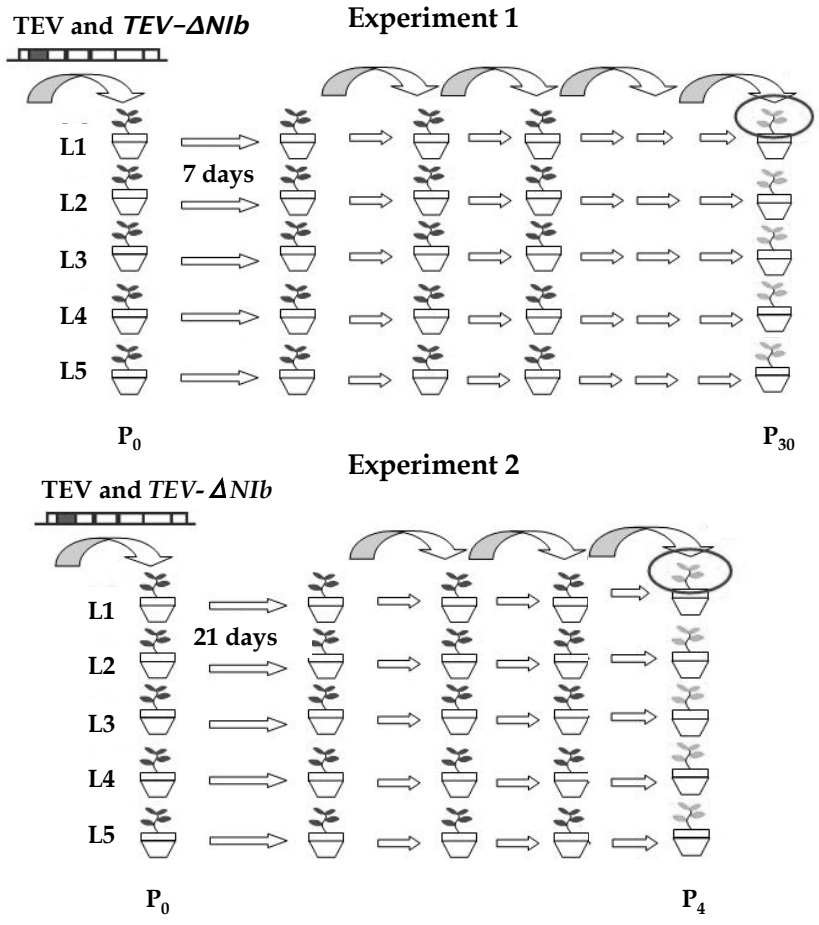


Figure 4-1. Schematic representation of the evolution process of TEV and TEV-Δ *NIb*.

Primers	TEV genome position (5′)	Sequence (5′ to 3′)
45-48-F	3542	TTGACGCTGAGCGGAGTGATGG
73-80-F	6377	TCATTACAAACAAGCACTTG
40-45-R	3877	ATCCAACAGCACCTCTCAC
67-77-R	6664	AATGCTTCCAGAATATGCC
92-96-R	8762	GCAAACTGCTCATGTGTGG
97-101-R	9473	CGCACTACATAGGAGAATTAG
F1	6911	GCAAACCTGAAGAGCCTTTTCAG
F2	7265	GCATGCTCATCACAAAGCTCAAG
F3	7591	GTGGATGATTTCAACAATCAATTTTATGAT
F4	8531	ACCAGCGTCAACACCAGCAC
F5	8215	GATCTGTCCCATTCCAAAATAGAAAC
TEV-CP-qPCR-F	9206	TTGGTCTTGATGGCAACGTG
TEV-CP-qPCR-R	9236	TGTGCCGTTCAGTGTCTTCCT

Table 4-3. Primers used to amplify or to sequence TEV genome.

RESULTS

Chapter 1

The rate and spectrum of spontaneous mutations in a plant RNA virus

The rate of spontaneous mutation is a key parameter to understand the genetic structure of populations over time. Mutation represents the primary source of genetic variation on which natural selection and genetic drift operate. Although the exact value of mutation rate is important for several evolutionary theories, accurate estimates are only available for a handful of organisms. In the case of plant RNA viruses, it has been repeatedly reported that their populations are highly genetically stable (Fraile *et al*., 1997; Herránz *et al*., 2008; Marco and Aranda, 2005; Rodríguez-Cerezo *et al*., 1991;) in comparison with their animal counterparts, although reports of higher substitution rates also exist (Fargette *et al*., 2008; Gibbs *et al*. 2008). This peculiar behavior might be due in part to stronger stabilizing selection, weaker immune-mediated positive selection (García-Arenal *et al*., 2001), the existence of strong bottlenecks during cell-to-cell movement and systemic colonization of distal tissues (Hall *et al*., 2001; Li and Roossinck, 2004; Sacristán *et al*., 2003;), severe bottlenecks during vector-mediated transmission (Ali *et al*., 2006; Betancourt *et al*., 2008; Moury *et al*., 2007), or differences in the replication mode

compared to lytic animal viruses (French and Stenger, 2003; Sardanyés *et al*., 2009). Indeed, the only two available direct estimates of mutation rates for plant viruses are both in the lower side of the range usually accepted for animal riboviruses: 0.10 - 0.13 per genome and generation for TMV (Malpica *et al*., 2002) and 0.28 for TEV (Sanjuán *et al*., 2009). However, none of these estimates is perfect. Although in the TMV experiments particular care was taken to measure mutation rate in a long target protected from the action of purifying selection (hence deleterious mutations remain in the population), uncertainties exist related to the number of infection cycles elapsed during the mutation-accumulation phase and the fraction of mutations that produced a selectable phenotype. In the case of TEV, the estimate should be taken as an upper limit because selection was operating during the mutation-accumulation phase. Furthermore, the estimate is in the same order of magnitude than the methodological error.

1a. Characterization of the mutant spectrum

Table 1-2 summarizes the spectrum of mutations characterized for the 472 clones sequenced (724992 nucleotides). Fifty-two mutations have been identified, 46 of which were nucleotide substitutions and six deletions. Four hundred twenty-six amplicons had no mutation, 41 carried a single mutation and five had two mutations. This distribution does not depart from the Poisson expectation (Kolmogorov-Smirnov test, $P = 1$). Among base substitutions, 33 were transitions and 13 transversions. Consistent with the principle that transitions are biochemically more likely than transversions, the maximum composite likelihood estimate of the overall transitions to transversions rates ratio was 2.161. This excess also occurs when purines (4.262) or pyrimidines (6.681) are considered separately. Indeed, the observed frequencies of transitions among purines and among pyrimidines are similar (Figure 1-4) and by far the most frequent type of mutation (Table 1-2, Figure 1-4). Therefore, we can conclude that TEV polymerase produces, on average, 2/3 transitions and 1/3 transversions. If purifying selection would

not be canceled out by NIb *trans*-complementation, the ratio would be more biased towards transitions because they are more often silent.

Under the observed mutational spectrum, the equilibrium base-composition achieved by mutation in the absence of selection would be 31.1% A, 25.4% U, 17.8% C, and 25.7% G. This distribution significantly deviates from what is expected just by sheer chance ($\chi^2 = 55.505$, 3 d.f., $P < 0.001$). The deviation is mainly driven by the unbalanced composition in purines, with a large excess of A (24.5%) that compensates for the large defect in G (–28.9%).

We have observed that 16 mutations were synonymous and 30 were nonsynonymous (including two stop codons). At least eight of the nonsynonymous substitutions could induce a major deformation on NIb folding by replacing polar or charged side chains by apolar ones (E20G, Q462P, H355L, and E507A) or apolar side chains by polar ones (F106S, G200S and W417R). Three substitutions (L143P, D146H and D276Y) lead to a strong change in the side chain length. Nine substitutions (D248N, A270V, D276Y, R283Q, I302L, D348N, H355L, T381I,

and Q387Stop) may be affecting the putative NIb active site (PFAM00680). Among deletions, three involved single nucleotides; one involved three contiguous nucleotides. To evaluate whether this pattern of synonymous and nonsynonymous changes is compatible with a model of neutral evolution, we have estimated the difference between substitution rates per nonsynonymous (d_N) and synonymous (d_S) sites (using Nei-Gojobori's modified method and bootstrap SEM). The observed value of $d_N - d_S = (5.537 \pm 4.133) \times 10^{-4}$ is not significantly different from zero ($z = 1.340$, $P = 0.090$), failing to reject the null hypothesis of neutral evolution and validating our methodology for protecting a viral sequence from purifying selection.

Type of mutation		Number		Substitution matrix			
Total		**51**		**A**	**U**	**G**	**C**
Base substitutions		**46**	**A**	**-**	**3**	**4**	**7**
	Transitions	**33**	**U**	**3**	**-**	**7**	**0**
	Transversions	**13**	**G**	**0**	**9**	**-**	**0**
	Synonymous	**16**	**C**	**10**	**2**	**1**	**-**
	Nonsynonymous	**30 (2 stops)**					
Deletions		**6**					
	1-nt	**3**					
	3-nt	**1**					

TABLE 1-2. Numbers of mutations by type and observed substitution matrix.

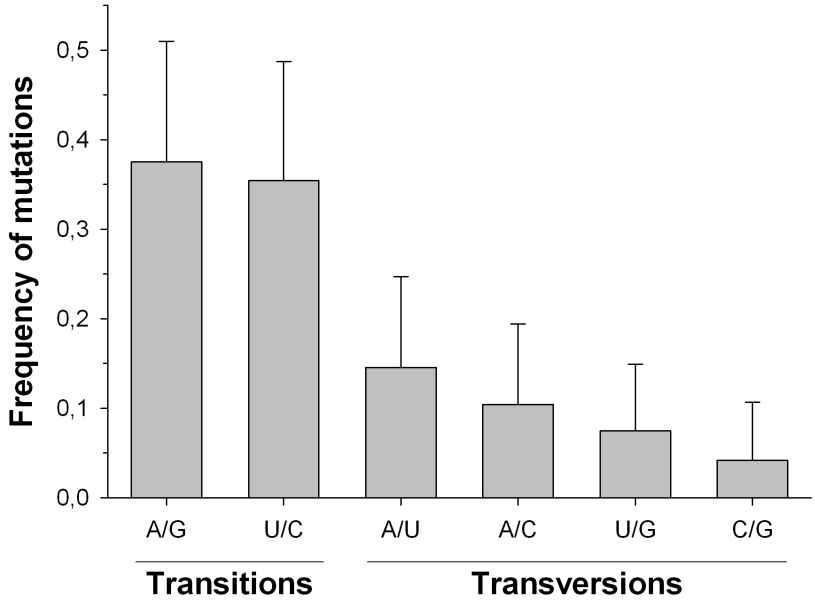


Figure 1-4. Observed frequencies for the different types of nucleotides substitutions. Each column groups mutations rendering complementary pairs and, thus can occur during the synthesis of the genomic or antigenomic strains. The LaPlace estimator of the frequency has been used to minimize the bias due to small sample size (Agresti and Coull, 1998). Error bars represent the 95% confidence interval for the estimator.

1b. Comparison of TEV mutant spectrum with that observed for other plant viruses

TEV spontaneous mutational spectrum differs in several aspects from the other only one reported for plant viruses, TMV (Malpica *et al*., 2002). First, TMV mutational spectrum is dominated by insertions and deletions (69% of all mutations belong to these categories). Deletions were both short (five cases with 1 – 3 nucleotides deleted) and long (seven cases with up to 100 nucleotides deleted). Insertions were also short (1 nucleotide) and large (4 cases with poly(A) insertions). In sharp contrast, only 9.8% of mutations in TEV mutational spectrum were short deletions, and not a single insertion has been observed. This significant difference (Fisher's exact test, $P < 0.001$) suggests either that TEV NIb is more processive *in vivo* than TMV replicase or that the difference is due to the experimental setup. In this regard, Malpica *et al*. (2002) used the MP protein expressed in *trans* as target for measuring mutation rate on the viral copy of MP. However, MP has a positive regulatory effect on the formation of TMV replication complex

(Beachy and Heinlein, 2000) and may favor template switching and a higher rate of deletions and insertions.

The second noticeable difference between both mutational spectra refers to the ratio of synonymous to nonsynonymous substitutions. TMV ratio is 1:10, whereas for TEV it is about five times larger (16:30). This difference may reflect that the method employed by Malpica *et al*. (2002) was less efficient than our method to protect deleterious point mutations from purifying selection, although this explanation is unsatisfactory given the large amount of deletions maintained in TMV populations. Nonetheless, it is honest saying that this 5-fold difference was not statistically significant (Fisher's exact test, $P = 0.146$).

A third difference is that the ratio of transitions to transversions was roughly 1.0 for TMV whereas it was > 2.0 for TEV. Given that it is biochemically easier to produce transitions than transversions, the deficit of the former type observed for TMV may reflect a preference of its replicase for transversions or imperfect sampling (Malpica *et al*., 2002).

Finally, Malpica *et al*. (2002) found striking the frequency of mutant genotypes carrying multiple mutations. The distribution of mutations per mutant TMV amplicon had a median of one and a range of 1 – 3. In our case, the distribution had also a median of one and a range of 1 – 2. From a statistical point of view, both distributions are undistinguishable in shape (Kolmogorov-Smirnov test, P = 0.199) and location (Mann-Whitney test, P = 0.929). Therefore, we would not consider striking finding a minor proportion of amplicons carrying more than one mutation: it is just what is expected from the Poisson model.

1c. Estimates of the mutation rate

Applying the first method described in the Material and Methods section, we have obtained 19 independent estimates of the spontaneous mutation rate. The estimates ranged from $0 \leq \mu \leq 1.340\times10^{-5}$ m/b/g. The distribution of estimates was Gaussian (Kolmogorov-Smirnov test, P = 0.944) with mean $\overline{\mu} = 4.754\times10^{-6}$ m/b/g and standard deviation $s_{\mu} = 3.540\times10^{-6}$ m/b/g. As a way to evaluate the statistical power associated with this estimate,

we constructed the 95% confidence interval around the mean as $3.048\times10^{-6} \leq \overline{\mu} \leq 6.460\times10^{-6}$ m/b/g, which excludes the zero. Therefore, according to these values, we conclude that the genomic mutation rate of TEV is 0.045±0.008 (SEM) per generation.

Next, we sought for applying the lethal alleles method. To compute the first term in Equation 2, we proceeded as above and fitted the observed number of deletions per amplicon per plant to a Poisson distribution, obtaining 19 independent estimates of μ_{indel}. The average rate of deletion mutations was $\mu_{indel} = (3.787\pm1.558)\times10^{-7}$ deletions/b/r. Next we focused in the computation of the second term in Equation 2. Only two out of the ~725 Kb sequenced were stop codons (hence $K = 2$ in Equation 2). As a consequence of codon usage bias, the actual number of NSMT in our sample is 7.46% instead of the expected 10.34%. Taking this source of bias into consideration and after correcting for the three possible nucleotide substitutions per site, the second term in Equation 2 results in $(6.295\pm0.556)\times10^{-5}$ m/NSMT/r. Therefore, the estimate of the spontaneous lethal mutation rate is $\mu_L = (6.299\pm0.558)\times10^{-5}$

m/b/r or, expressed into the per genome scale, 0.601±0.053 per replication event.

This μ_L value is 13.4 times higher than the μ estimate obtained using the first method, being the difference highly significant (2-samples *t*-test, $t_{36} = 10.328$, $P < 0.001$). What may produce this discrepancy? The lethality method has the advantage of being independent from generation time. However, it is strongly dependent on whether the mutations considered are truly lethal. Deviations from this assumption imply that the estimate immediately becomes an upper-limit of the true value. In infected cells wherein multiple genomes may coexist, genomes carrying deletions or NSMTs can still be replicated by the pool of polymerases, encapsidated into wild-type capsides and moved cell-to-cell and even systemically. In other words, complementation with functional proteins makes lethal alleles behave as effectively neutral. An alternative consideration is that, as defined above, one generation involves many replication rounds. Assuming that μ_L has not been biased by complementation, the 13.4 fold difference between estimates can be interpreted as the number of replication events within

an infected cell. Nonetheless, we can conservatively conclude that the above μ_L estimate must be taken as an upper-limit estimate of the true mutation rate.

1d. Comparison of TEV mutation rate with those obtained for other RNA viruses

The only previous direct estimate of mutation rate for another plant virus, TMV, was in the range 1.452×10^{-5} – 2.060×10^{-5} m/b/g (Malpica *et al*., 2002), values lying well within our two estimates. In a more recent study, Sanjuán *et al*., (2009) estimated TEV upper-limit mutation rate as $(2.96\pm0.32)\times10^{-5}$ m/b/g, a value also within our both estimates. In the same study, these authors performed a literature survey for upper-limit estimates of per site mutation rates for four plant viruses. All the compiled studies were methodologically similar and relied on characterizing the mutant spectrum from individual plants inoculated with a viral clone (i.e., close to zero starting genetic diversity). In neither of these studies was genetic variation protected from purifying selection (Sanjuán *et al*.,

2009). The median upper-limit mutation rate estimated was 7.74×10^{-4} m/b/g, which was in the range of values estimated for animal RNA viruses and some bacteriophages (Drake and Holland, 1999) but still 12.3-fold larger than our upper-limit estimate.

