Facultad de Ciencias Biológicas

Instituto Cavanilles de Biodiversidad y Biología Evolutiva

Variability and host-dependency of RNA virus mutation rates

Marine Combe

2015

Programa Oficial de Postgrado en Biodiversidad

Director: Dr. Rafael Sanjuán Verdeguer

PhD thesis presented by Marine Combe to get the degree of Doctor of Philosophy from the University of Valencia.

Marine Combe.

This work has been supervised by Doctor Rafael Sanjuán Verdeguer.

Rafael Sanjuán Verdeguer.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Parce qu'une thèse est un apprentissage et un acheminement professionnel mais aussi un accomplissement personnel, je ne pouvais commencer sans remercier toutes les personnes qui m'ont soutenue tout au long de ces années.

Un grand merci à mes parents, ma maman Nadine, mon papa Patrick, ma seconde maman Isabelle, mes grand parents, mon frère Loïc, mes sœurs Julie et Léa, ainsi qu'à Clémounet et Maëlou, qui m'ont épaulée durant toutes ces années, quels que soient mes choix. Merci pour votre soutien, vos visites à Valencia, et votre présence lors de mes brefs séjours en France. Malgré la distance et mon absence, vous êtes et serez toujours à mes côtés!

Je tenais à remercier tout particulièrement mes amis de la vie, Passette, Lamberte, Le bras et Alex, qui m'ont soutenue lors de mon départ en Espagne, et ont su m'écouter dans les bons comme dans les mauvais moments. A Nîmes, à Lattes, à Paris ou à Valencia, merci d'avoir répondu présent et un grand merci pour nos moments de délire et de folie!

L'expérience Valencienne a été une découverte personnelle et m'a permis de faire de très belles rencontres. Coco et Nico, vous avez été bien plus que des amis, une vraie famille! Merci pour votre accueil, votre présence, votre soutien, votre

écoute et vos conseils. Julie et Lara, merci d'avoir pris le relais. Grâce à vous j'ai découvert Zaragoza, Bilbao, les Pyrénées espagnols, sans oublier Las Fallas, une seule fois mais à fond! Elena, Dadou, Aroa et Amaya, merci pour m'avoir intégrée dans votre groupe. Las chicas, Kelly, Gwen, Anne-So, Elo, et les Cocos, quel plaisir de vous retrouver pour des soirées inoubliables et des nuits de folie. Sans vous tous à mes côtés l'aventure Valencienne n'aurait pas eu le même charme!

Mr. Mth Pey, merci de m'avoir soutenue et épaulée jour après jour, et ce malgrè la distance. Toi en Guyane, moi en Espagne et un Atlantique pour nous faciliter la tâche, nous avons malgrè tout parcouru un bout de chemin ensemble, et sans ta présence, ton écoute, tes conseils et ton grain de folie, il aurait été moins agréable. Un grand merci pour ta patience, maintenant commençons!

Je souhaitais également remercier mon directeur de thèse, Rafael Sanjuán Verdeguer, qui m'a offert la possibilité d'intégrer son équipe au sein del Instituto Cavanilles de Biodiversidad y Biologia Evolutiva, et sans qui cette thèse n'aurait pu aboutir. Un grand merci à mes collègues, Raquel, José, Ron, Silvia, Marianoel et Pablo, pour m'avoir enseignée, aidée et conseillée durant ces trois années. Manuel, merci à toi aussi pour ton écoute, ta patience, ta compréhension et tes judicieux conseils. J'espère que tu continueras à m'emmener dans tes périlleux voyages. Merci à toute l'équipe du Cavanilles et à tous ceux qui ont contribué à mon intégration et à mon bien-être au sein de l'institut.

RESUMEN

Los virus de ARN pueden infectar todo tipo de organismos, desde los procariotas a los eucariotas superiores, y estos agentes infecciosos parecen particularmente propensos a causar enfermedades emergentes tanto en humanos, animales, como en plantas. Su habilidad para escapar del sistema inmunitario, evadir estrategias antivirales o infectar a nuevas especies son aspecto más de su rápida evolución. Por lo tanto, comprender los procesos básicos del la evolución de los virus de ARN podría ayudar en el diseño de nuevas estrategias antivirales.

Una de las principales características de los virus de ARN es su tasa de mutación extremadamente alta. De hecho, la alta diversidad genética de las poblaciones virales da lugar a una nube de variantes que interactúan y contribuyen colectivamente, conocido también por el término de cuasiespecies, y que permite a las poblaciones virales adaptarse rápidamente a entornos dinámicos. Estudios anteriores han demostrado que la tasa de mutaciones espontáneas de los virus de ARN varía de 10^{-6} a 10^{-4} sustituciones por nucleótido por célula infectada (s/n/c) y puede variar considerablemente, incluso para el mismo virus, aunque se sabe poco sobre las causas de esta variabilidad. Consecuentemente, el conocimiento de la tasa de mutación y el espectro molecular de mutaciones espontáneas son importantes para entender la evolución de la composición genética de las poblaciones virales.

En los virus de ARN las tasas de mutación están determinadas por factores codificados por el virus, como la fidelidad de la polimerasa viral, la presencia/ausencia de mecanismos correctores, o el modo de replicación viral. Sin embargo, sólo unos pocos estudios se han centrado en las características celulares que podrían explicar la variabilidad en la producción de la diversidad genética viral, o si esta variabilidad pudiera ser distribuida heterogéneamente entre células únicas.

En este estudio, se propuso analizar algunos de los factores subyacentes a la variabilidad en las tasas de mutación de los virus de ARN. En primer lugar, se estudió el efecto del tipo de célula que se infecta, y la variación en función del huésped en el que el virus se replica. A continuación, nos centramos en la variabilidad que ocurre a nivel de una célula única, mediante la caracterización de la diversidad genética de los virus liberados a partir de células individuales. Por último, nos fijamos en el potencial efecto de factores celulares de tipo ADAR, mediante la determinación del tipo de mutaciones espontáneas que se acumulan en el genoma de un norovirus.

Se utilizó la prueba de fluctuación de Luria-Delbrück para comprobar la variabilidad en la tasa de mutación del virus de la estomatitis vesicular (VSV) entre diferentes células de mamíferos, así como entre diferentes condiciones de cultivo. Se encontró una tasa de mutación similar entre las células BHK-21, así como en células embrionarias de ratón (MEF), células MEF inmortalizadas mediante deleción del gen p53, células de cáncer de colon murino

(CT26) y de neuroblastoma (Neuro-2A), sugiriendo que VSV se replica con la misma fidelidad en estos diferentes tipos de células de mamíferos. Por otra parte, debido a que el ciclo de vida de VSV no sólo implica su replicación en mamíferos, sino también en insectos, comprobamos su tasa de mutación en células de *Drosophila melanogaster* (S2), de *Spodoptera frugiperda* (sf21) y del mosquito *Aedes albopictus* (C6/36). Curiosamente, la tasa de mutación de VSV fue cuatro veces mas baja en células de insectos en comparación con las células de mamíferos probadas. Curiosamente, los arbovirus parecen tener una evolución más lenta que los virus que no estan transmitidos por un vector, y nuestros resultados sugieren que en los insectos esto puede ser en parte debido a una tasa de mutación más baja.

Se desarrolló un enfoque que combina micromanipulación y secuenciación masiva para estudiar la diversidad genética de los virus producidos y liberados por células únicas, usando de nuevo VSV como sistema modelo. Demostramos que el virus presenta gran diversidad genética en células únicas, aunque esta diversidad no fue homogéneamente contribuida por todas las células. Llama la atención que la variabilidad existente en el inóculo viral demostró que la unidad infecciosa mínima (PFU) también puede transmitir diversidad genética viral. En efecto, no sólo se observó complementación genética entre alelos de las variantes albergadas dentro de la misma PFU, sino también que el efecto de una mutación se determinó colectivamente por otras variantes durante su co-transmisión, un proceso que podría seguir durante varias generaciones. Esta co-transmisión de diversidad genética

dentro de la misma unidad infecciosa tiene implicaciones importantes para la evolución viral, como por ejemplo la de mantener la diversidad genética durante fuertes cuellos de botella de las poblaciones virales, o crear una asociación espacial entre los alelos, y sugiere que la selección natural actúa a nivel de conjuntos de partículas virales diversas en lugar de sobre genotipos individuales. Por último, hemos visto que la variabilidad en la diversidad genética producida entre células únicas estaba estrechamente asociada con el rendimiento viral, sugiriendo un compromiso entre la eficacia y la fidelidad de la replicación del virus, que puede ser determinado por un modelo de replicación geométrica. Aunque este tipo de replicación alimenta la aparición de nuevas mutaciones, también aumenta la carga genética. Para equilibrar ambos mecanismos, hemos demostrado que los mutantes que muestran un fenotipo mutador deberían ser encontrados en una baja frecuencia en las poblaciones virales.

Se usó un clon infeccioso del virus Norwalk (NV). Puesto que NV no es capaz de iniciar subsiguientes ciclos de infección en las células aquí utilizadas (HEK293T), tuvimos la posibilidad de observar la aparición de mutaciones espontáneas en un único ciclo, descartando así posibles efectos de la selección o la acumulación de mutaciones a lo largo de varias generaciones. Obtuvimos así una tasa de mutación de 9×10^{-5} s/n/c consistente con las estimaciones publicadas anteriormente para otros virus de ARN. Curiosamente, se observó que de los 128 clones moleculares secuenciados con el Sanger, dos mostraron múltiples cambios T -> C. Sugerimos que estas hipermutaciones podrían ser el resultado

de la edición del RNA viral por parte de enzimas celulares de tipo ADAR. Esta hipótesis fue confirmada por secuenciación masiva, sugiriendo que los factores celulares de tipo ADAR también pueden actuar en la variabilidad de la tasa de mutación de los virus de ARN.

INTRODUCTION

1. Introduction to RNA viruses

By the end of the 19th century, the infectious agent responsible for tobacco mosaic disease, able to go through filters retaining bacteria was discovered by Adolf Mayer and Dimitri Ivanofsky. Soon after this discovery, the existence of entities smaller than bacteria, not observable with a light microscope and replicating only in living cells was introduced by Martinus Beijerinck (Manrubia and Lázaro, 2006). Later in the 20th century, these infectious agents were named viruses, and the first isolation and description of bacteriophages (i.e. viruses that infect bacteria) and animal viruses, such as foot-and-mouth disease virus (FMDV) (Loeffler and Frosch, 1998), initiated the development of the virology field.

Although all cellular organisms use DNA to store genetic information, viruses are composed either of a DNA or an RNA molecule. Whereas DNA viruses replicate their genome with similar enzymes as the host cell, in RNA viruses the genetic information is stored in an RNA molecule. Therefore, to reproduce, RNA viruses have to use specific replication mechanisms, which are characterized by an extremely low fidelity, and are mainly responsible for their high rates of mutation and highly heterogeneous populations. Indeed, a key feature of RNA viruses is their extremely high mutation rate, which allows them to evolve at timescales that can be recorded by human

observation, thus making them an ideal model to study evolution (Holmes, 2008; Holmes, 2009; Biek et al., 2015).

RNA viruses are able to infect a wide range of hosts, from prokaryotes to higher eukaryotes, and are well situated in the list of human infectious diseases. Indeed, RNA viruses include the second and sixth most deadly pathogens worldwide (HIV-1 and measles, respectively) and contribute significantly to the first (lower respiratory infections) and third (diarrhea) biggest causes of mortality (World Health Statistics, www.who.int/gho/publication/world-health-statistics/2013). Consequently, the economic cost of antiviral treatments, vaccination campaigns and epidemiological interventions against RNA viruses are considerable, and direct loss of productivity is an additional concern in livestock breeding and agriculture. For instance, during the 2002-2003 severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak, only 8437 people were infected, 813 died, but the global economic cost was about US$50 billion. Also, the 2001 epidemic of foot-and-mouth disease in the UK (due to FMDV) resulted in the death or slaughter of over 3.5 million cattle and a total cost of approximately $4 billion (Holmes, 2009).