Our data allow us to conclude that the mutation rate of TEV is slightly lower than previously estimated by Sanjuán *et al*. (2009) and very similar to the only other direct estimation available for another RNA plant virus, TMV (Malpica *et al.*, 2002).

1e. Potential pitfalls and considerations

In this study we have used a high-fidelity DNA polymerase to minimize the probability that observed mutations may be due to PCR errors. PrimeSTAR HS DNA polymerase is about two times more accurate than *Pfu* due to its improved and robust 3′ → 5′ exonuclease activity and its error rate has been estimated to be 1.60×10^{-6} m/b/PCR cycle (catalog.takara-bio.co.jp). Since we run PCRs for 30 cycles, we

expect an error rate per amplicon of 4.8×10^{-5} m/b. Henceforth, we may expect ≈ 34 mutations in our sample to be due to errors during PCR. Unfortunately, this is not the only source of error; the error rate of MMuLV RT is around 3.3×10^{-5} m/b/r (Arezi and Hogrefe, 2007), which means that we may expect as well ≈ 24 mutations to be produced during retrotranscription. Since we have obtained 52 mutations, someone may argue that all of them must result from errors during either retrotranscription or PCR amplification. This being the case, the mutation rate of TEV would be $< 10^{-9}$ m/b/g, a value that is, by all means, absurdly low. Furthermore, the estimate of the error rate of PrimeSTAR HS polymerase should be taken with strong precaution. It is surprising the manufacturer' claims that the enzyme has improved fidelity compared with *Pfu* but the estimate they provide is undistinguishable from values reported for *Pfu*, 1.3×10^{-6} m/b/PCR cycle (Cline *et al*., 1996; Bracho *et al*., 1998). Therefore, we can conclude that even if unwanted mutations are produced during the RT-PCR amplification, the estimated mutation rates are still on the low side of previous reports.

Chapter 2

Estimation of the *in vivo* recombination rate for a plant RNA virus

Mechanistically, recombination in RNA viruses can be defined as an exchange of genetic material between at least two different viral genomes subjected to a replicase-driven template-switching action. Various RNA virus recombination rates have been quantitatively estimated and showed high variability, ranging from 1.4×10^{-5} recombination events per site and generation for HIV-1 within a host (Neher and Leitner, 2010) to 4×10^{-8} in the case of HCV (Reiter *et al*., 2011). Froissart *et al*. (2005) reported the first *in planta* recombination rate for a plant virus during colonization of its individual multicellular host for CaMV. They found that recombination for this pararetrovirus was frequent and estimated its rate to be 4×10^{-5} events per nucleotide site and per replication cycle. Unfortunately, *in planta* estimates for recombination rate of single-stranded positive-sense RNA viruses, the most common described plant viruses, are still missing. In a very recent report, Pita and Roossinck (2013) describe frequent recombination events for *Cucumber mosaic virus* (CMV), but their experimental design does not allow performing an estimate of the recombination rate.

In this study, we provide an estimation of TEV recombination rate during a single infection passage in its primary hosts *N. tabacum*. Our strategy was based in inoculating equimolecular mixtures of pairs of engineered genotypes carrying different neutral markers and characterizing the progeny resulting after the systemic infection of tobacco plants. Potyviruses represent a particularly interesting model system for studying recombination, since two confronting observations have been made for this group of viruses. On the one side, phylogenetic evidences support specially high frequencies of recombination (Chare and Holmes, 2005; Reverse *et al*., 1996). On the other side, it has been reported that during mixed infections isolates from the same potyvirus form spatial patches in leaves, with no or very limited mixing (Dietrich and Maiss, 2003). The question that rises is: if different genotypes rarely coinfect the same cell, how can recombinant genotypes be so often created?

2a. Results of the experiment

Table 2-3 summarizes the results of the experiment. Notice that we failed to produce infection in two of the plants inoculated with the mixture *Pme*I/*Eco*47III and in one inoculated with the mixture *Sal*I/*Sgf*II. Averaging across plants infected with the same mixture, the lowest frequency of recombinant genotypes was observed 15 dpi for the *Sal*I/*Sgf*II pair (f = 2.04%), whereas the largest was observed 15 dpi for the *Asc*I/*Pme*I combination of markers (f = 16.29%). When frequencies of recombinant genotypes were transformed into recombination rates per site, the lowest value still corresponded to the *Sal*I/*Sgf*II pair 15 dpi (r = 8.88×10^{-6} r/s). However, the highest recombination rate was observed for the *Pme*I/*Eco*47III pair 45 dpi (1.23×10^{-4} r/s).

Data were fitted to a GLM using a linear response and taking time (dpi) and marker combination as factors. The grand mean r was estimated to be $(5.90\pm1.53)\times10^{-5}$ r/s and significantly different from zero (χ^2 = 14.857, 1 d.f., P < 0.001).

Combination	Fragment size (nt)	dpi	Plants analyzed	*f* (±1 SEM)	*r* (±1 SEM) (×10^{-5})
*Asc*I/*Pme*I	3334	15	5	0.163±0.046	4.886±1.237
		45	5	0.076±0.047	2.292±1.422
*Pme*I/*Eco*47III	1234	15	4	0.145±0.088	11.893±7.218
		45	4	0.149±0.098	12.247±8.041
*Eco*47III/*Sal*I	2197	15	5	0.041±0.003	1.867±0.149
		45	5	0.074±0.020	3.350±0.920
*Sal*I/*Sgf*I	2298	15	5	0.020±0.000	0.888±0.013
		45	4	0.037±0.000	1.610±0.000

Table 2-3. Observed frequency of recombinants and recombination rates

2b. Stability of recombination rate along infection time

No significant time effect was observed neither in *f* nor in *r* (GLM: $\chi^2 = 0.016$, 1 d.f., $P = 0.900$), thus suggesting that the frequency of recombinant genotypes did not increase with time after the first 15 dpi. A plausible explanation for this result is that the chances for recombination do not increase as infection progress and that all recombination events take place early during infection. This may be caused by several possible mechanisms. One of such mechanisms being that the multiplicity of infection (*MOI*) remains constant after 15 dpi. Although this possibility has not been yet explored for TEV or

other potyviruses, it was shown that *MOI* is higher during the early stages of the infection and then decreases as infection progresses in the case of plant viruses as different as TMV (González-Jara *et al*., 2009) and CaMV (Gutiérrez *et al*., 2010). Since TEV has fully colonized systemic tissues well before 15 dpi, it is reasonable that MOI would remain more or less constant from 15 to 45 dpi. A possible second mechanism to minimize the chances of recombination is the above-mentioned spatial segregation observed between strains of the same potyvirus during mixed infections (Dietrich and Maiss, 2003; Zwart *et al*., 2011, 2012). Obviously, strains carrying each marker must coinfect the same cell in order for recombination to occur. In such scenario, coinfection will be restricted to the very residual number of cells found in the contact edge between growing foci. Furthermore, if different leaves are colonized by different viral strains, as amply described for plant viruses (*e.g*., French and Stenger, 2005; Hall *et al*., 2001; Jridi *et al*., 2006), then chances for coinfection will be further reduced.

2c. Variability in recombination rate along TEV genome

The above GLM found a significant effect of the marker combination in the observed recombination rate (χ^2 = 13.103, 3 d.f., P = 0.004). This difference was entirely driven by the ca. 6-fold larger recombination rate observed for the *Pme*I/*Eco*47III genomic region (Table 2-3): 1.58×10^{-4} r/s versus an average of 2.62×10^{-5} r/s across the other three regions. All pairwise comparisons involving *Pme*I/*Eco*47III were significant ($P \leq 0.007$), whereas all other were not ($P \geq 0.538$). The 1215 nucleotide fragment defined by the markers *Pme*I and *Eco*47III is entirely within the CI cistron. The CI is an ATP-dependent helicase (Laín *et al*., 1990, 1991) that binds to mature virions and promotes cell-to-cell movement via direct interaction with the movement protein P3N-PIPO (Vijayapalani *et al*., 2012; Wei *et al*., 2010). CI helicase has a modular structure formed by seven conserved motifs (Kadaré and Haenni, 1997). Modularity may facilitate the viability of recombinants CI proteins as far as the junctions between modules correspond to the precise sites for recombination. This being the case, recombination will bring together combinations of modules from different sources, each

module retaining its functionality. By contrast, recombination in a non-modular protein would result in sequences that may not work well together. Unfortunately, GenBank contains not enough information on CI intraspecific sequence variability for TEV as to test this tantalizing hypothesis. Nonetheless, it is interesting that the 5'-terminal half of *Turnip mosaic virus* (TuMV) CI cistron has been previously described as a recombination hotspot (Ohshima *et al.*, 2007).

2d. A possible drawback and a solution: unbalanced mixtures of markers

The efficiency of the method we have used to estimate recombination rate is strongly dependent on the mixture of the two markers being well balanced during infection. We carefully quantified and mixed both genotypes to ensure the 1:1 initial ratio. However, this ratio may have changed during infection for many reasons, *e.g.* due to subtle fitness differences, or because bottlenecks during inoculation or during systemic movement. If one of the markers dominates over the other, then recombinants will be detected less often than in both markers

remain equally abundant. In other words, if the two marked genotypes are in proportion 1:1 during the whole experiment, *r* will be properly estimating recombination rate. In contrast, if the ratio between both marked genotypes largely deviates from 1:1 as infection progresses, *f* will be largely underestimated and so does *r*. Indeed, this effect of unweighted mixtures on the estimate of recombination rate has already been reported for MMuLV (Anderson *et al*., 1998). To evaluate the effect of the marker composition we plotted each estimate of recombination rate against the observed ratio of the less abundant to the most abundant parental genotype (Figure 2-3). As shown in Figure 2-3, the mixtures between both parental strains were not always well balanced at the moment of sampling, with ratios spreading in the wide range 0.01 to 1. A significant linear regression exists between this ratio and *r* (R = 0.772, $F_{1,35}$ = 51.699, P < 0.001), confirming that the more departure from the hypothetical 1:1 parental ratio, the lower the estimated recombination rate. The regression equation can be used to correct for the effect of the unbiased mixtures. In the worse scenario, that is when the ratio has the largest possible departure from 1:1 ratio, the

(understimated) expected recombination rate is simply a = (2.19±1.25)×10^{-5} r/s (regression model parameters shown in the legend of Figure 2-3). By contrast, in the optimal situation, namely when the ratio between markers equals one, then the expected recombination rate is $a + b$ = (4.50±0.72)×10^{-4} r/s, that is, 7.6 times larger than the genome-average value we reported in the previous sections.

Taking this last number as our best estimate for TEV overall recombination rate, and given the observed lack of time effect on recombination rate, we can rescale the estimate to the more biologically meaningful units of per generation. Given a generation time of 2.91±0.58 generations per day (Martínez *et al.*, 2011), 15 dpi are equivalent to 43.65±8.70 generations and therefore, rg = (1.03±0.37)×10^{-5} r/s/g.

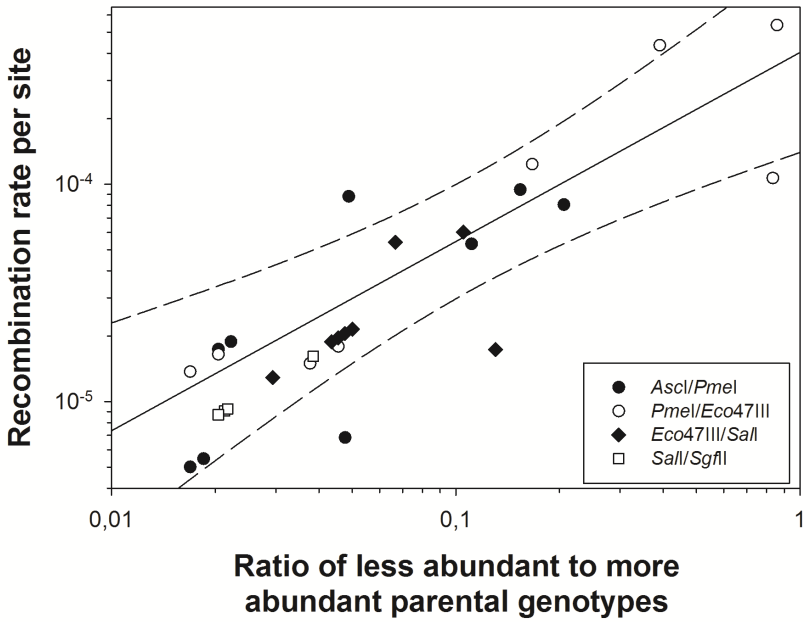


Figure 2-3. **Effect of the mixture bias in the estimate of recombination rate**. The ratio between the less- and the most-abundant genotypes is shown in the ordinates axis, whereas the observed recombination rates are shown on abscises. Different symbols are used for each marker pair. The plot is shown in log-log scale to stretch out the data points. The solid line represents the regression line $a + b \times RATIO$, whereas the dashed lines correspond to the 95% confidence intervals. The parameter values (±1 SEM) are: $a = (2.190 \pm 1.249) \times 10^{-5}$ and $b = (4.276 \pm 0.595) \times 10^{-4}$.

2f. Dealing with the formation of recombinant molecules during RT-PCR

A worrisome aspect of PCR based studies of virus variability is the phenomenon of PCR-mediated recombination, or chimera formation (Meyerhans *et al*., 1990). In a recent study, Lahr and Katz (2009) have shown that chimera formation was minimized by using the Phusion DNA polymerase, no more than 30 amplification cycles and, very critical, a low initial template concentration. To determine whether the RT-PCR conditions used in our experiments may have favor the formation of chimeras, we proceeded as follows. First, we prepared TEV RNAs by in vitro transcription of the plasmids containing the *Pme*I and *Eco*47III markers and mixed them in a 1:1 (w:w) ratio. These markers were chosen for this experiment because they show the highest recombination rate. The concentration of this RNA mixture was spectrophotometrically determined to be 3.92×10^{11} RNA molecules/ μ L. Five 10-fold serial dilutions of this mixture were made; thus the largest dilution contained 3.92×10^{3} RNA molecules/ μ L. Each of these dilutions was then used as template for an RT-PCR experiment

(30 cycles, Phusion DNA polymerase). The PCR products were purified, cloned, transformed into *E. coli*, and a number of clones per amplification experiment (ranging 28 – 44, median 35) were analyzed by digestion with *Pme*I and *Eco*47III. Figure 2-4 shows the frequency of chimera molecules as a function of the number of TEV template RNA molecules in the RT-PCR reaction. The relationship between these two variables is sigmoidal: at RNA template concentrations smaller than 5.46×10^{7} molecules, the estimated inflexion point of the fitted curve, no chimeras have been observed. However, for template concentrations larger than this critical value, the frequency of chimera molecules sharply increases to ca. 50%.

This result confirms that, all else being equal, the concentration of template RNA molecule determines the chances of generating chimera molecules during RT-PCR. Therefore, the key question now is how many TEV template RNA molecules we used in our experiments. The quantitative RT-PCR experiments done to test the neutrality of the five markers showed that, on average, the concentration of TEV on infected plants was of 3.77±0.91 pg per ng of total RNA

extracted, which is equivalent to $(7.38\pm1.77)\times10^5$ TEV genomes per ng of total RNA. Since we seeded our RT-PCR experiments on 100 ng of total RNA, this means we added 10^7 - 10^8 TEV genomes. Therefore, we cannot rule out that our estimates may be affected to some extent by PCR artifacts.

Lets make two additional considerations. First, the sigmoidal relationship shown in Figure 2-4 tells that for 10^7 template RNA molecules the expected frequency of recombinant molecules would be approximately $f = 0.23$. As Table 2-3 shows, the observed f values are all smaller than this figure and, therefore, PCR artifacts seem not to be as common as suggested in the previous paragraph. Second, if only PCR errors would contribute to generate recombinant molecules, then observed frequency would be more or less the same regardless the pairs of markers used. The observed differences among pairs further diminish a possible role of PCR artifacts in our observations.

In conclusion, despite all these considerations, the truth is that we cannot completely rule out a possible contribution of PCR to the generation of chimeric molecules, we cautiously

would consider our estimate as an upper limit value of TEV recombination rate.