During their lifecycle, RNA viruses infect different tissues within the same host, different hosts of the same species, but also jump between host species, and therefore have to constantly adapt to new environments (Stern et al., 2014). Although the high genetic diversity of RNA virus populations allows them to quickly respond to these changing environments, these viruses also experience several evolutionary constraints. As a consequence,

RNA viruses frequently cause transient “spill-over” infections in new host species, where only a single individual is infected, because they can not sustain host-to-host transmission (Holmes, 2009).

2. Evolutionary mechanisms: mutation, selection, genetic drift and adaptation

Evolution through natural selection necessitates the generation of genetic diversity on which it can act. For riboviruses, this source of diversity is mainly a consequence of the occurrence of mutations during genome replication within their host, such as the substitution, deletion or insertion of a nucleotide into the nascent chain. This mechanism is the basis of the spontaneous mutation rate, defined as the probability of a genetic change per copied nucleotide per viral generation (Combe and Sanjuán, 2014a), and must be distinguished from the mutation frequency that is the actual proportion of mutations in a given population (Drake, 1993; Sanjuán et al., 2010). Since RNA replicases and reverse transcriptases lack proofreading activities, mutation rates in RNA viruses are order of magnitude higher than those of DNA viruses, and are close to 1 mutation per genome per replication event, leading to highly heterogeneous populations (Domingo and Holland, 1997; Drake and Holland, 1999; Drake, 1993; Schrag et al., 1999; Duffy et al., 2008).

High mutation rates can represent a selective advantage because they generate a huge diversity of variants that allow RNA viruses to divert the host immune system, adapt to rapid

environment changes (Giraud et al., 2001) and confront vaccine and antiviral drugs (Gerrish and García-Lerma, 2003). However, the majority of spontaneous mutations are deleterious, as demonstrated with VSV, where up to 40% of spontaneous mutations were lethal whereas 29.2% were deleterious (Sanjuán et al., 2004), and with TEV where 40.9% were lethal and 36.4% were deleterious (Carrasco et al., 2007). As a consequence, high mutation rates are also associated to increased genetic load and fitness loss (Chao, 1990; Duarte et al., 1992; de la Iglesia and Elena, 2007; García-Arenal et al., 2001; Pybus et al., 2007).

The evolutionary mechanisms driving high riboviral mutation rates have also been interpreted in terms of fitness trade-offs. Indeed, it has been proposed that there is a trade-off between replication speed and replication fidelity, such that selection for rapid replication results in higher mutation rates (Elena and Sanjuán, 2005). This trade-off was exemplified by the observation of a fitness cost associated with increased replication fidelity that resulted in lower replication rates for both vesicular stomatitis virus (Furió et al., 2005) and HIV-1 (Furió o et al., 2007). Also, there could be a possible trade-off between the rates of production of advantageous *versus* deleterious mutations. Although natural selection should favor lower mutation rates in stable environments, because of the reduced load of deleterious mutations (Duffy et al., 2008), RNA viruses experience constantly changing environments, and thus high error-prone RNA polymerases represent a selective advantage (Mansky and Cunningham, 2000; Vignuzzi et al., 2006). However, RNA viruses exist close to their maximum tolerable error rates (i.e. error

threshold, Manrubia et al., 2010) and slight elevations in mutation rate lead to their extinction because the fittest genotype is lost (Eigen, 1971; Eigen, 1992). This has been illustrated with studies of lethal mutagenesis, where artificially increased mutation rates using a chemical mutagen, such as 5'-fluorouracil or ribavirin, resulted in important fitness loss but also extinction (Mansky and Bernard, 2000; Sierra et al., 2000; Anderson et al., 2004; Domingo et al., 2005; Bull et al., 2007). Therefore, these observations suggested that RNA virus high mutation rates might also be the results of a trade-off between the production of sufficient beneficial mutations to rapidly adapt to changing environments and the risk of mutational meltdown (Sniegowski et al., 2000; Duffy et al., 2008; Holmes, 2009).

An important consequence of the idea that riboviruses exist at the limit of the error threshold is that it might also have an upper limit in their genome sizes, described as Eigen's paradox (Holmes, 2009). Under the same rate of mutation per nucleotide, RNA viruses with larger genomes will suffer of more deleterious mutations than those with smaller genomes (Eigen, 1992). However, if smaller genomes are selected, they cannot package the additional genetic material needed for proofreading activities (Maynard Smith and Szathmáry, 1995). Counterbalancing the burden of deleterious mutations imposed by their quick replication and the lack of proofreading mechanisms, RNA viruses exist as enormous population sizes. For instance, at the intra-host level, it has been estimated for HIV-1 that up to 10^7-10^8 cells can be infected, with about 10^{10} virions produced every day (Perelson et al., 1996), whereas in the case of HCV, the daily production is

about 10^{12} virions (Neumann et al., 1998). For plant viruses, there may also be up to 10^{12} virus particles produced in a single leaf (Gibbs et al., 2008). Finally, for poliovirus the burst size (i.e. the number of infectious viruses produced per infected cell) could be about 10^{4} (Kew et al., 2005). Moreover, because of huge population sizes, and to compensate for the accumulation of deleterious mutations, natural selection may also favour mutational robustness in RNA virus populations (Krakauer and Plotkin, 2002; Elena et al., 2006; Lenski et al., 2006; Sanjuán et al., 2007), defined as the phenotypic constancy of a viral population in the face of genetic changes.

In addition to selection, the frequency of a given mutation in a viral population can also be influenced by genetic drift. This evolutionary mechanism results from randon changes in allele frequencies in finite populations, which can lead to the stochastic fixation of mutations irrespectively of their selective value (Ohta, 1992). In viral populations, bottleneks frequently take place during transmission from one individual host to another, thus increasing the power of genetic drift. Transmission bottleneks have been estimated in several animal and plant viruses (Wahl et al., 2002; Sacristán et al., 2003; Li and Roossinck, 2004; Betancourt et al., 2008) and, in most cases, they can be substantial, with only one or few viral particles being transmitted.

3. RNA virus mutation rates

The rate of spontaneous mutation is a key parameter determining the genetic structure of populations over time.

However, it was only in the late 1970´s that the pioneer work demonstrating the extremely high genetic diversity of bacteriophage Qβ (Domingo et al., 1978) allowed us to have a better understanding of the causes and implications of viral mutation rates. Afterwards, mutation rates have been measured for over 20 different RNA viruses (Sanjuán, 2010), although quantifying these rates remains technically difficult. For instance, many studies are limited to only a few and potentially unrepresentative mutational targets, and estimates are often biased by the action of purifying selection, which removes deleterious mutations from the population (Sanjuán, 2010). To divert this selection bias, one possibility is to use reporter genes or *in vitro* biochemical assays using recombinant enzymes, but these types of studies do not accurately reflect the sequence context and the cellular environment of viral replication in nature.

Importantly, all viral RNA-dependent RNA polymerases (RdRps), except those of coronaviruses, and all reverse transcriptases (RT) studied to date lack 3´exonuclease proofreading activity and incorporate mismatches into the nascent chain relatively efficiently (Roberts et al., 1988; Minskaia et al., 2006).

RNA virus mutation rates published to date range two orders of magnitude, from 10^{-6} to 10^{-4} substitutions per nucleotide per cell infection (s/n/c) and can vary considerably even for the same virus (Duffy et al., 2008; Sanjuán et al., 2010). However, little is known about the causes of this variability, aside from pure measurement errors. Since slight changes of the viral mutation

rate can have important implications for pathogenesis (Pfeiffer and Kirkegaard, 2005; Vignuzzi et al., 2006), vaccine development (Vignuzzi et al., 2008; Weeks et al., 2012), antiviral therapy (Anderson et al., 2004; Domingo, 2006), and epidemiological disease management (Holmes, 2008; Woolhouse and Gaunt, 2007), it is necessary to have a better idea of the factors underlying the variability in viral mutation rates.

3.1 Virus-encoded factors acting on mutation rates

3.1.1 Mutation rate versus genome size

Among DNA viruses, bacteria, and also unicellular eukaryotes, mutation rates have been shown to correlate inversely with genome size, such that the number of mutations per genome stays roughly constant at 0.003 substitutions per round of copying (Drake, 1991). However, this rule was recently rejected for multicellular eukaryotes, in which mutation rates actually tend to increase with genome size (Lynch, 2010). Because RNA virus genome sizes vary only by tenfold, as opposed to a thousand-fold variation across DNA viruses and microorganisms, the relationship between mutation rate and genome size has remained less clear. However, a negative correlation between RNA virus mutation rates and genome sizes was first suggested by the discovery of a 3´exonuclease proofreading activity in the RdRp of coronaviruses, the RNA viruses with the largest genomes (Denison et al., 2011; Eckerle et al., 2010; Eckerle et al., 2007; Ulferts and Ziebuhr, 2011; Minskaia et al., 2006). Moreover, there is a negative correlation between RNA virus genome sizes and evolution rates (measured substitutions per nucleotide per year in

a phylogeny), which are in turn partially determined by mutation rates (Sanjuán, 2012). Finally, the recent mutation rate estimate for bacteriophage Qβ (1.4×10^{-4} s/n/r), which has one of the smallest genome sizes, was the highest described among riboviruses, thus supporting the hypothesized negative association with genome size (Figure 1) (Bradwell et al., 2013).

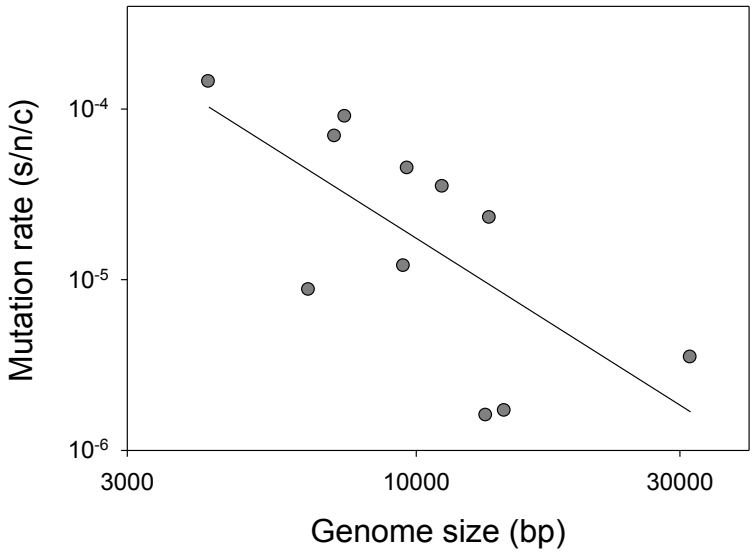


Figure 1. Negative correlation between RNA virus mutation rate and genome size (bp: base-pair). The least-squares regression line has slope -2.06 ± 0.79 in log-log scale (Pearson r = - 0.655, p = 0.029). Reverse transcribing viruses were not included in this plot. Adapted from Combe and Sanjuán (2014a).

The evolutionary explanation for this negative correlation between mutation rates and genome sizes remains unclear. As discussed before, spontaneous mutations in RNA viruses are more

often deleterious than beneficial, and to reduce the genetic load selection should favour lower mutation rates, but on the other hand, higher mutation rates allow adaptation (Sniegowski et al., 2000). To resolve this situation, theory predicts that there should be an optimum value of the mutation rate that maximizes adaptability and correlates inversely with genome size (Orr, 2000; Johnson and Barton, 2002).

3.1.2 Mutation rate versus genome polarity

According to Baltimore's classification, there are the following groups of viruses: positive-stranded, negative-stranded, double stranded (ds), and reverse-transcribing viruses. The first two groups of viruses are single-stranded, and are denoted ss(+)RNA and ss(–)RNA, respectively. In ss(+)RNA and ss(–)RNA viruses, replication occurs by synthesis of complementary strands, which are used as templates for the synthesis of new genomic RNAs. In dsRNA viruses replication can occur either through a positive-strand intermediate used for synthesizing dsRNA (i.e. rotaviruses) or following the standard semi-conservative replication mode of cellular DNA (i.e. partitiviruses).