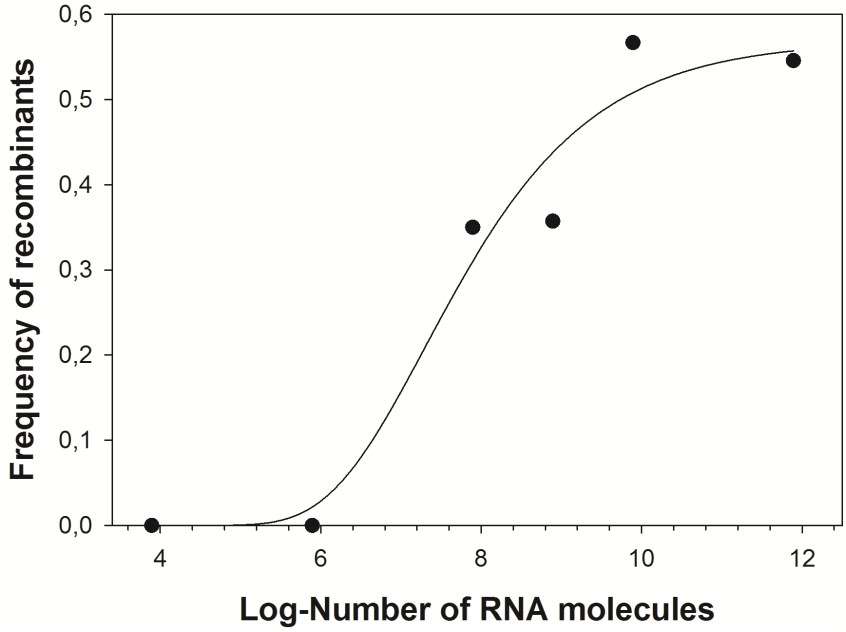


Figure 2-4. **Effect of the number of TEV template RNA molecules in the formation of chimera molecules during RT-PCR**. The solid line represents the Chapman sigmoidal equation $a\left(1-e^{-b\times RNA}\right)^{c}$. The parameter values (±1 SEM) are: a = 0.569±0.073, b = 0.840±0.415 and c = 459.521±1444.342 (R^2 = 0.962, $F_{2,3,}$ = 37.650, P = 0.008).

2g. Comparison with estimates of recombination rate for other plant RNA viruses

The only previous estimate of recombination rate for a plant virus during real infection conditions was reported by Froissart *et al*. (2005) for CaMV, being rg in the range 2×10^{-5} - 4×10^{-5} r/s/g. This value is in the same order of magnitude than our estimation for TEV.

The recombination rate for the tripartite *Brome mosaic virus* (BMV) has been evaluated in several studies (Bruyere *et al*., 2000; Olsthoorn *et al*., 2002; Urbanowicz *et al*., 2005). Unfortunately, comparison of these studies and our own one is not straightforward for several reasons: (i) none of these studies made a rigorous statistical data analyses and just reported counts of recombinant and parental genomes, (ii) each study focused on a particular genomic region, which may or may not be representative for the whole genome, and (iii) no insight on the number of generations per day exist for BMV. Nonetheless, it is still possible to compute *f* and *r* from the numbers provided in different tables and figures of these studies. Bruyere *et al*. (2000) introduced several restriction site markers in BMV

RNA3, mixed at equal proportions pairs of markers and inoculated them in the non-natural host *Chenopodium quinoa*. Variable numbers of local lesions were analyzed for the presence of parental and recombinant genomes. From the data contained in their Table 1, and averaging across the four experimental replicate, we estimated $r = (1.41 \pm 0.63) \times 10^{-4}$ r/s. In a follow-up study Urbanowicz *et al*. (2005) used the same method (with some minor experimental variations) to estimate the recombination rates on RNA1 and RNA2. Using the data contained in their Figure 3, we estimated $r = (1.74 \pm 0.04) \times 10^{-4}$ r/s for RNA1. Similarly, using the data shown in their Figure 2, and averaging across experiments, we found that $r = (2.49 \pm 0.40) \times 10^{-4}$ r/s for RNA2. The results for the three segments are homogeneous (Kruskal-Wallis test, $H = 2.444$, 2 d.f., $P = 0.295$) and thus we can estimate an average genome-wide recombination rate per site for BMV of $r = (2.10 \pm 0.27) \times 10^{-4}$ r/s. Interestingly, this value is very much consistent with our estimate for TEV. In an additional study, Olsthoorn *et al*. (2002) evaluated the probability of recombination between the homologous 3' non-coding regions of the three BMV genomic

segments (*i.e.*, recombination between segments). Using the data reported in their Table 1, we calculated $r = (1.62 \pm 0.15) \times 10^{-2}$ r/s for this small regulatory region, a value that is significantly larger than those observed for BMV coding sequences.

2h. Relationship between mutation and recombination rates for RNA viruses

A previous theory suggested that RNA virus high mutation rates are also a likely consequence of the tradeoff between fast replication and accuracy (Belshaw *et al*., 2007; Elena and Sanjuán, 2005), then a positive correlation must exist between mutation and recombination rates across RNA viruses. To test this prediction, we search the literature for RNA viruses for which estimates of mutation and recombination rates exist. Unfortunately, the set is just limited to the following seven cases: HIV-1, HCV, MHV, MMuLV, PV, SNV, and TEV. The source of recombination rates values are provided in the legend of Figure 2-5, whereas mutation rates for all these viruses were taken from Sanjuán *et al*. (2010). When more than one estimate existed for one virus, the average was taken. Figure 2-5

illustrates the relationship between mutation and recombination rates. A positive correlation exists between both traits, which becomes highly significant if the discordant data point for HCV is removed from the computation (Spearman's ρ = 1.000, 4 d.f., $P < 0.001$). Removing HCV may be justified because the low recombination rate measured by Reiter *et al*. (2011) has already been questioned by others (González-Candelas *et al*., 2011). Therefore, we provide support to the hypothesis that TEV recombination rate may be a side effect of selection for fast but error prone replication, rather than being selected for the fitness advantages it may provide in the long run.

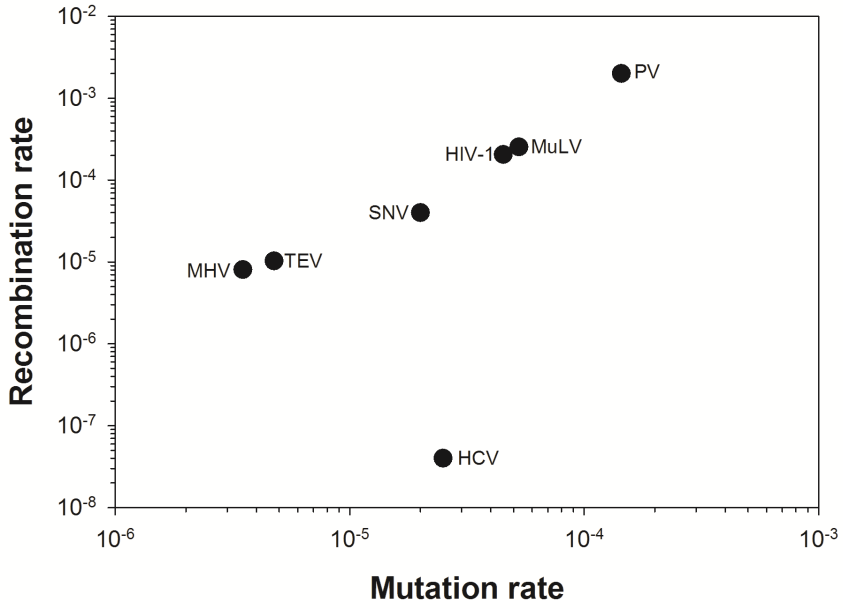


Figure 2-5. Relationship between mutation and recombination rates for seven RNA viruses and retroviruses. Mutation rates were taken from Sanjuán *et al*. (2010). Recombination rates were taken from: HIV-1 (Batorsky *et al*., 2011; Jezt *et al*., 2000), HCV (Reiter *et al*., 2011), MHV (Baric *et al*., 1990), MMuLV (Anderson *et al*., 1998; Zhuang *et al*., 2006), PV (Duggal *et al*., 1997; Jarvis and Kirkegaard, 1992; King, 1988), SNV (Hu and Temin, 1989), and TEV (this study).

Chapter 3

Estimation of the multiplicity of infection (*MOI*) during RNA virus dynamic infection in a multicellular host

The cellular multiplicity of infection (*MOI*) is a key parameter for describing the interactions between virions and cells, for predicting the dynamics of mixed-genotype infections, and for understanding virus evolution. Two recent studies have reported *in vivo MOI* estimates for TMV and CaMV (González-Jara *et al*., 2009; Gutiérrez *et al*., 2010), using different approaches to measure the distribution of two virus genotypes over host cells. Here, we perform a series of experiments to estimate the cellular *MOI* during TEV infection of tobacco taking into account two important parameter in the dynamics: space and time.

We isolated protoplasts from the third, fifth, sixth, and seventh true leaf at three, five, seven, and 10 dpi, with five replicate plants for each time point. We did not analyze the fourth leaf because under the conditions used this leaf does not show any infection (data not shown). We subsequently refer to these leaves as leaf one (inoculated leaf), three, four and five. Flow cytometry was subsequently used to determine which cells were uninfected, infected with only TEV-BFP or TEV-Venus, or infected with both variants (see *Materials and*

Methods). We could therefore quantitatively track the number of TEV-infected cells in an infected plant over space and time, for two virus genotypes. Then, we used flow cytometry to detect the presence of fluorescently labelled virus variants in thousands of individual cells and we used model selection to identify the best-supported model for making *MOI* estimates.

Overall, our data and analyses confirmed that *MOI* estimation methods must take into account differences in susceptibility or exposure of cells to virions, whereas there was, surprisingly, no evidence for spatial segregation of viral genotypes.

3a. Total number of TEV-infected cells over space and time

In contrast to previous reports (González-Jara *et al*., 2009; Gutiérrez *et al*., 2010), the frequency of virus-infected cells was low (mean ± SD: 0.100±0.073), with the highest level of infection observed in any sample being 0.306 (leaf four at 7 dpi) (Figure 3-2a-d). The frequency of cells infected by both virus variants was also low (mean ± SD: 0.012±0.023), with the highest level of

coinfection observed in any sample being 0.112 (leaf four at 7 dpi) (Figure 3-2a-d). *Prima facie* these results suggest that the cellular *MOI* is low, given that if a small number of virions infect each cell there will be few cellular coinfections with both viral variants. At 3 dpi, few cells were infected in any leaf, with the greatest number of infections being found in the leaves one and four (Figure 3-2e). This surprising observation can be explained by the occurrence of limited, relatively slow TEV expansion at the macroscopic level in the inoculated leaf (Zwart *et al.,* 2012), combined with fast egress from the inoculated leaf to leaf four (Lafforgue *et al*., 2012). Both infection and coinfection appear to increase over time in the different leaves, although leaf three only shows very low levels of infection. The lack of infected cells in leaf three probably comes about because this leaf has already made a sink-source transition, impeding access to a phloem-transported virus. Infection progresses slower in leaf one than in leaves four and five. Leaf four becomes infected before leaf five, but the dynamics in these two leaves appear to otherwise be very similar. These results are therefore in good agreement with macroscopic observations of

infection dynamics (Dolja *et al.*, 1992; Lafforgue *et al.*, 2012; Zwart *et al.*, 2012).

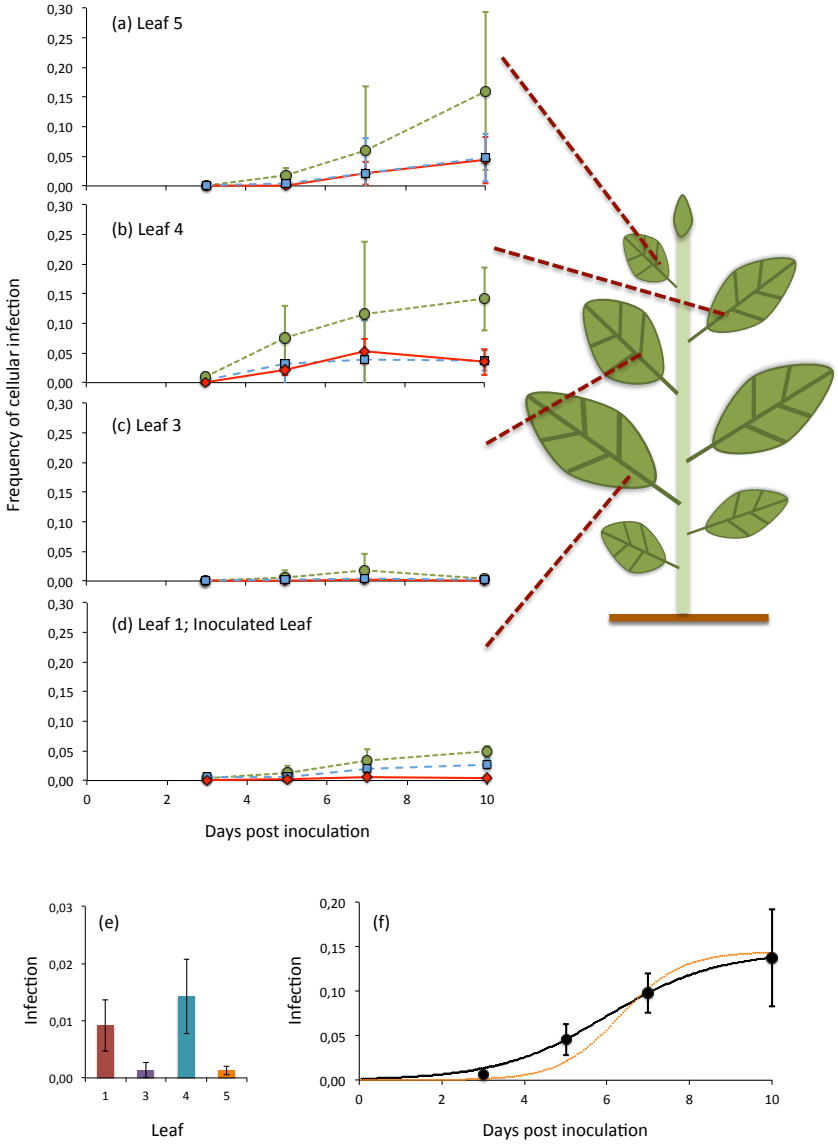


Figure 3-2: Infected cells over space and time. In panels (a-d), the observed frequencies of cellular infection in leaves five,

four, three and one are given for TEV-Venus only (finely dotted green line with circles), TEV-BFP only (coarsely dotted blue line with squares) and coinfections with both variants (continuous red line with diamonds). The abscissae represents dpi, whilst the ordinate is the frequency at which a particular infected-cell type was observed. Error bars represent one standard deviation for all panels in the figure. Note that leaf two was not included because it does not become infected. In Panel (e), the frequency of infected cells (TEV-Venus, TEV-BFP and both virus variants) at day three is given for different leaves. In Panel (f), the frequency of infected cells (ordinate) over time (abscissae) is given, for the data pooled from all leaves (black circles). The black line indicates a logistic model fitted to the data, whereas the brown line indicates a scaled logistic growth curve fitted to RT-qPCR data in a similar experiment (Zwart *et al.*, 2012). The logistic model fitted the data well, and suggests that in the final time point the plant is approaching κ, the carry capacity. The RT-qPCR-based curve is surprisingly similar to infected-cell curve, although at 5 dpi predicted TEV RNA levels appear to be relatively lower than the proportion of infected cells. This discrepancy may depend on the methodology used, or there may be a large number of cells that are in early infection, when the fluorescent marker protein is expressed by viral RNA accumulation levels are still low.

3b. Cellular MOI estimation

The cellular *MOI* can be estimated from our data, as has been previously done for two plant viruses with a similar experimental setup (González-Jara *et al*., 2009; Gutiérrez *et al*., 2010). Estimates of *MOI* can, however, be strongly influenced by the estimation method (Zwart, Tromas and Elena, unpublished manuscript). Model selection was therefore performed on a set of nine *MOI*-predicting models, by testing which Poisson-based model best predicted the relationship between the fractions of uninfected and mixed-variant-infected cells. The models incorporated spatial segregation of virus genotypes, spatial aggregation of infected cells, coinfection exclusion at the cellular level and combinations of these effects. We could thereby identify the best model to make *MOI* estimates (Tables 3-1 and 3-2).

The best-supported model incorporated a leaf-dependent aggregation factor ϕ (Table 3-2) suggesting that the aggregation of virus-infected cells is a key element for describing *MOI*. Infected cells are probably aggregated by virtue of their immobility and because TEV spreads by cell-to-

cell movement within a leaf. Surprisingly, models incorporating spatial segregation of genotypes were never selected. Our results therefore suggest that although virus-infected cells are clearly aggregated, the lower than expected rates of mixed-variant infections are not due to aggregation of different viral variants. In other words, aggregation in viral infection disrupts the Poisson-based relationship between the fractions of uninfected and mixed-variant-infected cells observed, whilst the well-mixed assumption for viral genotypes is never rejected during the course of infection.

We then derived predictions of *MOI* using the best-supported model (Figure 3-3a). As could be expected from the low rates of cellular infection and mixed-variant infection (Figure 3-2a-d), the predicted *MOI*s were low, ranging from 1.006 (leaf three, 3 dpi) to 1.646 (leaf four, 7 dpi). Note that we report the estimated *MOI* value in infected cells only (*i.e.*, m_I in *Materials and Methods*), which has a minimum value of 1. The corresponding range of *MOI* values calculated over the whole population of infected and uninfected cells (m_T) is 0.012 (leaf three, 3 dpi) to 1.071 (leaf four, 7 dpi). To help interpret

estimated *MOI* values, we provide model predictions of the frequencies at which cells are infected by a particular number of virions, for low, intermediate and high levels of infection observed here (Figure 3-3c-e).