Genome polarity has many implications in RNA virus biology. In ss(+)RNA viruses, the genomic RNA acts directly as mRNA, whilst in ss(–)RNA and dsRNA viruses it needs to be transcribed first. As a consequence, ss(–)RNA and dsRNA viruses have to package a RdRp in the virion and generally protect their genomic RNA with nucleoproteins, potentially protecting them from RNA damaging agents or host-mediated editing. In contrast, the genome of ss(+)RNA viruses and retroviruses is more naked

and susceptible to form secondary structures, thus increasing the probability of recombination and frame-shift mutation, as showed for HIV-1 (Galetto et al., 2004; Simon-Loriere et al., 2009; Pathak and Temin, 1992).

Recently, we showed that ss(+)RNA viruses exhibit the highest variability in mutation rates, since they ranged by two orders of magnitude (Figure 2). However, this observation could be due to the fact that there are (i) more studies for ss(+)RNA viruses compared to ss(–)RNA and dsRNA viruses, (ii) estimates for a broader host range in ss(+)RNA viruses (animals, plants and bacteria), (iii) more variable genome sizes in ss(+)RNA, from 4.2 kb for the phage Qβ to 31.3 kb for the murine hepatitis coronavirus, *versus* a 7.8-15.9 kb range for all other RNA viruses for which mutation rates are available, and (iv) the anomalously low mutation rate for the bacteriophage Φ6, representing the only estimate for dsRNA viruses (Chao et al., 2002).

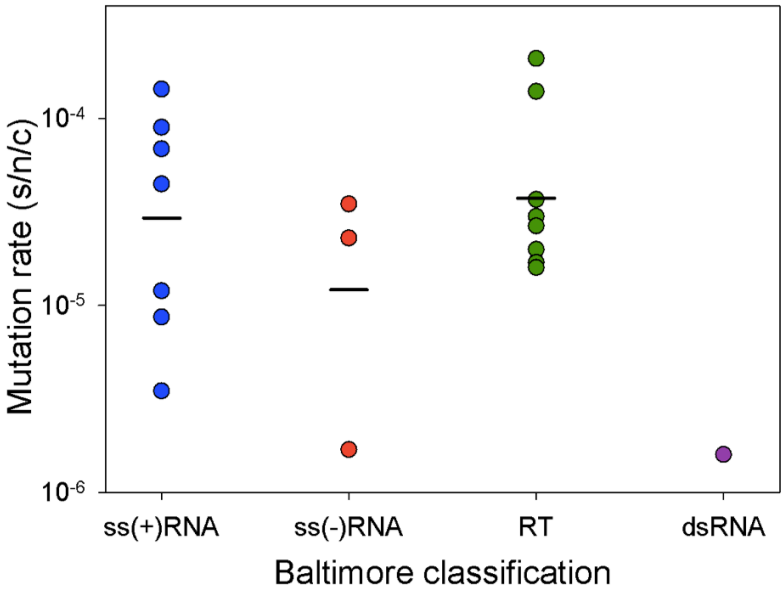


Figure 2. Mutation rates *versus* genome polarity. Mutation rates are represented for the four groups of RNA viruses established by the Baltimore classification. Horizontal black lines represent the geometric mean (log-scale mean) for each category. Adapted from Combe and Sanjuán (2014a).

3.1.3 Viral replication mode

Riboviruses can replicate following several models (Figure 3). Under the stamping machine (SM) model, a single template is repeatedly used to produce all progeny strands within a cell, which do not become template until they infect another cell. Therefore under the SM model due to the complementarity of the RNA, there are only two rounds of genome copying per cell, allowing a linear accumulation of the progeny and thus low mutation frequencies are expected (Safari and Roossinck, 2014; García-Villada and Drake, 2012). In contrast, under geometric

replication (GR), once a new progeny strand is synthesized it also becomes template. There are thus multiple (i.e. more than two) rounds of genome copying per cell, leading to an exponential growth of the virus (Luria, 1951; Sardanyés et al., 2009; Thébaud et al., 2010), and thus higher mutation frequencies are expected (Safari and Roossinck, 2014).

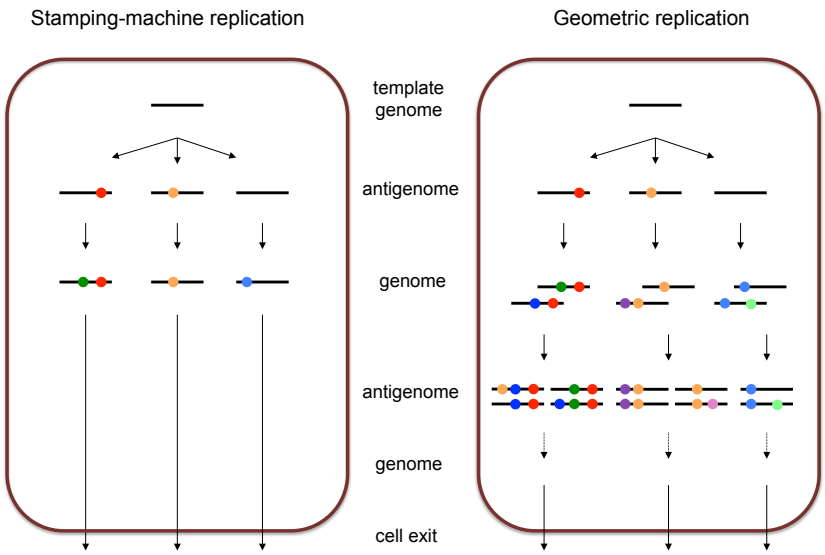


Figure 3. Viral replication modes. Under the stamping machine replication model, a single template is used to produce all the progeny strands within a cell. There are only two rounds of genome copying per cell, thus progeny accumulate linearly, as well as mutations (coloured dots). In contrast, under geometric replication, once a new progeny strand is synthesized it also becomes template within the same cell, thus there are multiple (i.e. more than two) rounds of genome -> antigenome -> genome copying per cell and progeny genomes and mutations accumulate exponentially. Adapted from Combe and Sanjuán (2014a).