(a)

Parameter	Estimate	95 % Conficence Interval
ψ_1	0.288	[0.224-0.314]
ψ_3	0.552	[0.487-0.608]
ψ_4	0.622	[0.536-0.648]
ψ_5	0.694	[0.627-0.716]

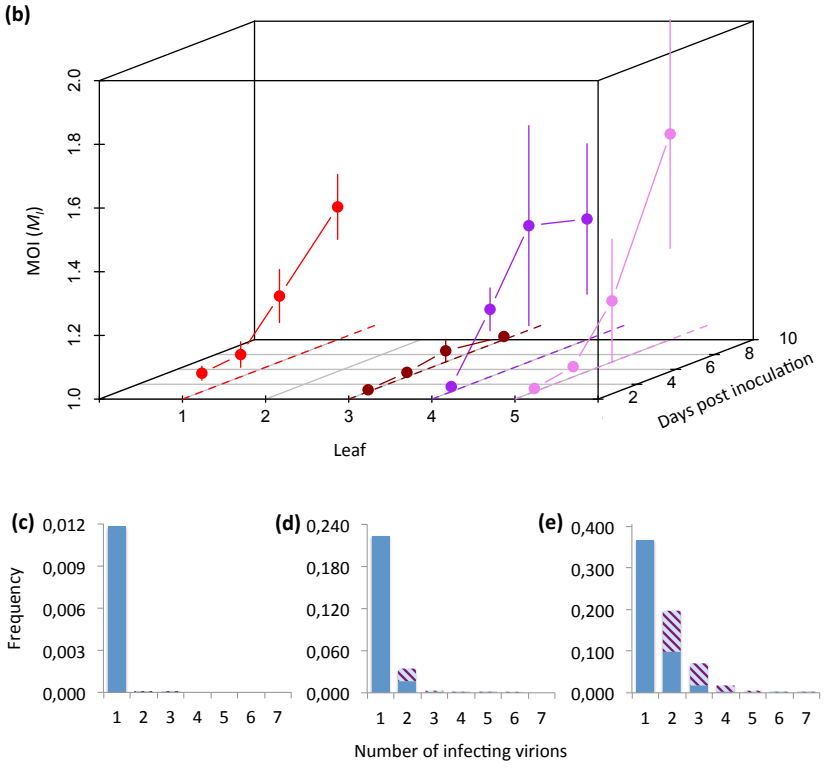


Figure 3-3: Cellular *MOI*. Panel (a) gives estimates of ϕ, the leaf-dependent infection aggregation parameter, for the best-

supported model. Panel (b) gives estimates of cellular *MOI*, for different times post inoculation and in different leaves. Note that the reported *MOI* value is m_I, the MOI in infected cells only, which has a minimum value of 1. *MOI* is initially very low and gradually increases, never reaching 1.7. *MOI* values for the final timepoint (10 dpi) are similar for leaves one, four and five, whilst remaining very low for leaf three, which hardly becomes infected. In panels (c-e), model predictions for the frequency at which cells are infected by a certain number of virions is given. The blue section of the bar indicates the frequency of infection by only one virus variant, whereas striped area indicates mixed-variant infection, assuming a 1:1 ratio of virus variants. Panel (c) gives this prediction for the lowest *MOI* value (m_I = 1.006), panel d for the mean *MOI* (m_I = 1.171), and panel e for the highest *MOI* value (m_I = 1.646). Estimated *MOI* values are low, but the number of infecting virions is assumed to follow a Poisson distribution. Hence even at the low mean *MOI* some cells will be infected by 2 or more virions, allowing for mixed-variant infections.

Model	Parameter estimates [95% CI]
1	-
2	ϕ = 0 [0-0.017]
3	ϕ = 0.650 [0.588-0.660]
4	ϕ_1 = 0.288 [0.224-0.314] ϕ_3 = 0.552 [0.487-0.608] ϕ_4 = 0.622 [0.536-0.648] ϕ_5 = 0.694 [0.627-0.716]

5	Ω = 1 [*] τ = 0 [*]
6	ϕ = 0 [0-0.063] ϕ_1 = 0.288 [0.224-0.328] ϕ_3 = 0.552 [0.442-0.608] ϕ_4 = 0.622 [0.415-0.646] ϕ_5 = 0.694 [0.446-0.715]
7	Ω = 1 [*] τ = 0 [0-0.018] ϕ_1 = 0.288 [0.249-0.333] ϕ_3 = 0.552 [0.466-0.607] ϕ_4 = 0.622 [0.506-0.648] ϕ_5 = 0.694 [0.581-0.718]
8	ϕ = 0 [0-0.014] Ω = 1 [*] τ = 0 [0-0.007]
9	ϕ = 0 [0-0.060] Ω = 1 [*] τ = 0 [0-0.002] ϕ_1 = 0.288 [0.242-0.329] ϕ_3 = 0.552 [0.458-0.615] ϕ_4 = 0.622 [0.477-0.647] ϕ_5 = 0.694 [0.519-0.716]

Table 3-1: Estimated model parameters

Model	Parameters	NLL	AIC	Δ AIC	AW
1	0	13,173.83	26,347.66	22,314.12	0
2	1	13,173.83	26,349.66	22,316.12	0
3	1	2560.25	5112.50	1088.96	0
4	4	2012.77	4033,54	-	0.644
5	2	13,173.83	26,351.66	22,318.12	0
6	5	2012.77	4035,54	2.00	0.237
7	6	2012.77	4037,54	4.00	0.087
8	3	13,173.83	26,353.66	22,320.12	0
9	7	2012.77	4039.54	6.00	0.032

Table 3-2: Model selection. Results of model selection using the Akaike Information Criterion (AIC). Model refers to the model number in *Materials and Methods*, with Model 5 being the model discussed in the results section of the paper. Parameters is the number of model parameters, NLL is the negative log likelihood, AIC is the Akaike Information Criterion, Δ AIC is the difference in AIC between the model in question and the best fitting model (Model 5), and AW is the Akaike Weight, which gives an indication of the probability that a model is best supported. Model 4 is the best-supported model. Although Models 6, 7 and 9 have similar AIC values and an appreciable AW, these models can be discounted since at the estimated parameter values (Table 3-1) the model collapses to Model 4.

Chapter 4

Evolution in case of genetic and functional redundancy in an RNA virus

RNA viruses can be defined by two evolutionary characteristics: high variability and extremely small and compacted genomes. Interestingly, a small genome leads to an apparently low complexity, to a lack of replication enzymes with fidelity mechanisms and so on to a high variability (Smith and Szathmáry, 1995). A clear advantage of a small genome is to replicate quick. Consequently, only few examples of host genetic material integrated into the RNA genome have been observed (Meyers *et al.*, 1989a). When a host gene is inserted into a viral genome, RNA genome size increases and would lead to mutational load increase and viral fitness reduction (Holmes, 2009). Dolja *et al.* (1993) observed large deletions into a foreign gene inserted into the TEV genome. On the contrary, it has been recently observed that NIRVs are integrated into mammalian (Belyi *et al.*, 2010; Geuking *et al.*, 2009; Horie *et al.*, 2010; Katzourakis and Gifford, 2010) and plant (Chiba *et al*, 2011) genomes. A particularly interesting situation is created when hosts carrying a given NIRV are infected by the virus from which the NIRV was taken. How would the virus population evolve under such situation of functional

redundancy? This is the question we want to tackle in this study. To do so, we are evaluating the patterns of genomic evolution of TEV in *N. tabacum* transgenic plants expressing TEV replicase NIb (Li and Carrington, 1995).

During infection of transgenic plants, the NIb copy coded by TEV genome is redundant with the polymerase produced by the transgenic plant and thus should experienced relaxed selection. As a redundant gene, we hypothesize that partial or total deletion of viral NIb gene could occur after several serial passages. In a second evolution experiment, we anticipate this deletion and study how a TEV-Δ*NIb* evolves when complemented *in trans* by the NIb protein expressed by the plants. In both cases, how would such a big genomic perturbation affect the regulation of expression and accumulation of all other viral components?

4a. Comparison between TEV and TEV-ΔNIb accumulation in transgenics plants

In a previous study, Li and Carrington (1995) showed that TEV-Δ*NIb* can be complemented *in trans* by the NIb protein

expressed by the plants and they demonstrated that at 8 dpi, GUS activity in leaves systemically infected by the TEV-Δ*NIb*-GUS mutant was comparable to the activity level of parental TEV-GUS. We have confirmed this result by RT-qPCR (Figure 4-2), and no significant differences were observed in virus accumulation between TEV and TEV-Δ*NIb* in transgenic *Nb*::NIb plants (t_8 = 1.394, P = 0.201). After 30 one-week passages, no significant differences were noticed between the ancestral TEV and the TEV lineages isolated at the end of the evolution phase. On the contrary, three lineages of TEV-Δ*NIb* showed a lower viral accumulation than the ancestral TEV-Δ*NIb* (Mann-Whitney test, P < 0,05) (Figure 4-3). Then, these results suggest that the use *in trans* of NIb produced by the plant could affect virus accumulation of TEV-Δ*NIb* genotypes after several short-time passages.

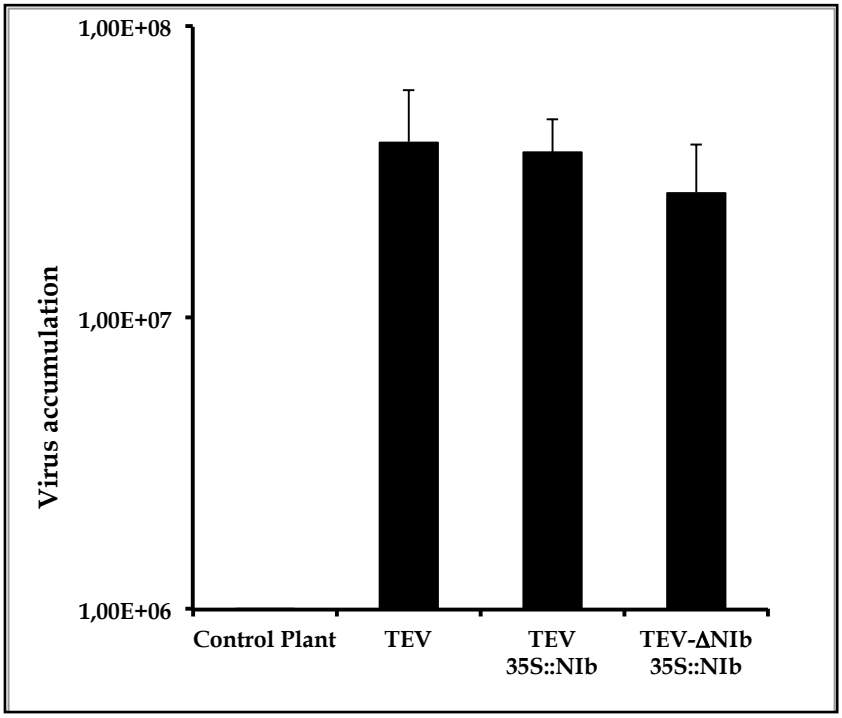


Figure 4-2. Virus accumulation by absolute quantification of TEV *vs* TEV-Δ*NIb* in *N. tabacum*::*NIb* after one passage. No significant difference in the accumulation of TEV and TEV-Δ*NIb* was observed.

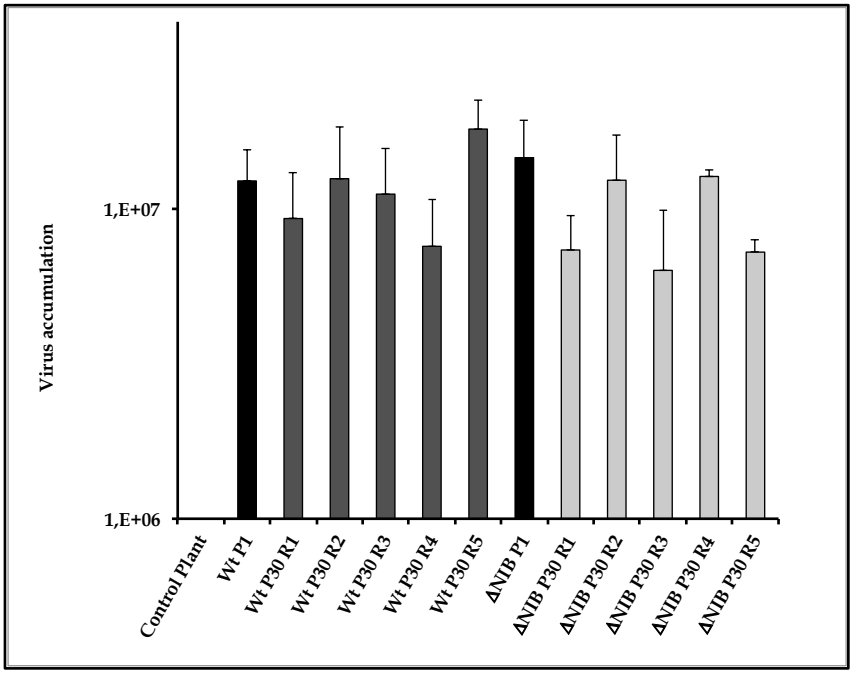


Figure 4-3. Virus accumulation by absolute quantification of TEV *vs* TEV-Δ*NIb* in *N. tabacum*::*NIb* after 30 passages.

4b. Mutational spectrum of TEV and TEV-ΔNIb after 30 one-week-passages

Surprisingly, mutational spectra for both viral genotypes are dominated by poiint mutations and not a single insertion nor deletion was observed (Table 4-4). Thus, the transfer of TEV and TEV-Δ*NIb* from plant to plant at of seven

days intervals resulted in a good stability of each genotype over 30 passages.

No difference was noticed between ratios of transitions to transversions; ratios were 7 for TEV and 6 for *TEV*-Δ*NIb*. Given that it is biochemically easier to produce transitions than transversions, it is not surprising to observe these biased ratios.

The first noticeable difference between both mutational spectra refers to the ratio of synonymous to non-synonymous substitutions. TEV ratio is 11:5, whereas for TEV-Δ*NIb* it is about four times lower (7:11), although the difference between viruses was not significant (Fisher's exact test, $P = 0.100$).

For both genotypes, mutations were spread equally among the genome (Figure 4-4) and only one mutation was shared by both genotypes. Indeed, the mutation A6805G (K2269E) localized in NIa-Pro was observed 4/5 lineages of TEV genotype and in 2/5 lineages of the TEV-Δ*NIb* genotype. This non-synonymous mutation; localized in the extern face of the NIa protein could affect the NIa protease; interfering in a putative interaction with other protein.

Of particular interest are three non-synonymous mutations localized in the 6K1/CI junction (Figure 4-5): G3618U (E1206D), A3622G (I1208V), and C3629U (T1210M), affecting residues P6, P4 and P2 of the proteolytic site, respectively (Adams *et al*., 2005a).

Previous study defined a conserved heptapeptide sequence [E-X-X-Y-X-Q/S (or G)] at the 6K1/CI junction (Adams *et al*., 2005a; Carrington and Dougherty, 1987, 1988; Carrington *et al.,* 1988). As described in Figure 4-5, these three mutations were observed after 30 one-week passages in position P6, P4 and P2. Each mutation was obtained in one lineage but none lineage got all the three mutations together. E1206D conservative mutation at P6 site should affect weakly the proteolytic processing as amino acids E and D are widely represented in *Potyviriade* family at this site (Adams *et al*., 2005a). On the contrary, P4 Mutation I1208V at position P4 should have a stronger impact on processing activity. Indeed, I is represented in position P4 at a low frequency of 7% among all *Potyviridae*, whereas V is extremely common (80%) (Adams *et al*., 2005a). Another mutation that could affect protease

processing is T1210M in P2; being M rare among the *Potyviridae* (2%). Generally, mutation on P2/P4 may involve regulation of the cleavage rate (Dougherty *et al.*, 1989).

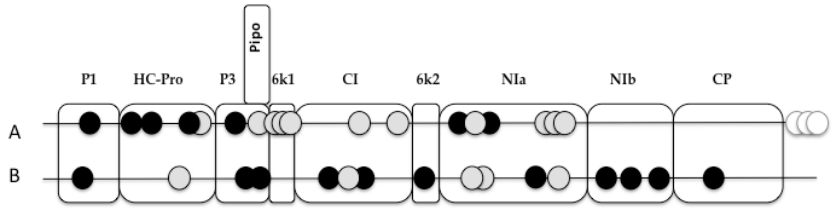


Figure 4-4. Mutations in TEV (B) and TEV-Δ*NI*b (A) evolved genome after 30 weekly passages in *Nt*::*NIb* plants. Black circles represent synonymous mutation, grey circles the non-synonymous mutation and white circle mutations localized in UTR.

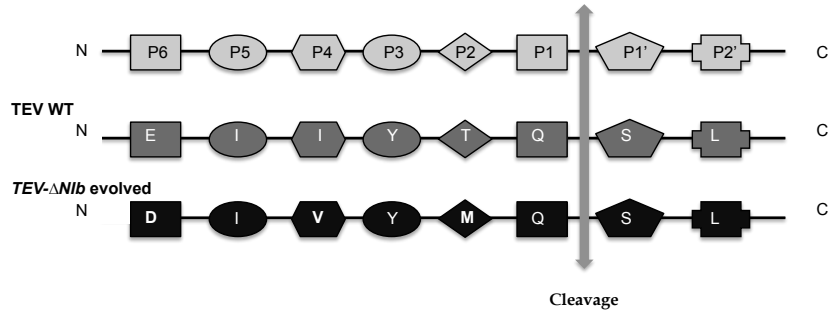


Figure 4-5. TEV-Δ*NIb* mutations on the NIa-Pro proteolytic 6K1/CI site. P6 to P2′ (E-x-x-Y-x-Q-G/S) are key amino acids constituting the NIa-Pro binding site. Cleavage site is between P1 and P1′. P6: E (26.5%) and D (12.8%) are the most frequent

amino acids at this position in the family *Potyviridae*; P4: V is the most frequent 80%; P2: H (40%) and F (15%) are the most represented in *Potyviridae*, M is represented with a frequency of only 2%. Generally, mutations at positions P2/P4 may involve regulation of the cleavage rate (Dougherty *et al*., 1989).

TEV type of mutation		TEV-ΔNIb type of mutation	
Total	16	Total	21
Base substitution	16	Base substitution	21
Transitions	14	Transitions	18
Transversions	2	Transversions	3
Synonymous	11	Synonymous	7
Non-synonymous	5	Non-synonymous	11
UTR	0	UTR	3

	A	U	G	C
A	_	0	2	0
U	2	_	0	2
G	6	0	_	0
C	0	4	0	_

	A	U	G	C
A	_	0	9	0
U	0	_	0	3
G	4	1	_	0
C	1	3	0	_

TABLE 4-4. Mutation spectrum and substitution matrix for TEV and TEV-ΔNIb evolved genome after 30-one-week passages in *Nt::NIb* plants

4c. Analysis of a selection of TEV-ΔNIb mutant after thirty weekly passages

The different mutants K2269E, E1206D, I1208V, and T1210M were generated by site-directed mutagenesis on the infectious plasmid pMTEV. In addition, since K2269E is common among different lineages, we also produced this mutation in combination with the other three. As a proxy to the fitness effect of these mutations, viral accumulation of each simple and double mutant was determined 64 hpi by RTq-PCR in 3 replicates.

Data were fitted to a GLM using "genotype" as factor and "biological replicate" nested within "genotype". A significant effect of the mutant in the observed viral accumulation (χ^2 = 134.648, 7 d.f., P < 0.001) and a Bonferroni sequential *post hoc* test found four different homogenous groups (Figure 4-6), A being the group with the lowest viral accumulation value and D the highest.

Group A was formed by the original TEV-Δ*NIb* and mutant E1206D (affecting the P6 position of 6K1/CI proteolitic site). Interestingly this result is consistent with our previous

hypothesis. Group B showed higher accumulation values and was formed by genotypes carrying mutations at positions P4 and P2 of the 6K1/CI proteolytic site. Both groups C and D carry mutation K2269E at the NIa-Pro. This mutation confers a significantly higher accumulation to the TEV-Δ*NIb* than mutations of the A and B group. Very interestingly, double mutants carrying K2269E and the corresponding mutations affecting the P6 or P4 (group D) showed the highest viral accumulation levels. Thus, we conclude that the NIa-Pro mutation K2269E observed in different evolved lineages of *TEV*-Δ*NIb* must confer a fitness advantage to the TEV-Δ*NIb* when combined with mutations affecting important residues of the proteolytic site.

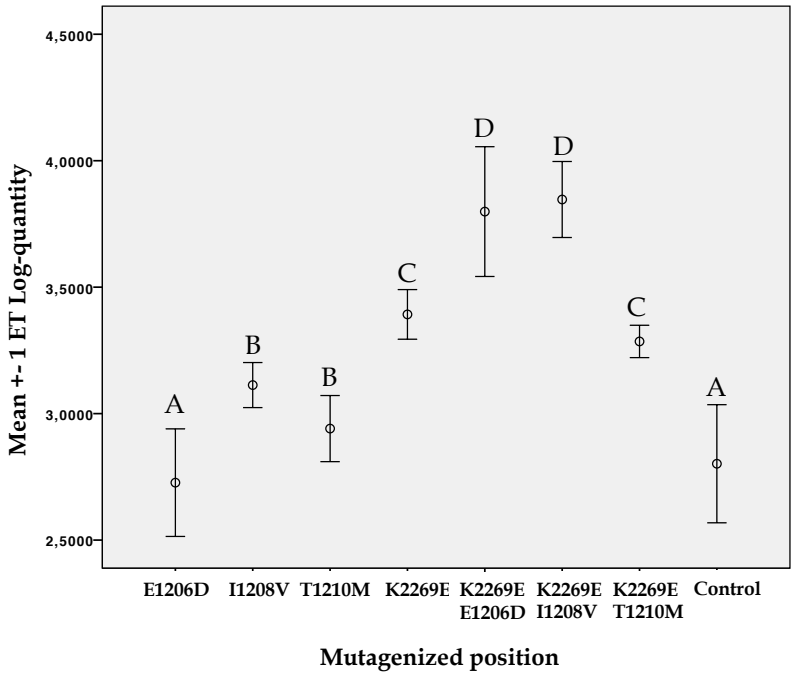


Figure 4-6. Accumulation of mutants E1206D, I1208V, T1210M, K2269E and the combinations K2269E/E1206D, K2269E/I1208V and K2269E/T1210M. Four groups with homogeneous accumulations (groups A to D).

4d. Mutational spectrum of TEV and TEV-ΔNIb after four three-week passages

Evolution experiments with longer passages were performed for both TEV and TEV-Δ*NIb* genotypes. Following Dolja *et al*. (1993), we expect that these longer passages will favor the fixation of more mutations given the increased

chances for competition and selection. A first remarkable observation is that no recombination was observed between TEV-Δ*NIb* and the transgene that restores the genome architecture and gene content of TEV, as previously observed for tombusvirus (Borja *et al.*, 1999). A second remarkable observation is that the first NIb deletion was observed for one TEV lineage after only one passage. Readly, additional deletions were detected by RT-PCR after four passages on other lineages (Figure 4-7). Each of these deletions conserved the reading frame and was situated within the NIb cistron.

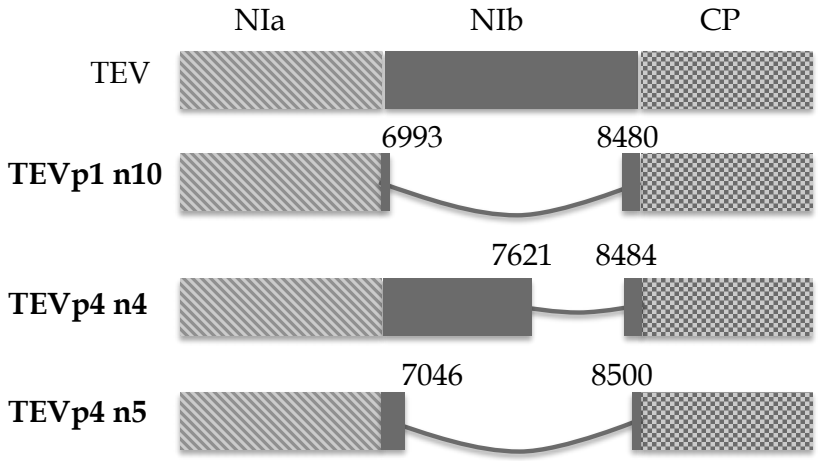


Figure 4-7. TEV and deletion variants detected by RT-PCR in different lineages at passages one and four. NIb (6982-8517) is rapidly deleted after one three-week passage (TEVp1 n10) in

only one lineage. After four three-week passages, deletion variants are observed in different lineages as *e.g*. TEVp4 n4 and TEVp4 n5. All deletions observed after RT-PCR are situated within NIb and retain the reading frame.

4e. High-throughput sequencing of TEV after four three-week passages

In order to assess the stability of TEV genotype and its tendency to loose the NIb gene during passages in transgenic plants *Nt*::NIb, a NGS experiment was performed on the five independent lineages of TEV evolved after four three-week passages. The TEV genome region containing NIb cistron (6377-9473) was amplified and sequenced as described in *Materials and Methods* section. To control the quality of our study, the same region from five *N. tabacum* infected by wild-type TEV was also sequenced by NGS and analysed. In these control plants, no deletions of NIb were observed. The frequency of each deletion variants described in Table 4-5, detected in systemically infected leaves of transgenic plants, is significantly different to our negative control (Fisher's exact test, $P < 0.001$). Our results is also consistent with the previous

deletion variants identified by RT-PCR, indeed we detected the same deletion in lineages four (7621-8484) and five (7046-8500) with the highest frequency. Twenty-two different deletion variants were observed, ranging in frequency from 2.597×10^{-6} (7445-8341; lineage 6) to 6.665×10^{-4} (7621-8484; lineage 4). Interestingly, all deletion variants have respected the reading frame of the coding sequence, meaning that these variant could be viable when *trans*-complemented by the NIb expressed by the plant.

Very surprisingly, in six deletion variants, the deletion extended in the 5′ to include part of the NIa-Pro. Also, in other six variants, the deletion covered from position 6399 in NIa-Pro to position 9282 of CP. These 12 genotypes carrying larger-than-NIb deletions will produce deffective NIa-Pro and/or CP and thus, must act as defective interfering particles. These defective particles must be complemented in co-infected cells by other TEV virus and their spread must be limited, then their frequency in viral population should be lower than ancestral and NIb deleted genotypes. Unexpectedly, one NIa-Pro-to-CP deletion variant (from positions 6438 to 8984) shows a

frequency of 3.13279×10^{-5}, comparable to that of most of NIb-only deletion variants. Finally, as described in Figure 4-8, all lineages except TEV number six are dominated by variants with NIb deletion only, which is consistent with the fact that these variants are viable. Nevertheless, it is surprising to detect so many defective genomes with deletions in NIa-Pro and/or CP, which lead to think that coinfection shall be a common phenomenon during TEV infection. In agreement with this idea, the preservation of defective viral particles by complementation has been widely described (Froissart *et al*., 2004; García-Arriaza *et al*., 2004; Moreno *et al*., 1997; Roux *et al*., 1991).

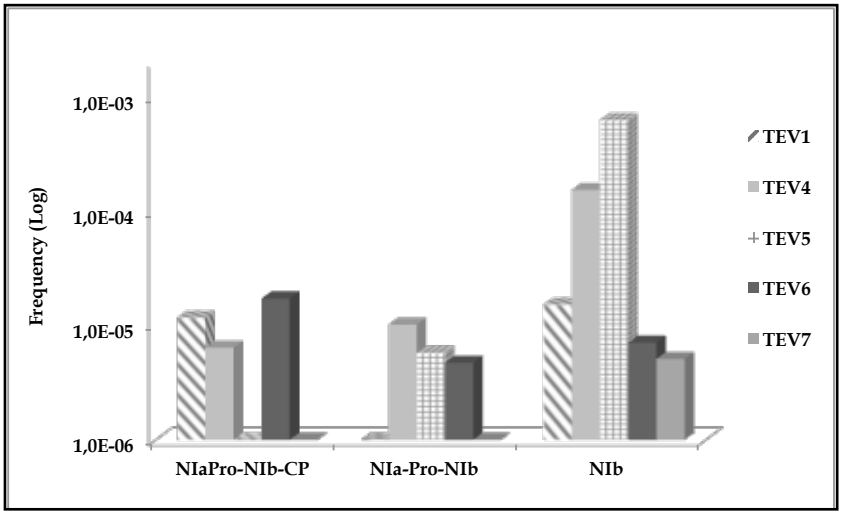


Figure 4-8. Frequency of deletion variants in TEV lineages.

Position	Deletion size	Repeats	Freq	Region	Lineage
6418-9282	2865	16,0	1.63E-05	NIaPro-NIb-CP	TEVn1
6569-8746	2178	12,5	1.27E-05	NIaPro-NIb-CP	TEVn1
6774-8795	2022	6,0	6.11E-06	NIaPro-NIb-CP	TEVn1
7291-8421	1131	15,0	1.53E-05	NIb	TEVn1
6418-8421	2004	8,0	6.39E-06	NIaPro-NIb	TEVn4
6448-8463	2016	17,5	1.40E-05	NIaPro-NIb	TEVn4
6399-9200	2802	8,0	6.39E-06	NIaPro-NIb-CP	TEVn4
7072-8385	1314	6,5	5.19E-06	NIb	TEVn4
7148-8386	1239	7,5	5.99E-06	NIb	TEVn4
7405-8478	1074	54,5	4.83E-05	NIb	TEVn4
7621-8484	864	835,0	6.66E-4	NIb	TEVn4
6457-8385	1929	6,0	5.78E-06	NIaPro-NIb	TEVn5
7046-8500	1455	650,0	6.26E-4	NIb	TEVn5

6460-8478	2019	26,5	8.60E-06	NIaPro-NIb	TEVn6
6488-8455	1968	8,0	2.60E-06	NIaPro-NIb	TEVn6
6590-8374	1785	9,0	2.92E-06	NIaPro-NIb	TEVn6
6429-8702	2274	9,5	3.08E-06	NIaPro-NIb-CP	TEVn6
6438-8984	2547	96,5	3.13E-05	NIaPro-NIb-CP	TEVn6
7155-8483	1329	9,5	3.08E-06	NIb	TEVn6
7221-8480	1260	30,5	1.10E-05	NIb	TEVn6
7445-8341	897	8,0	2.60E-06	NIb	TEVn6
7655-8407	753	7,0	5.09E-06	NIb	TEVn7

Table 4-5. Diversity and frequency of deletions in TEV genotype after four thee-week passages in five different lineages.

DISCUSSION

1. The rate and spectrum of spontaneous mutations in a plant RNA virus

Plant RNA virus mutation rate estimates are within a narrow range of values in the lower side of estimates reported for RNA animal viruses and bacteriophages (Malpica *et al.*, 2002; Tromas and Elena, 2010). This agreement suggests that plant RNA viruses have lower mutation rates than their animal counterparts. Indeed, this difference in mutation rates may help to partially explain why the rates of molecular evolution of most RNA plant viruses are apparently lower than those observed for RNA animal viruses (Fraile *et al.,* 1997; Herránz *et al.,* 2008; Marco and Aranda, 2005; Rodríguez-Cerezo *et al*., 1991). This difference between animal and plant RNA viruses raises an intriguing question: given that plant and animal RNA viruses do not form separated phylogenetic groups and that they are replicated by similar polymerases, why plant RNA viruses show lower mutation rates? We can imagine several scenarios. First, obviously, this may not be the rule and just by chance TEV and TMV turn out to have polymerases of particularly good fidelity. A second obvious possibility is that

most values for animal RNA viruses are upper-limit estimations. In this sense, it has been reported that *Yellow fever virus* polymerase has an error rate as low as 1.9×10^{-7} m/b/g (Pugachev *et al*., 2002). Third, the difference is real and results from differences in the selective pressures that modulated the evolution of mutation rates in both types of hosts; this implying that mutation rate is higher for animal RNA viruses because animals represent a more stressful environment, perhaps in the form of more diverse cell types or stronger antiviral responses (*e.g*., the adaptive immune system; Kamp *et al*., 2002). However, whether virus mutation rates have been optimized by natural selection or are byproducts of a parasitic fast lifestyle needs to be confirmed (Elena and Sanjuán, 2005).

2. Estimation of the *in vivo* recombination rate for a plant RNA virus

In this thesis we have evaluated the point mutation rate for TEV to be in the range 4.754×10^{-6} - 6.299×10^{-5} m/s/g, a range that, given the uncertainties associated to both estimates,

is in the same order of magnitude than the estimate of recombination rate here provided. In any case, this similarity suggests that mutation and recombination may have a similar impact on the generation of genetic diversity for TEV.

Mutation and recombination rates being on the same order of magnitude have important evolutionary implications. For example, high mutation rates in combination with small population sizes during viral transmission turn on Muller's ratchet, a phenomenon already shown to operate in experimental TEV populations transmitted throughout very dramatic bottlenecks (De la Iglesia and Elena, 2007). A large recombination rate would counter act the effect of Muller's ratchet by recreating mutation-free genomes. This example points towards one of the most claimed evolutionary advantages of recombination, as a form of sexual reproduction, namely the purging of deleterious mutations from populations (Muller, 1964; Simon-Loriere and Holmes, 2011). In a similar vein, large recombination rates may have also evolved to speed up the rate of adaptation by bringing together beneficial mutations that otherwise would exist in different genomes,

thus minimizing the effect of clonal interference (Fisher, 1930; Muller, 1932; Simon-Loriere and Holmes, 2011). In addition to these "fitness advantage" theories, other models have been brought forward to explain viral recombination. A particularly intriguing yet appealing of such theories argues that RNA virus recombination may have not evolved because direct fitness benefits but simply as a byproduct of the high processivity of viral RNA polymerases (Simon-Loriere and Holmes, 2011): the faster the replicase moves, the more slippery it is and the more likely to suffer a jump to a different template. If the tradeoff between fast replication and accuracy induced RNA virus high mutation rates as previously suggested (Belshaw *et al.*, 2007; Elena and Sanjuán, 2005); the same correlation could be done for the RNA virus recombination rates. A positive correlation was defined between mutation and recombination rates across RNA virus and could open new perspective to understand the evolution of recombination rates.