Viral replication under the stamping machine or the geometric model impact differently the pattern of intra-cellular mutation accumulation (García-Villada and Drake, 2012; Schulte et al., 2015), and the mean and variance of the number of mutations after one infection cycle will be higher under the geometric model (Combe and Sanjuán, 2014a). However, scarce information on the mode by which RNA viruses replicate within a host cell is available. Only few studies published to date, conducted with the bacteriophage φ6 (Chao et al., 2002), φX174 (Denhardt and Silver, 1966), Qβ (García-Villada and Drake, 2012) and the ss(+)RNA turnip mosaic virus (Sardanyés et al., 2011) and Wheat streak mosaic virus (French and Stenger, 2003), suggested that riboviruses replicate following the stamping machine model. However, the mutant distribution for φ6 differed from the expected Poisson distribution, suggesting an exponential component of the replication dynamics (Chao et al., 2002). Also, Martínez et al. (2011) investigated the dynamics of (-)- and (+)-strands RNA of tobacco mosaic virus in the protoplast of *Nicotiana benthamiana*, and proposed that the mode of replication of this virus might be a mixed strategy of geometric and stamping machine, although 90% of the synthesized genomes derived from the stamping machine model. For poliovirus, a recent study combining a mathematical model with quantitative data on the production of RNA and viral particles showed that multiple generations elapsed within a single cell infection (Schulte et al., 2015). For ss(-)RNA viruses, although it has been proposed that the replication mode of VSV might be mixed (Sanjuán et al., 2010),

the high ratio of (-)- to (+)-strands also support a stamping machine model (Safari and Roossinck, 2014).

3.2 Host factors acting on mutation rates

In addition to virus-encoded factors, such as the intrinsic viral polymerase fidelity (Menéndez-Arias, 2009; Pfeiffer and Kirkegaard, 2003; Arnold et al., 2005; Korneeva and Cameron, 2007), the presence/absence of proofreading mechanisms (Roberts et al., 1988; Steinhauer et al., 1992; Denison et al., 2011), or the mode of replication (Chao et al., 2002; Sardanyes et al., 2009), host factors may also determine viral mutation rates.

3.2.1 Variability at the cellular level

Viruses depend on intra-cellular resources for nucleic acid and protein biosynthesis. For instance, variation among cell types of the total intra-cellular concentration and balance of dNTPs may affect replication fidelity in retroviruses (Diamond et al., 2004; Bebenek et al., 1992; Julias and Pathak, 1998). The impact of cell metabolism on viral mutation rates has also been exemplified in a study showing that ethanol-derived reactive oxygen species (ROS) are mutagenic for hepatitis C virus, whereas glutathione or iron chelators were found to have the opposite effect (Seronello et al., 2011). Moreover, the expression of specific cellular genes, such as APOBEC3G cytidine deaminases, can edit and produce G-to-A hypermutation in retroviruses (Harris et al., 2003; Lecossier et al., 2003; Mangeat et al., 2003; Bonnac et al., 2013) and thus impact viral mutation rates. Indeed, it has been shown that expression levels of APOBEC3G could determine the rate at which

hypermutated HIV-1 genomes accumulate throughout the course of an infection and influence disease progression, with increased hypermutant frequency associated with slower progression rates (Land et al., 2008).

Cellular editing is not restricted to retroviruses, since the RNA-dependent adenosine deaminase (ADAR) has been shown to induce A-to-G hypermutation in several RNA viruses, including measles virus (Cattaneo et al., 1988; Schmid et al., 1992), human parainfluenza virus (Murphy et al., 1991), respiratory syncytial virus (O'Hara et al., 1984; Martinez et al., 1997), influenza virus (tenOever et al., 2007), vesicular stomatitis virus (Bass et al., 1989), paramyxoviruses (Rueda et al., 1994), lymphocytic choriomeningitis virus (Zahn et al., 2007), Rift Valley fever virus (Suspène et al., 2008), mumps virus (Chambers et al., 2009) and hepatitis C virus (Taylor et al., 2005). The ADAR enzyme, expressed in a wide variety of mammalian cell types, is a double-stranded RNA (dsRNA) binding protein that modifies cellular and viral RNA sequences by adenosine (A) deamination to inosine (I) (Nie et al., 2007). Inosine preferentially base pair with cytosine, leading thus to the replacement of genome-encoded adenosine (A) by guanosine (G) (Suspène et al., 2008). In the cytoplasm of the host, viral dsRNAs bind and activate cytoplasmic viral sensor proteins (Takeuchi and Akira, 2010) that need the presence of dsRNA ends and 5' triphosphates to discriminate virus and pathogen RNAs from host cytoplasmic RNAs (Mannion et al., 2014). The stimulation of these sensors further activate the cellular antiviral response *via* activation of (i) type I interferon

(IFN) and proinflammatory cytokines (Geiss et al., 2001) and (ii) dsRNA-activated enzymes such as ADAR that inhibit viral replication (Samuel, 2001).

Whilst the formation of dsRNA has been long thought to be a common feature to all viruses (Jacobs and Langland, 1996; Kumar and Garmichael, 1998), Weber et al. (2006) confirmed its production in infected cells by (+)-strand RNA viruses, dsRNA viruses and DNA viruses. In contrast, dsRNA intermediates were not detected for (-)-strand RNA viruses, although these viruses are able to induce interferon synthesis (Weber et al., 2004), activate PKR (Bergmann et al., 2000; Stojdl et al., 2000; Streitenfeld et al., 2003) or even express a dsRNA-binding protein (Hatada et al., 1992; Lu et al., 1995). Alternatively, in (-)-strand RNA viruses other viral structures may play the role of dsRNA as an antiviral signal for the host cell (Weber et al., 2006), such as the ribonucleoprotein particles of VSV and measles virus capable of triggering IFN induction (tenOever et al., 2002; tenOever et al., 2004). Because ADAR is induced by viral infection and IFN activation, it has been proposed to play a role in host defense mechanisms (Patterson et al., 1995; Rabinovici et al., 2001), although recent studies have demonstrated that it can also promote the replication of many RNA viruses (Gélinas et al., 2011). However, the RNA editing by ADAR might affect the outcome of virus-host interactions (Samuel, 2011), such as in measles infection where hypermutations were associated with changes in viral pathogenesis and persistent infection (Cattaneo et al., 1988), or as showed for human hepatitis virus where ADAR

inhibited viral replication and compromised virus stability as a potential antiviral mechanism (Hartwig et al., 2006; Jayan and Casey, 2002; Taylor et al., 2005).

3.2.2 Variability at the host species level

The slower evolution of arthropod-borne viruses compared to directly transmitted viruses (Jenkins et al., 2002; Woelk and Holmes, 2002; Hicks and Duffy, 2014) has been interpreted in terms of fitness trade-offs, where beneficial mutations in one host can be costly in the other host, thus restricting viral adaptation in the long term (Weaver et al., 1999; Coffey et al., 2008). This trade-off has been well documented in dengue virus, in which the analysis of sequence diversity revealed strong purifying selection (Holmes, 2003; Holmes and Rambaut, 2004). However, previous work has not tested for possible differences between the rates of spontaneous mutation of arboviruses and those of other viruses. This may, though, offer an alternative explanation for the relatively slow evolution of arboviruses. Also, the relatively high genetic homogeneity of plant RNA viruses may be explained by the fact that positive selection of new beneficial mutations is weaker compared to their animal counterpart (Hall et al., 2001; Rodriguez-Cerezo et al., 1991; Kim et al., 2005; Marco and Aranda, 2005; Fraile et al., 1997). This can be the result of (i) weaker immune pressure than in animal viruses (García-Arenal et al., 2001), or (ii) increased genetic drift, relative to selection, during population bottlenecks operating in leaf-to-leaf (Li and Roossinck, 2004) and inter-host transmission (Ali et al., 2006). Moreover, it has been proposed that plant-

specific spatial structures could impose a limit to the fixation of beneficial mutations by reducing the opportunities for competition among genetic variants (Elena et al., 2011). Recent studies showed that the mutation rates of turnip mosaic virus and tobacco etch virus were slightly above 10^{-6} s/n/r (Tromas and Elena, 2010; de la Iglesia et al., 2012), which, similar to the rate for dsRNA bacteriophage Φ6, fall in the lower part of the typical RNA virus range (Figure 4). In support of this finding was the observation that the fidelity of the viral replicase of cucumber mosaic virus (CMV) depends on the host plant in which the virus replicates, since insertions/deletions were more frequent in pepper than in tobacco plants (Pita et al., 2007). These different levels of genetic diversity between host plants have been attributed to either differences in viral replicase fidelity, that could be host-dependent, or, alternatively, to host-specific selection pressures and bottlenecks associated to cell-to-cell movement that limited the accumulation of diversity (Schneider and Roossinck, 2001).

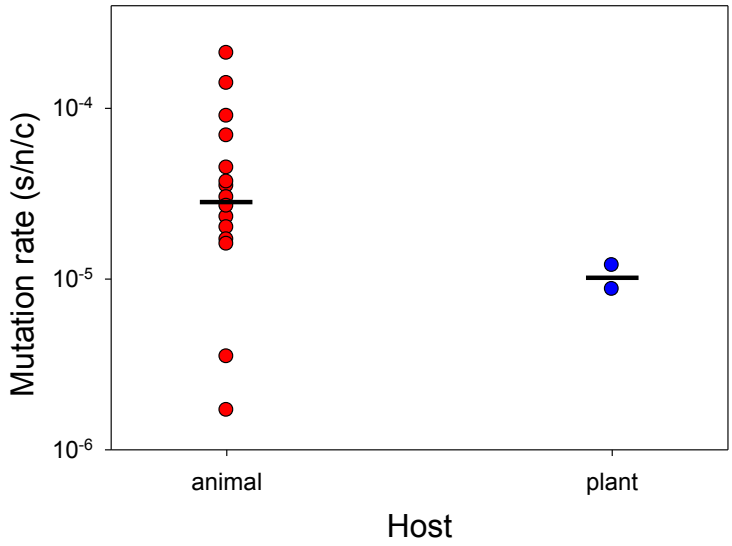


Figure 4. Animal *versus* plant RNA virus mutation rates. Horizontal black lines represent the geometric mean (log-scale mean) for each category. Adapted from Combe and Sanjuán (2014a).

4. Experimental models

4.1 Rhabdovirus: Vesicular stomatitis virus

4.1.1 History and generalities

Vesicular stomatitis virus (VSV) is a lytic, enveloped, non-segmented RNA virus of negative polarity, belonging to the order *Mononegavirales*, family *Rhabdoviridae* and genus *Vesiculovirus* (group V in Baltimore's classification). The virion is described as a bullet-shaped particle of 180 nm in length and 75 nm in diameter (Figure 5).

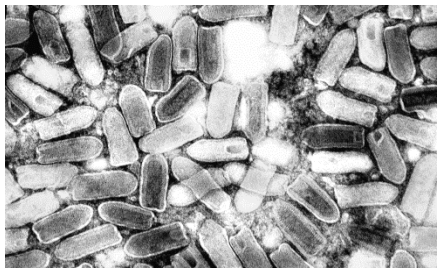


Figure 5. Electron microscope photograph of VSV particles, courtesy of Professor FA Murphy and UC Davis (http://www.vetmed.ucdavis.edu/viruses/download.html).

VSV is the causative agent of the zoonotic vesicular stomatitis disease in multiple livestock species, as well as a variety of wild animals, that gives rise to important public health, veterinary and agricultural impacts (Kuzmin et al., 2009; Reis et al., 2009). Although rhabdoviruses are distributed worldwide, vesiculoviruses are endemic between southern North America and northern South America, where they cause annual outbreaks usually associated with the change between rainy and dry seasons. Vesicular stomatitis is not a newly emerged disease since outbreaks of "sore tongue" in horses between 1801-1802, and then in 1817, were reported in the easter USA, that were followed by important epidemics between 1862-1995 (Letchworth et al., 1999).

The two main VSV serotypes are the Indiana serotype (VSIV) and the New Jersey serotype (VSNJV), which account for 80% of the outbreaks in the US (Kuzmin et al., 2009). In infected animals, the clinical disease signs are like those of foot-and-mouth disease and are characterized by initial vesicular formations on

the feet and in the mouth that progress to erosions and ulcerations on the tongue, palate, gum, lips, snout, teats, prepuce, interdigital space and coronary band, whereas in humans symptoms are like those of influenza (Letchworth et al., 1999). The incubation time is variable but usually vesicles are visible within 24-72 h after viral infection, and recovery occurs within two to three weeks of vesicle formation (Reis et al., 2009).