3. Estimation of the multiplicity of infection (*MOI*) during RNA virus dynamic infection in a multicellular host

Contrary to what has been described in TMV (González-Jara *et al*., 2009) and CaMV (Gutiérrez *et al*., 2010), low rates of infection and coinfection were observed for TEV. Nevertheless, both infection and coinfection appear to increase over time in the different leaves reaching the 30% of infected cells and 10% of coinfection. Consequently, the predicted *MOI* values, which are extremely dependent of the estimation method, were low, ranging approximately from 1 to 1.65 viruses per cell. Our results should be consistent with previous study describing strong spatial separation pattern in mixed potyvirus infection (Dietrich and Maiss, 2003), limiting then the coinfection to "a small number of cells at the border of different fluorescent cell clusters". However, our models incorporating spatial segregation of genotypes were never selected, and suggest that virus-infected cells are clearly aggregated. Moreover, the rate of mixed-genotype infection is higher than that expected at a given rate of infection for the Poisson null model. In other

words, our results contradict Dietrich and Maiss (2003) conclusions. The main difference between their experiment and our is on the scale of the study, we evaluated infection and coinfection at a cellular level, looking which of ca. 50000 viable protoplasts were non infected, infected or coinfected.

Finally, low coinfection could be also an advantage for viral population, limiting cheating of complemented defective particles (Turner and Chao, 1998) or fitness reduction provocated by intrahost competition between viruses (Froissart *et al*., 2004).

4. Evolution in case of genetic and functional redundancy in an RNA virus

Dolja *et al*. (1993) tested for the first time the stability of TEV genome carrying a foreing gene by using a TEV-GUS engineered genotype. In this genotype, GUS gene is clearly useless for TEV. They studied the stability of the engineered genotype after two different infection periods. TEV-GUS weekly transfers from plant to plant showed a surprising retention of the GUS insert over 21 passages. However, one

passage every four-weeks was enough to observe various large deletions including the sequences coding for most of GUS and of the N-terminal region of HC-Pro. In our experiment, we tested the genomic perturbation that occurs when TEV infects cells that are transgenically producing one of its essential genes, NIb. This study is especially relevant in the context of NIRVs, where no evolution study has been performed yet to understand the impact on the donor virus of having functional redundancy provided by the NIRV. In a first experiment, we made the hypothesis that if NIb is expressed by transgenic plant, then the viral NIb copy could be unnecessary for TEV and then could be deleted, as it was the case with GUS in Dolja *et al*. (1993) experiments. In a second experiment, we anticipate this deletion and study how a TEV-Δ*NIb* evolves when *trans*-complemented by the NIb protein expressed by the plants (*i.e*., NIb acts as a NIRV). Surprisingly, during short periods of infection *i.e.* 30 one-week passages, no multiple mutations were observed in both TEV and TEV-Δ*NIb* genotypes. This result suggests that TEV genotypes showed a high stability and a capacity to evolve using NIb polymerase *in trans*. Furthermore,

single mutations on the 6K1/CI proteolitic site, that should probably affect the NIa-Pro activity, were detected in TEV-Δ*NIb*. TEV is translated into a large polyprotein, thus a mechanism to regulate concentration of effective viral proteins in infected cells may be regulation of cleavage process (Merits *et al*., 2002). Most of these mutations showed a significant higher viral accumulation than the ancestral genotype. In conclusion, these mutations localized in the NIa-Pro substrate may positively regulate the cleavage 6K1/CI in order to produce more mature 6K1 and CI. Proteins NIb and CI are contituents of the viral replication complex (Carrington *et al*., 1998), and thus increasing effective CI concentration may balance a lower production of NIb from the transgene in comparison to what should be produced by the virus in normal conditions. Furthermore, a single mutation in NIa-Pro was interestingly observed on most of TEV and TEV-Δ*NIb* lineages. Mutants carrying this positively selected mutation showed a significantly higher viral accumulation than the ancestral genotypes. This mutation localized on the surface of the NIa-

Pro may facilitate interactions between NIa-Pro and NIb, for example, or simply improve NIa-Pro binding to the substrate.

During long period of infection, *i.e.* four three-week passages, NIb deletions were detected by RT-PCR and confirmed by NGS. We identified 22 different deletion variants, all conserving the reading frame. Twelve additional deletion variants, that must be defective particles, extended the deletion to a large part of 3′ NIa-Pro; six of them also had part of the 5′ extreme of CP deleted. Interestingly, these various defective particles with no negligible frequency were able to colonize the host during infection being complemented with viable genotypes. As described in Montville *et al.* (2005), co-infection is a common viral strategy for RNA virus that is advantageous in the short term but prejudicial in the long term. Indeed, complementation can slow down the rate at which deleterious alleles are eliminated from the virus population (Froissart *et al.*, 2004) and promote selection for defective-interfering particles as cheater genotypes that reduce mean population fitness (Turner and Chao, 1999).

Finally, this experimental system allowed NIb deletion variants to be *trans*-complemented by the plant NIb NIRV and our results support our original hypothesis that the NIRV will relax selection on its homologous viral copy, making it redundant and, therefore, prescindible. Deleting viral NIb gene leads to a genome reduction size of about 15%, which could provide a replicative advantage to the deletion mutants. Recently, a direct competition study between TEV-GFP and TEV was performed (Zwart *et al.*, unpublished data); TEV-GFP genome size is 7% larger that TEV. After only one three-week passage, TEV-GFP completely disappeared in all lineages to the benefit of TEV.

CONCLUSION

Along this thesis we have estimated two essential evolutionary parameters of TEV: the *in vivo* mutation and recombination rates. Interesingly, both rates are in the same range, which reflect the possibility that recombination and mutation are linked by the low fidelity and high processivity of viral RdRps (Simon-Loriere and Holmes, 2011): the faster the replicase moves, the less accuracy it is and the more likely to suffer a jump to a different template. Furthermore, as previous study (Jarvis and Kirkegaard, 1992), our results showed that a correlation exist between RNA concentration and recombination frequency *in vivo*.

Nevertheless, an increase of recombination frequency was not observed between 15 and 45 dpi, suggesting that RNA concentration did not varied or had already reached a plateau. Moreover, recombination mechanisms are strongly dependent of coinfection, and then frequency of recombination should be closely related to *MOI*. Our results show that during host colonization and over time, TEV cellular *MOI* is quite low; so does the frequency of coinfection. These findings suggest that beyond what is minimally necessary (*i.e.*, for each cell to be

infected with at least one of each virus), *MOI* has relatively little effect on recombination frequency, which is consistent with previous study on virus recombination (Kamita *et al*., 2003).

Finally, our result showed a great TEV genome plasticity under conditions of functional redundancy. Indeed, a large number of deletion variants, complemented *in trans* by NIb produced by the plant (NIRV), were observed indicating that non-homologous recombination can also occur during replication. Then, recombination should also act as a mechanism cleaning useless gene, generating quickly strong changes.

Quasispecies theory predicts that the genome length, stably maintained without losing information, is negatively correlated with mutation rate (Eigen, 1971). High processivity of viral RNA polymerase should be coupled with a high mutation rate, generating high genetic diversity, including variation in genome size. Coinfection favouring complementation should allow viral population to retain these size variants in the viral population. These variants composed

of viable and defective particles form a substantial part of the mutant swarm that may be important for evolvability.

REFERENCES

Literature cited

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

Adams, M.J., Antoniw, J.F., and Fauquet, C.M. (2005b). Molecular criteria for genus and species discrimination within the family *Potyviridae*. Arch. Virol. *150*, 459–47.

Agresti, A., and Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. Am. Stat. *52*, 119-126.

Agrios, G.N. (1997). Plant pathology, 4th edition. Academic Press, San Diego, USA.

Ali, A., Li, H., Schneider, W.L., Sherman, D.J., Gray, S., Smith, D., and Roossinck, M.J. (2006). Analysis of genetic bottlenecks during horizontal transmission of *Cucumber mosaic virus*. J. Virol. *80*, 8345–8350.

Anderson, J.A., Bowman, E.H., and Hu, W.S. (1998). Retroviral recombination rates do not increase linearly with marker distance and are limited by the size of the recombining subpopulation. J. Virol. *72*, 1195–1202.

Antia, R., Regoes, R.R., Koella, J.C., and Bergstrom, C.T. (2003). The role of evolution in the emergence of infectious diseases. Nature *426*, 658–661.

Aranda, M.A., Fraile, A., Dopazo, J., Malpica, J.M., and García-Arenal, F. (1997). Contribution of mutation and RNA recombination to the evolution of a plant pathogenic RNA. J. Mol. Evol. *44*, 81–88.

Arezi, B., and Hogrefe, H.H. (2007). Escherichia coli DNA polymerase III epsilon subunit increases *Moloney murine*

leukemia virus reverse transcriptase fidelity and accuracy of RT-PCR procedures. Anal. Biochem. *360*, 84–91.

Bald, J.G. (1937). The use of numbers of infections for comparing the concentration of plant virus suspensions: the effect of carbon on the production of lesions by viruses of the tobacco mosaic group. Annals Appl. Biol. *24*, 77–86.

Baric, R.S., Fu, K., Schaad, M.C., and Stohlman, S.A. (1990). Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. Virology *177*, 646-656.

Batorsky, R., Kearney, M.F., Palmer, S.E., Maldarelli, F., Rouzine, I.M., and Coffin, J.M. (2011). Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. Proc. Natl. Acad. Sci. USA *108*, 5661–5666.

Batschelet, E., Domingo, E., and Weissmann, C. (1976). The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. Gene *1*, 27–32.

Beachy, R.N., and Heinlein, M. (2000). Role of P30 in replication and spread of TMV. Traffic *1*, 540–544.

Beck, D.L., and Dawson, W.O. (1990). Deletion of repeated sequences from *Tobacco mosaic virus* mutants with two coat protein genes. Virology *177*, 462–469.

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

Bedoya, L.C., Martinez, F., Orzaez, D., and Daròs, J.A. (2012). Visual tracking of plant virus infection and movement using a reporter MYB transcription factor that activates anthocyanin biosynthesis. Plant Physiol. *158*, 1130-1138.

Belshaw, R., Gardner, A., Rambaut, A., and Pybus, O.G. (2007). Pacing a small cage: mutation and RNA viruses. Trends Ecol.

Evol. *23*, 188–193.

Belyi, V.A., Levine, A.J., and Skalka, A.M. (2010). Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS Pathog. *6*, e1001030.

Bergmann, M., García-Sastre, A., and Palese, P. (1992). Transfection-mediated recombination of *Influenza A virus*. J. Virol. *66*, 7576–7580.

Bertsch, C., Beuve, M., Dolja, V.V., Wirth, M., Pelsy, F., Herrbach, E., and Lemaire, O. (2009). Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. Biol. Direct *4*, 1-11.

Betancourt, M., Fereres, A., Fraile, A., and García-Arenal, F. (2008). Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. J. Virol. *82*, 12416–12421.

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

Boni, M.F., de Jong, M.D., van Doorn, H.R., and Holmes, E.C. (2010). Guidelines for identifying homologous recombination events in *Influenza A virus*. PLoS ONE *5*, e10434.

Bonnet, J., Fraile, A., Sacristán, S., Malpica, J.M., and García-Arenal, F. (2005). Role of recombination in the evolution of natural populations of *Cucumber mosaic virus*, a tripartite RNA plant virus. Virology. *332*, 359–368.

Borja, M., Rubio, T., Scholthof, H.B., and Jackson, A.O. (1999). Restoration of wild-type virus by double recombination of tombusvirus mutants with a host transgene. Mol. Plant Microbe Interact. *12*, 153–162.

Bracho, M.A., Moya, A., and Barrio, E. (1998). Contribution of Taq polymerase-induced errors to the estimation of RNA virus

diversity. J. Gen. Virol. *79*, 2921–2928.

Brandes, J., and Wetter, C. (1959). Classification of elongated plant viruses on the basis of particle morphology. Virology *8;* 99-115.

Bruyere, A., Wantroba, M., Flasinski, S., Dzianott, A., and Bujarski, J.J. (2000). Frequent homologous recombination events between molecules of one RNA component in a multipartite RNA virus. J. Virol. *74*, 4214–4219.

Bujarski, J.J., and Kaesberg, P. (1986). Genetic recombination between RNA components of a multipartite plant virus. Nature *321*, 528–531.

Bull, J.C., Godfray, H.C., and O'Reilly, D.R. (2001). Persistence of an occlusion-negative recombinant nucleopolyhedrovirus in *Trichoplusia ni* indicates high multiplicity of cellular infection. Appl. Environ. Microbiol. *67*, 5204–5209.

Burch, C.L., and Chao, L. (2004). Epistasis and its relationship to canalization in the RNA virus Phi6. Genetics *167*, 559–567.

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

Caride, E., Brindeiro, R.M., Kallas, E.G., De Sá, C.A.M., Eyer-Silva, W.A., Machado, E., and Tanuri, A. (2002). Sexual transmission of HIV-1 isolate showing G-->A hypermutation. J. Clin. Virol. *23*, 179–189.

Carney, J., Daly, J.M., Nisalak, A., and Solomon, T. (2012). Recombination and positive selection identified in complete genome sequences of Japanese encephalitis virus. Arch. Virol. *157*, 75–83.

Carpenter, C.D., and Simon, A.E. (1996). *In vivo* repair of 3'-end deletions in a TCV satellite RNA may involve two abortive synthesis and priming events. Virology. *226*, 153–160.

Carrasco, P., Daròs, J.A., Agudelo-Romero, P., and 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.

Carrington, J.C. and Dougherty, W.G. (1987). Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. J. Virol. *61*, 2540 –2548.

Carrington, J.C. and Dougherty, W.G. (1988) A viral cleavage site cassette: identification of amino acid sequences required for *Tobacco etch virus* polyprotein processing. Proc. Natl Acad. Sci. USA *85*, 3391– 3395.

Carrington, J.C., Cary, S.M., and Dougherty, W.G. (1988). Mutational analysis of *Tobacco etch virus* polyprotein processing: *cis* and *trans* proteolytic activities of polyproteins containing the 49-kilodalton proteinase. J. Virol. *62*, 2313 –2320.

Carrington, J.C., Cary, S.M., Parks, T.D., and Dougherty, W.G. (1989). A second proteinase encoded by a plant potyviral genome. EMBO J. *8*, 365 –370.

Carrington, J.C., Freed, D.D., and Sanders, T.C. (1989). Autocatalytic processing of the potyvirus helper component proteinase in *Escherichia coli* and *in vitro*. J. Virol. *63*, 4459–4463.

Carrington, J.C., Kasschau, K.D., Mahajan, S.K., and Schaad, M.C. (1996). Cell-to-cell and long-distance transport of viruses in plants. Plant Cell *8*, 1669–1681.

Carrington, J.C., Jensen, P.E., and Schaad, M.C. (1998). Genetic evidence for an essential role for potyvirus CI protein in cell-to-cell movement. Plant J. *14*, 393–400.

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

Chao, L. (1992). Evolution of sex in RNA viruses. Trends Ecol. Evol. *7*, 147–151.

Chao, L., Tran, T., and Matthews, C. (1992). Muller's ratchet

and the advantage of sex in the RNA virus 46. Evolution *46*, 289- 299.

Chao, L., Rang, C.U., and Wong, L.E. (2002). Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage Phi6. J. Virol. *76*, 3276–3281.

Chare, E.R., Gould, E.A., and Holmes, E.C. (2003). Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. J. Gen. Virol. *84*, 2691–2703.

Chare, E.R., and Holmes, E.C. (2005). A phylogenetic survey of recombination frequency in plant RNA viruses. Arch. Virol. *151*, 933–946.

Chen, Y.-K., Goldbach, R., and Prins, M. (2002). Inter- and intramolecular recombinations in the *Cucumber mosaic virus* genome related to adaptation to alstroemeria. J. Virol. *76*, 4119–4124.

Chew, V. (1971). Point estimation of the parameter of the Binomial distribution. Am. Stat. *25*, 47–50.

Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S., and Suzuki, N. (2011). Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. PLoS Pathog. *7*, e1002146.

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

Cline, J., Braman, J.C., and Hogrefe, H.H. (1996). PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. Nucl. Acids Res. *24*, 3546–3551.

Clune, J., Misevic, D., Ofria, C., Lenski, R.E., Elena, S.F., and Sanjuán, R. (2008). Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. PLoS Comput. Biol. *4*, e1000187.

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

Codoñer, F.M., and Elena, S.F. (2008). The promiscuous evolutionary history of the family *Bromoviridae*. J. Gen. Virol. *89*, 1739–1747.

Cronin, S., Verchot, J., Haldeman-Cahill, R., Schaad, M.C., and Carrington, J.C. (1995). Long-distance movement factor: a transport function of the potyvirus helper component proteinase. Plant Cell *7*, 549–559.

Crotty, S., Maag, D., Arnold, J.J., Zhong, W., Lau, J.Y., Hong, Z., Andino, R., and Cameron, C.E. (2000). The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nat. Med. *6*, 1375–1379.

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

Crow, J.F. (1992). The high genomic mutation rate. Curr. Biol. *2*, 605–607.

Cuevas, J.M., González-Candelas, F., Moya, A., and Sanjuán, R. (2009). Effect of ribavirin on the mutation rate and spectrum of *Hepatitis C virus in vivo*. J. Virol. *83*, 5760–5764.

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

Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D., and Baric, R.S. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. *8*, 270–279.

Dietrich, C., and Maiss, E. (2003). Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. J. Gen. Virol. *84*, 2871–2876.

Ding, X.S., Carter, S.A., Deom, C.M., and Nelson, R.S. (1998). Tobamovirus and potyvirus accumulation in minor veins of inoculated leaves from representatives of the Solanaceae and Fabaceae. Plant Physiol. *116*, 125–136.

Ding, X., Shintaku, M.H., Carter, S.A., and Nelson, R.S. (1996). Invasion of minor veins of tobacco leaves inoculated with *Tobacco mosaic virus* mutants defective in phloem-dependent movement. Proc. Natl. Acad. Sci. USA *93*, 11155–11160.

Dolja, V.V., McBride, H.J., and Carrington, J.C. (1992). Tagging of plant potyvirus replication and movement by insertion of beta-glucuronidase into the viral polyprotein. Proc. Natl. Acad. Sci. USA *89*, 10208–10212.

Dolja, V.V., Herndon, K.L., Pirone, T.P., and Carrington, J.C. (1993). Spontaneous mutagenesis of a plant potyvirus genome after insertion of a foreign gene. J. Virol. *67*, 5968–5975.

Dolja, V.V., Haldeman, R., Robertson, N.L., Dougherty, W.G., and Carrington, J.C. (1994). Distinct functions of capsid protein in assembly and movement of *Tobacco etch potyvirus* in plants. EMBO J. *13*, 1482–1491.

Dougherty, W.G., Cary, S.M., and Parks, T.D. (1989). Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. Virology *171*, 356 – 364.

Dougherty, W.G., and Semler, B.L. (1993). Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol. Rev. *57*, 781–822.

Drake, J.W. (1993). Rates of spontaneous mutation among RNA viruses. Proc. Natl. Acad. Sci. USA *90*, 4171–4175.

Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Rates of spontaneous mutation. Genetics *148*, 1667–1686.

Drake, J.W., and Holland, J.J. (1999). Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. USA *96*, 13910–13913.

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

Duggal, R., Cuconati, A., Gromeier, M., and Wimmer, E. (1997). Genetic recombination of poliovirus in a cell-free system. Proc. Natl. Acad. Sci. USA *94*, 13786–13791.

Eigen, M. (1971). Molecular self-organization and the early stages of evolution. Q. Rev. Biophys. *4*, 149–212.

Eigen, M. (1996). On the nature of virus quasispecies. Trends Microbiol. *4*, 216–218.

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

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

Escarmís, C., Dávila, M., Charpentier, N., Bracho, A., Moya, A., and Domingo, E. (1996). Genetic lesions associated with Muller's ratchet in an RNA virus. J. Mol. Biol. *264*, 255–267.

Eshel, I., and Feldman, M.W. (1970). On the evolutionary effect of recombination. Theor. Pop. Biol. *1*, 88–100.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzera, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., Purcell R.H., and Alter H.J. (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science *288*, 339–344.

Fargette, D., Pinel, A., Rakotomalala, M., Sangu, E., Traoré, O., Sérémé, D., Sorho, F., Issaka, S., Hébrard, E., Séré, Y., Kanyeka Z, Konaté G. (2008). *Rice yellow mottle virus*, an RNA plant virus, evolves as rapidly as most RNA animal viruses. J. Virol. *82*, 3584–3589.

Fellers, J., Wan, J., Hong, Y., Collins, G.B., and Hunt, A.G. (1998). *In vitro* interactions between a potyvirus-encoded, genome-linked protein and RNA-dependent RNA polymerase. J. Gen. Virol. *79*, 2043–2049.

Fernández-Cuartero, B., Burgyán, J., Aranda, M.A., Salánki, K., Moriones, E., and García-Arenal, F. (1994). Increase in the relative fitness of a plant virus RNA associated with its recombinant nature. Virology *203*, 373–377.

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

Flegel, T.W. (2009). Hypothesis for heritable, anti-viral immunity in crustaceans and insects. Biol. Direct *4*, 32.

Flint, S.J., Enquist, L.W., Racaniello, V.R., and Skalka, A.M. (2004). Principles of Virology. Molecular biology, pathogenesis, and control of animal viruses. 2nd Edition, ASM Press, Washington, DC, USA.

Folimonova, S.Y. (2012). Superinfection exclusion is an active virus-controlled function that requires a specific viral protein. J. Virol. *86*, 5554–5561.

Fraile, A., Escriu, F., Aranda, M.A., Malpica, J.M., Gibbs, A.J., and García-Arenal, F. (1997). A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. J. Virol. *71*, 8316–8320.

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

French, R., and Stenger, D.C. (2005). Population structure within lineages of *Wheat streak mosaic virus* derived from a common founding event exhibits stochastic variation inconsistent with the deterministic quasispecies model. Virology *343*, 179–189.

Froissart, R., Wilke, C.O., Montville, R., Remold, S.K., Chao, L., and Turner, P.E. (2004). Co-infection weakens selection against epistatic mutations in RNA viruses. Genetics *168*, 9–19.

Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S., and Michalakis, Y. (2005). Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. PLoS Biol. *3*, e89.

Furió, V., Moya, A., and Sanjuán, R. (2005). The cost of replication fidelity in an RNA virus. Proc. Natl. Acad. Sci. USA *102*, 10233–10237.

Furió, V., Moya, A., and Sanjuán, R. (2007). The cost of replication fidelity in *Human immunodeficiency virus* type 1. Proc. R. Soc. B *274*, 225–230.

Gal-On, A., Meiri, E., Raccah, B., and Gaba, V. (1998). Recombination of engineered defective RNA species produces infective potyvirus in planta. J. Virol. *72*, 5268–5270.

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

García-Arriaza, J., Manrubia, S.C., Toja, M., Domingo, E., and Escarmís, C. (2004). Evolutionary transition toward defective RNAs that are infectious by complementation. J. Virol. *78*, 11678–11685.

García-Díaz, M., and Bebenek, K. (2007). Multiple functions of DNA polymerases. CRC Crit. Rev. Plant Sci. *26*, 105–122.

Gerrish, P.J., and García-Lerma, J.G. (2003). Mutation rate and the efficacy of antimicrobial drug treatment. Lancet Infect. Dis *3*, 28–32.

Geuking, P., Narasimamurthy, R., Lemaitre, B., Basler, K., and Leulier, F. (2009). A non-redundant role for *Drosophila* Mkk4 and hemipterous/Mkk7 in TAK1-mediated activation of JNK. PLoS ONE *4*, e7709.

Gibbs, A.J., Ohshima, K., Phillips, M.J., and Gibbs, M.J. (2008). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. PLoS ONE *3*, e2523.

Gibbs, A.J., and Ohshima, K. (2010). Potyviruses and the digital revolution. Annu. Rev. Phytopathol. *48*, 205–223.

Gillespie, T., Boevink, P., Haupt, S., Roberts, A.G., Toth, R., Valentine, T., Chapman, S., and Oparka, K.J. (2002). Functional analysis of a DNA-shuffled movement protein reveals that microtubules are dispensable for the cell-to-cell movement of *Tobacco mosaic virus*. Plant Cell *14*, 1207–1222.

Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., and Taddei, F. (2001). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. Science *291*, 2606–2608.

Godfray, H.C.J., O'Reilly, D.R., and Briggs, C.J. (1997). A model of nucleopolyhedrovirus (NPV) population genetics applied to co-occlusion and the spread of the few polyhedra (FP) phenotype. Proc R. Soc. B *264*, 315–322.

González-Candelas, F., López-Labrador, F.X., and Bracho, M.A. (2011). Recombination in *Hepatitis C virus*. Viruses *3*, 2006–2024.

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

Gray, S.M., and Banerjee, N. (1999). Mechanisms of arthropod transmission of plant and animal viruses. Microbiol. Mol. Biol. Rev. *63*, 128–148.

Greene, A.E., and Allison, R.F. (1994). Recombination between viral RNA and transgenic plant transcripts. Science *263*, 1423–1425.

Gritsun, T.S., and Gould, E.A. (2006). Direct repeats in the 3' untranslated regions of mosquito-borne flaviviruses: Possible

implications for virus transmission. J. Gen. Virol. *87*, 3297–3305.

Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y., and Blanc, S. (2010). Dynamics of the multiplicity of cellular infection in a plant virus. PLoS Pathog. *6*, e1001113.

Hall, J.S., French, R., Hein, G.L., Morris, T.J., and Stenger, D.C. (2001). Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. Virology *282*, 230–236.

Harrison, B.D., Finch, J.T., Gibbs, A.J., Hollings, M., Shepherd, R.J., Valenta, V., and Wetter, C. (1971). Sixteen groups of plant viruses. Virology *45*, 356–363.

Herránz, M.C., Al Rwahnih, M., Sánchez-Navarro, J.A., Elena, S.F., Choueiri, E., Myrta, A., and Pallás, V. (2008). Low genetic variability in the coat and movement proteins of *American plum line pattern virus* isolates from different geographic origins. Arch. Virol. *153*, 367–373.

Hirst, G.K. (1962). Genetic recombination with *Newcastle disease virus*, polioviruses, and influenza. Cold Spring Harbor Symposia on quantitative Biology *27*, 303.

Holmes, E.C. (2009). The evolution and emergence of RNA Viruses. Oxford Series in Ecology and Evolution (OSEE). Oxford University Press, Oxford, UK.

Hong, Y. and Hunt, A. G. (1996). RNA polymerase activity catalyzed by potyvirus encoded RNA-dependent RNA polymerase. Virology *226*, 146-151.

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M., and Tomonaga, K. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature *463*, 84–87.

Hu, W.S., and Temin, H.M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc.

Natl. Acad. Sci. USA *87*, 1556–1560.

Jagadish, M.N., Huang, D., and Ward, C.W. (1993). Site-directed mutagenesis of a potyvirus coat protein and its assembly in *Escherichia coli*. J. Gen. Virol. *74*, 893–896.

Jarvis, T.C., and Kirkegaard, K. (1992). Poliovirus RNA recombination: mechanistic studies in the absence of selection. EMBO J. *11*, 3135–3145.

Jetzt, A.E., Yu, H., Klarmann, G.J., Ron, Y., Preston, B.D., and Dougherty, J.P. (2000). High rate of recombination throughout the *Human immunodeficiency virus* type 1 genome. J. Virol. *74*, 1234–1240.

Johnson, E. M. (1930). Virus diseases of tobacco in Kentucky. Ky. Agric. Exp. Sta. Bull. *306*, 289-415.

Jridi, C., Martin, J.-F., Marie-Jeanne, V., Labonne, G., and Blanc, S. (2006). Distinct viral populations differentiate and evolve independently in a single perennial host plant. J. Virol. *80*, 2349–2357.

Kadaré, G., and Haenni, A.L. (1997). Virus-encoded RNA helicases. J. Virol. *71*, 2583–2590.

Kamita, S. G., Maeda, S., and Hammock, B. D. (2003). High-frequency homologous recombination between baculoviruses involves DNA replication. J. Virol. *77*, 13053-13061

Kamp, C., Wilke, C.O., Adami, C., and Bornholdt, S. (2002). Viral evolution under the pressure of an adaptive immune system: Optimal mutation rates for viral escape. Complexity *8*, 28–33.

Katzourakis, A., and Gifford, R.J. (2010). Endogenous viral elements in animal genomes. PLoS Genet. *6*, e1001191.

King, A.M., McCahon, D., Slade, W.R., and Newman, J.W. (1982). Recombination in RNA. Cell *29*, 921–928.

King, A.M. (1988). Preferred sites of recombination in poliovirus RNA: an analysis of 40 intertypic cross-over sequences. Nucl. Acids Res. *16*, 11705–11723.

Kirkegaard, K., and Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. Cell *47*, 433–443.

Koetzner, C.A., Parker, M.M., Ricard, C.S., Sturman, L.S., and Masters, P.S. (1992). Repair and mutagenesis of the genome of a deletion mutant of the coronavirus *Mouse hepatitis virus* by targeted RNA recombination. J. Virol. *66*, 1841–1848.

Koonin, E.V. (2010). Taming of the shrewd: novel eukaryotic genes from RNA viruses. BMC Biol. *8*, 2.

Kosambi, D.D. (1944). The estimation of map distance from recombination values. Ann. Eugen. *12*, 172-175.

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

Krakauer, D.C., and Plotkin, J.B. (2002). Redundancy, antiredundancy, and the robustness of genomes. Proc. Natl. Acad. Sci. USA *99*, 1405–1409.

Lafforgue, G., Tromas, N., Elena, S.F., and Zwart, M.P. (2012). Dynamics of potyvirus systemic infection establishment: independent yet cumulative action of primary infection sites. J. Virol. *86*, 12912-12922

Lahr, D. J. G., and Katz, L. A. (2009). Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. BioTechniques *47*, 857-866.

Lai, M.M., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L., and Stohlman, S.A. (1985). Recombination between nonsegmented RNA genomes of murine coronaviruses. J. Virol. *56*, 449–456.

Lai, M.M. (1992). RNA recombination in animal and plant viruses. Microbiol. Rev. *56*, 61–79.

Lai, M.M. (1995). Transcription, replication,recombination, and engineering of coronavirus genes. Adv. Exp. Med. Biol. *380*, 463–471.

Laín, S., Riechmann, J.L., and García, J.A. (1990). RNA helicase: a novel activity associated with a protein encoded by a positive strand RNA virus. Nucl. Acids Res. *18*, 7003–7006.

Laín, S., Martín, M.T., Riechmann, J.L., and García, J.A. (1991). Novel catalytic activity associated with positive-strand RNA virus infection: nucleic acid-stimulated ATPase activity of the plum pox potyvirus helicaselike protein. J. Virol. *65*, 1–6.

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

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

Leisner, S.M., Turgeon, R., and Howell, S.H. (1993). Effects of host plant development and genetic determinants on the long-distance movement of *Cauliflower mosaic virus* in *Arabidopsis*. Plant Cell *5*, 191–202.

Levy, D.N., Aldrovandi, G.M., Kutsch, O., and Shaw, G.M. (2004). Dynamics of HIV-1 recombination in its natural target cells. Proc. Natl. Acad. Sci. USA *101*, 4204–4209.

Li, H., and Roossinck, M.J. (2004). Genetic bottlenecks reduce population variation in an experimental RNA virus population. J. Virol. *78*, 10582–10587.

Li, X.H., and Carrington, J.C. (1995). Complementation of *Tobacco etch potyvirus* mutants by active RNA polymerase expressed in transgenic cells. Proc. Natl. Acad. Sci. USA *92*, 457–461.

Li, X.H., Valdez, P., Olvera, R.E., and Carrington, J.C. (1997). Functions of the *Tobacco etch virus* RNA polymerase (NIb): subcellular transport and protein-protein interaction with VPg/proteinase (NIa). J. Virol. *71*, 1598–1607.

Loeb, L.A., Essigmann, J.M., Kazazi, F., Zhang, J., Rose, K.D., and Mullins, J.I. (1999). Lethal mutagenesis of HIV with mutagenic nucleoside analogs. Proc. Natl. Acad. Sci. USA *96*, 1492–1497.

López-Moya, J.J., and Pirone, T.P. (1998). Charge changes near the N terminus of the coat protein of two potyviruses affect virus movement. J. Gen. Virol. *79*, 161–165.

Loverdo, C., Park, M., Schreiber, S.J., and Lloyd-Smith, J.O. (2012). Influence of viral replication mechanisms on within-host evolutionary dynamics. Evolution *66*, 3462–3471.

Luria, S.E., and Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. Genetics *28*, 491–511.

Maia, I.G., and Bernardi, F. (1996). Nucleic acid-binding properties of a bacterially expressed *Potato virus Y* helper component-proteinase. J. Gen. Virol. *77*, 869–877.

Malpica, J.M., Fraile, A., Moreno, I., Obies, C.I., Drake, J.W., and García-Arenal, F. (2002). The rate and character of spontaneous mutation in an RNA virus. Genetics *162*, 1505–1511.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99–103.

Marco, C.F., and Aranda, M.A. (2005). Genetic diversity of a natural population of *Cucurbit yellow stunting disorder virus*. J. Gen. Virol. *86*, 815–822.

Martin, D.P., Van der Walt, E., Posada, D., and Rybicki, E.P. (2005). The evolutionary value of recombination is constrained

by genome modularity. PLoS Genet. *1*, e51.

Martín, S., Sambade, A., Rubio, L., Vives, M.C., Moya, P., Guerri, J., Elena, S.F., and Moreno, P. (2009). Contribution of recombination and selection to molecular evolution of *Citrus tristeza virus*. J. Gen. Virol. *90*, 1527–1538.

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

Merits, A., Rajamaki, M.L., Lindholm, P., Runeberg-Roos, P.,Kekarainen, T., Puustinen, P., Makelainen, K., Valkonen, J.P.T. and Saarma, M. (2002). Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. J. Gen. Virol. *83*, 1211–1221.

Meyerhans, A., Vartanian, J. P. and Wain-Hobson, S. (1990). DNA recombination during PCR. Nucl. Acids Res. *18*, 1687-1691.

Meyers, G., Rümenapf, T., and Thiel, H.J. (1989a). Molecular cloning and nucleotide sequence of the genome of *Hog cholera virus*. Virology *171*, 555–567.

Meyers, G., Rümenapf, T., and Thiel, H.J. (1989b). Ubiquitin in a togavirus. Nature *341*, 491.

Miyashita, S., and Kishino, H. (2010). Estimation of the size of genetic bottlenecks in cell-to-cell movement of *Soil-borne wheat mosaic virus* and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. J. Virol. *84*, 1828–1837.

Montville, R., Froissart, R., Remold, S.K., Tenaillon, O., and Turner, P.E. (2005). Evolution of mutational robustness in an RNA virus. PLoS Biol. *3*, e381.

Moreno, I.M., Malpica, J.M., Rodríguez-Cerezo, E., and García-Arenal, F. (1997). A mutation in *Tomato aspermy cucumovirus*

that abolishes cell-to-cell movement is maintained to high levels in the viral population by complementation. J. Virol. *71*, 9157–9162.

Morra, M.R., and Petty, I.T. (2000). Tissue specificity of geminivirus infection is genetically determined. Plant Cell *12*, 2259–2270.

Moury, B., Fabre, F., and Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. Proc. Natl. Acad. Sci. USA *104*, 17891–17896.

Muller, H.J. (1932). Some genetic aspects of sex. Am. Nat. *66*, 118-138.

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

Munishkin, A.V., Voronin, L.A., and Chetverin, A.B. (1988). An *in vivo* recombinant RNA capable of autocatalytic synthesis by Qβ replicase. Nature *333*, 473–475.

Nagai M., Sakoda Y., Mori M., Hayashi M., Kida H., and Akashi H. (2003). Insertion of cellular sequence and RNA recombination in the structural protein coding region of cytopathogenic *Bovine viral diarrhea virus*. J. Gen. Virol. *84*, 447-452

Nagy, P.D., and Bujarski, J.J. (1993). Targeting the site of RNA-RNA recombination in brome mosaic virus with antisense sequences. Proc. Natl. Acad. Sci. USA *90*, 6390–6394.

Nagy, P.D., Dzianott, A., Ahlquist, P., and Bujarski, J.J. (1995). Mutations in the helicase-like domain of protein 1a alter the sites of RNA-RNA recombination in *Brome mosaic virus*. J. Virol. *69*, 2547–2556.

Nagy, P.D., and Bujarski, J.J. (1995). Efficient system of homologous RNA recombination in *Brome mosaic virus*: sequence and structure requirements and accuracy of crossovers. J. Virol. *69*, 131–140.

Nagy, P.D., and Bujarski, J.J. (1996). Homologous RNA recombination in *Brome mosaic virus*: AU-rich sequences decrease the accuracy of crossovers. J. Virol. *70*, 415–426.

Nagy, P.D., and Simon, A.E. (1997). New insights into the mechanisms of RNA recombination. Virology *235*, 1–9.

Nagy, P.D., Zhang, C., and Simon, A.E. (1998). Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase. EMBO J. *17*, 2392–2403.

Nagy, P.D., Pogany, J., and Simon, A.E. (1999). RNA elements required for RNA recombination function as replication enhancers *in vitro* and *in vivo* in a plus-strand RNA virus. EMBO J. *18*, 5653–5665.

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

Neher, R.A., and Leitner, T. (2010). Recombination rate and selection strength in HIV intra-patient evolution. PLoS Comput. Biol. *6*, e1000660.

Nobusawa, E., and Sato, K. (2006). Comparison of the mutation rates of human influenza A and B viruses. J. Virol. *80*, 3675–3678.

Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F., and Hance, A.J. (2007). Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J. Virol. *81*, 7620–7628.

Ohshima, K., Tomitaka, Y., Wood, J.T., Minematsu, Y., Kajiyama, H., Tomimura, K., and Gibbs, A.J. (2007). Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. J. Gen. Virol. *88*, 298–315.

Olkin, I., Gleser, L.J. and Derman, C. (1994). Probability Models and Applications, 2nd edition. Macmillan Press, New York.

Olkkonen, V.M., and Bamford, D.H. (1989). Quantitation of the adsorption and penetration stages of bacteriophage Phi6 infection. Virology *171*, 229–238.

Olsthoorn, R.C.L., Bruyere, A., Dzianott, A., and Bujarski, J.J. (2002). RNA recombination in *Brome mosaic virus*: effects of strand-specific stem-loop inserts. J. Virol. *76*, 12654–12662.

Onodera, S., Qiao, X., Gottlieb, P., Strassman, J., Frilander, M., and Mindich, L. (1993). RNA structure and heterologous recombination in the double-stranded RNA bacteriophage Phi6. J. Virol. *67*, 4914–4922.

Otto, S.P., and Lenormand, T. (2002). Resolving the paradox of sex and recombination. Nat. Rev. Genet. *3*, 252–261.

Pamilo, P., Nei, M., and Li, W.H. (1987). Accumulation of mutations in sexual and asexual populations. Genet. Res. *49*, 135–146.

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

Pfeiffer, J.K., and Kirkegaard, K. (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. *1*, e11.

Pita, J. S., and Roossinck, M.J. (2013). Fixation of emerging interviral recombinants in *Cucumber mosaic virus* populations. J. Virol. *87*, 1264-1269.

Pugachev, K.V., Guirakhoo, F., Ocran, S.W., Mitchell, F., Parsons, M., Penal, C., Girakhoo, S., Pougatcheva, S.O., Arroyo, J., Trent, D.W., and Monath, T.P. (2004). High fidelity of *Yellow fever virus* RNA polymerase. J. Virol. *78*, 1032–1038.

Raju, R., Subramaniam, S.V., and Hajjou, M. (1995). Genesis of *Sindbis virus* by *in vivo* recombination of nonreplicative RNA precursors. J. Virol. *69*, 7391–7401.

Rao, K.S., and Prakash, A.H. (1995). A simple method for the isolation of plant protoplasts. J. Biosci. *20*, 645–655.

Reiter, J., Pérez-Vilaró, G., Scheller, N., Mina, L.B., Díez, J., and Meyerhans, A. (2011). *Hepatitis C virus* RNA recombination in cell culture. J. Hepatol. *55*, 777–783.

Restrepo-Hartwig, M.A., and Carrington, J.C. (1994). The *Tobacco etch potyvirus* 6-kilodalton protein is membrane associated and involved in viral replication. J. Virol. *68*, 2388–2397.

Revers, F., Le Gall, O., Candresse, T., Le Romancer, M., and Dunez, J. (1996). Frequent occurrence of recombinant potyvirus isolates. J. Gen. Virol. *77*, 1953–1965.

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

Riechmann, J.L., Cervera, M.T., and García, J.A. (1995). Processing of the *Plum pox virus* polyprotein at the P3-6K1 junction is not required for virus viability. J. Gen. Virol. *76*, 951–956.

Rodríguez-Cerezo, E., Elena, S.F., Moya, A., and García-Arenal, F. (1991). High genetic stability in natural populations of the plant RNA virus *Tobacco mild green mosaic virus*. J. Mol. Evol. *32*, 328–332.

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

Roossinck, M.J. (2008) Mutant clouds and bottleneck events in plant virus evolution. In *Origin and Evolution of Viruses* 2nd edition. Elsevier Academic Press, Amsterdam.

Roux, L, Simon, A.E., and Holland, J.J. (1991). Effects of defective interfering viruses on virus replication and pathogenesis *in vitro* and *in vivo*. Adv. Virus Res. *40*, 181–211.

Rubio, T., Borja, M., Scholthof, H.B., Feldstein, P.A., Morris, T.J., and Jackson, A.O. (1999). Broad-spectrum protection against tombusviruses elicited by defective interfering RNAs in transgenic plants. J. Virol. *73*, 5070–5078.

Sacristán, S., Malpica, J.M., Fraile, A., and García-Arenal, F. (2003). Estimation of population bottlenecks during systemic movement of *Tobacco mosaic virus* in tobacco plants. J. Virol. *77*, 9906–9911.

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

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

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

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

Sanjuán, R., Agudelo-Romero, P., and Elena, S.F. (2009). Upper-limit mutation rate estimation for a plant RNA virus. Biol. Lett. *5*, 394–396.

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

Sankara Rao, K., and Prakash, A. H. (1995). A simple method for the isolation of plant protoplasts. J. Biosci. *20*, 645-655

Santa Cruz, S., Roberts, A.G., Prior, D.A.M., Chapman, S., and Oparka, K.J. (1998). Cell-to-cell and phloem-mediated transport of *Potato virus X*: the role of virions. Plant Cell *10*, 495–510.

Sardanyés, J., Solé, R.V., and Elena, S.F. (2009). Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. J. Virol. *83*, 12579–12589.

Schaad, M.C., Jensen, P.E., and Carrington, J.C. (1997a). Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. EMBO J. *16*, 4049–4059.

Schaad, M.C., Lellis, A.D., and Carrington, J.C. (1997b). VPg of *Tobacco etch potyvirus* is a host genotype-specific determinant for long-distance movement. J. Virol. *71*, 8624–8631.

Shackelton, L.A., and Holmes, E.C. (2004). The evolution of large DNA viruses: combining genomic information of viruses and their hosts. Trends Microbiol. *12*, 458–465.

Shapiro, B., Rambaut, A., Pybus, O.G., and Holmes, E.C. (2006). A phylogenetic method for detecting positive epistasis in gene sequences and its application to RNA virus evolution. Mol. Biol. Evol. *23*, 1724–1730.

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

Shukla, D.D., Ward, C.W., Brunt, A.A., and Berger, P.H. (1998). *Potyviridae* family. AAB Descriptions of Plant Viruses. *366*.

Shun Ding, X., Carter, S.A., Michael Deom, C., and Nelson, R.S. (1998). *Tobamovirus* and *Potyvirus* accumulation in minor veins of inoculated leaves from representatives of the *Solanaceae* and *Fabaceae*. Plant Physiol. *116*, 125–136.

Sierra, S., Dávila, M., Lowenstein, P.R., and Domingo, E. (2000).

Response of *Foot-and-mouth disease virus* to increased mutagenesis: influence of viral load and fitness in loss of infectivity. J. Virol. *74*, 8316–8323.

Silva, M.S., Wellink, J., Goldbach, R.W., and Van Lent, J.W.M. (2002). Phloem loading and unloading of *Cowpea mosaic virus* in *Vigna unguiculata*. J. Gen. Virol. *83*, 1493–1504.

Simon-Loriere, E., and Holmes, E.C. (2011). Why do RNA viruses recombine? Nat. Rev. Microbiol. *9*, 617–626.

Steinhauer, D.A., Domingo, E., and Holland, J.J. (1992). Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene *122*, 281–288.

Swetina, J., and Schuster, P. (1982). Self-replication with errors. A model for polynucleotide replication. Biophys. Chem. *16*, 329–345.

Smith, J.M., and Szathmáry, E. (1995). The major evolutionary transitions. Nature *374*, 227–232.

Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M., and Bujarski, J.J. (2011). RNA-RNA recombination in plant virus replication and evolution. Annu. Rev. Phytopathol. *49*, 415–443.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. *24*, 1596–1599.

Taylor, D.J., and Bruenn, J. (2009). The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biol. *7*, 88.

Taylor, D.R., Zeyl, C., and Cooke, E. (2002). Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *99*, 3690–3694.

Thébaud, G., Chadoeuf, J., Morelli, M.J., McCauley, J.W., and Haydon, D.T. (2010). The relationship between mutation frequency and replication strategy in positive-sense single-

stranded RNA viruses. Proc. R. Soc. B *277*, 809–817.

Tromas, N., and Elena, S.F. (2010). The rate and spectrum of spontaneous mutations in a plant RNA virus. Genetics *185*, 983–989.

Turner, P.E., and Chao, L. (1999). Prisoner's dilemma in an RNA virus. Nature *398*, 441–443.

Turner, P.E., Burch, C.L., Hanley, K.A., and Chao, L. (1999). Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage Phi6. J. Virol. *73*, 2420–2424.

Turner, P.E., and Chao, L. (2003). Escape from Prisoner's Dilemma in RNA phage Phi6. Am. Nat. *161*, 497–505.

UNAIDS (2010). Global report fact sheet. www.unaids.org/documents/20101123_FS_Global_em_en.pdf

Urbanowicz, A., Alejska, M., Formanowicz, P., Blazewicz, J., Figlerowicz, M., and Bujarski, J.J. (2005). Homologous crossovers among molecules of *Brome mosaic bromovirus* RNA1 or RNA2 segments *in vivo*. J. Virol. *79*, 5732–5742.

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

Van Nimwegen, E., Crutchfield, J.P., and Huynen, M. (1999). Neutral evolution of mutational robustness. Proc. Natl. Acad. Sci. USA *96*, 9716–9720.

Van der Walt, E., Rybicki, E.P., Varsani, A., Polston, J.E., Billharz, R., Donaldson, L., Monjane, A.L., and Martin, D.P. (2009). Rapid host adaptation by extensive recombination. J. Gen. Virol. *90*, 734–746.

Van der Werf, W., Hemerik, L., Vlak, J.M., and Zwart, M.P. (2011). Heterogeneous host susceptibility enhances prevalence of mixed-genotype micro-parasite infections. PLoS Comput. Biol. *7*, e1002097.

Verchot, J., Koonin, E.V., and Carrington, J.C. (1991). The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. Virology *185*, 527–535.

Verchot, J., Herndon, K.L., and Carrington, J.C. (1992). Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: identification of essential residues and requirements for autoproteolysis. Virology *190*, 298–306.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature, *439*, 344–348.

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

Ward, C.W., Weiller, G.F., Shukla, D.D., and Gibbs, A. (1995). Molecular systematics of the *Potyviridae*, the largest plant virus family. In Molecular Basis of Virus Evolution. Cambridge University Press, Cambridge.

Wei, T., Zhang, C., Hong, J., Xiong, R., Kasschau, K.D., Zhou, X., Carrington, J.C., and Wang, A. (2010). Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. PLoS Pathog. *6*, e1000962.

White, K.A., and Morris, T.J. (1994). Recombination between defective tombusvirus RNAs generates functional hybrid genomes. Proc. Natl. Acad. Sci. USA *91*, 3642–3646.

Wilke, C.O. (2001). Selection for fitness versus selection for robustness in RNA secondary structure folding. Evolution *55*, 2412–2420.

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

World Health Organization. (2003). Fact Sheet no.211.Influenza. www.who.int/mediacentre/factsheets/2003/fs211/.

Yusa, K., Kavlick, M.F., Kosalaraksa, P., and Mitsuya, H. (1997). HIV-1 acquires resistance to two classes of antiviral drugs through homologous recombination. Antiviral Res. *36*, 179–189.

Yuste, E., Sánchez-Palomino, S., Casado, C., Domingo, E., and López-Galíndez, C. (1999). Drastic fitness loss in *Human immunodeficiency virus* type 1 upon serial bottleneck events. J. Virol. *73*, 2745–2751.

Zhuang, J., Mukherjee, S., Ron, Y., and Dougherty, J.P. (2006). High rate of genetic recombination in *Murine leukemia virus*: implications for influencing proviral ploidy. J. Virol. *80*, 6706–6711.

Zwart, M.P., Erro, E., van Oers, M.M., de Visser, J.A.G.M., and Vlak, J.M. (2008). Low multiplicity of infection in vivo results in purifying selection against baculovirus deletion mutants. J. Gen. Virol. *89*, 1220–1224.

Zwart, M.P., Hemerik, L., Cory, J.S., de Visser, J.A.G.M., Bianchi, F.J.J.A., Van Oers, M.M., Vlak, J.M., Hoekstra, R.F., and Van der Werf, W. (2009). An experimental test of the independent action hypothesis in virus-insect pathosystems. Proc. Biol. Sci. *276*, 2233–2242.

Zwart, M.P., Daròs, J.A., and Elena, S.F. (2011). One is enough: in vivo effective population size is dose-dependent for a plant RNA virus. PLoS Pathog. *7*, e1002122.

Zwart, M.P., Daròs, J.A., and Elena, S.F. (2012). Effects of potyvirus effective population size in inoculated leaves on viral accumulation and the onset of symptoms. J. Virol. *86*, 9737–9747.