4.1.2 VSV host range and lifecycle

VSV has a wide host range since it can infect a very large number of mammal species including livestock (cattle, horse, swine, goats, etc.) and wild animals (rodents, bear, lynx, bats, etc.), and also infects insects (sandflies, blackflies, mosquitoes, etc.) (Letchworth et al., 1999; Kuzmin et al., 2009) which act as transmission vectors (Comer et al., 1990; Mead et al., 2004; Tesh et al., 1972). Indeed, VSV is an arthropod-borne virus (arbovirus) and its lifecycle involves horizontal transmission from insects to mammals during blood meals (Comer et al., 1990; Mead et al., 2004; Mead et al., 1999) and vertical (i.e. transovarian) transmission from infected female insects to their offspring (Tesh et al., 1972) (Figure 6). Although transovarian transmission in insects can be maintained for several generations, the virus can also be lost. Therefore, replication in both arthropods and vertebrates is necessary in the lifecycle of arboviruses, explaining the necessity of mammalian infections (Zárate and Novella, 2004).

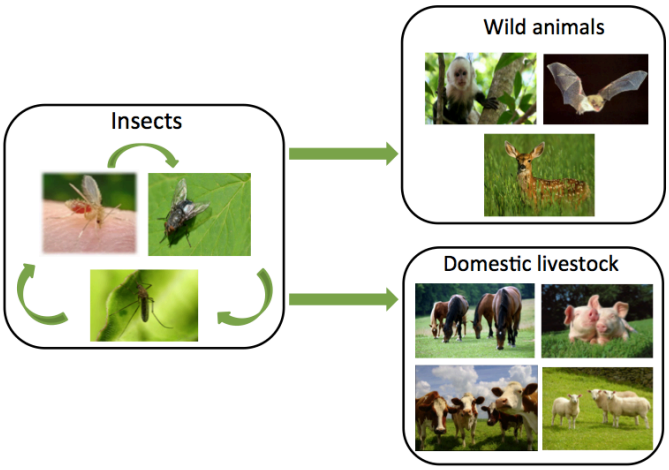


Figure 6. Natural life cycle of VSV. Within the insect reservoir, arrows indicate vertical (i.e. transovarian) transmission of the virus from infected female insects to their offspring. Horizontal arrows indicate transmission of the virus between insects and mammals (wild and domestic animals) during blood meals. In mammals, VSV can transmit by contact between an infected and a non-infected animal, although the mechanisms of transmission are unknown.

4.1.3 VSV genome organization

The VSV RNA genome has 11,161 nucleotides in length and packaged in the form of an internal ribonucleoprotein (RNP) complex. The genome comprises five genes, a leader and a trailer regulatory sequences organized in the following order 3'-le, N, P, M, G, L, tr -5' (Abraham and Banerjee, 1976, Ball and White, 1976) (Figure 7). The genome is never found as a naked RNA during the replication cycle in the cytoplasm of infected cells, but rather is always packaged by the nucleocapsid (N) protein to form a nuclease-resistant helical core, that is the functional template for RNA synthesis (Barr et al., 2002). Because negative-sense RNA

viral genomes cannot be used as mRNA, and host cells do not have appropriate enzymes to catalyze transcription (Lichty et al., 2004), initiation of the infection cycle relies on a viral RNA-dependent RNA-polymerase (RdRp), composed by the large (L) protein and a phosphoprotein (P), which are both bound to and interact with the RNP core to perform viral transcription and replication. The remaining two components of the VSV virion, the matrix (M) protein and the glycoprotein (G) are not essential for RNA synthesis. Nevertheless, M protein serves as condensation of the RNP core during assembly and participates in the budding of the virion from the cell membrane, while the G protein is the functional unit for both assembly into virions and virus binding and entry *via* host cell receptors (Kuzmin et al., 2009). Also, the M protein has a crucial role in the early phases of viral infection since it avoids the cellular antiviral programs by interrupting cellular transcription and blocking mRNA export from the nucleus (Lichty et al., 2004; Kuzmin et al., 2009).

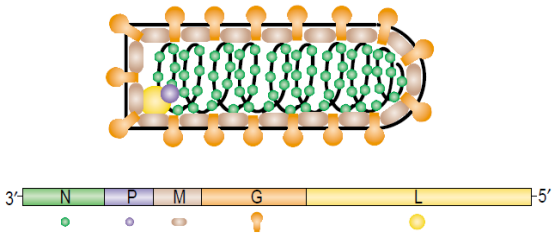


Figure 7. Vesicular stomatitis virus genome organization. The ss(-)RNA genome is composed of five structural genes, plus leader and trailer regulatory sequences: 3'-(leader), the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G), the large polymerase protein (L), and (trailer)-5'. From Lichty et al. (2004).

4.1.4 VSV genome replication

The replication cycle of the VSV genome is similar to those of most negative-stranded RNA viruses and requires both transcription and replication of the (-)-strand RNA molecule. Intra-cellular replication involves the synthesis of a full-length (+)-strand RNA, which then serves as the template for the production of progeny (-)-strand RNA (Peluso and Moyer, 2002).

The adsorption of VSV to the cell membrane is mediated by the G protein using nonspecific electrostatic and hydrophobic interactions. VSV penetration uses clathrin-dependent endocystosis into coated vesicles, which become an endosome 1-2 min after adsorption. Then, a pH drops below 6.5 allows the fusion of the endosome membrane with the viral envelope (Letchworth et al., 1999). Following attachment, penetration and uncoating, the genome is delivered into the cytoplasm, where RNA synthesis occurs (Follett et al., 1974). This process starts by a primary transcription, where the virion polymerase responds to specific signals to transcribe six discrete RNAs: a 47-nucleotide leader RNA that is neither capped nor polyadenylated, and five mRNAs that are capped and methylated at the 5'end and polyadenylated at the 3' end. Following translation of the mRNAs into sufficient yield of viral proteins, genome replication starts. The RdRp binds to the 3' end of the genome, ignores all the signals for production of discrete monocistronic mRNAs, and synthesizes a full-length complementary antigenome. This antigenomic RNA serves as template for synthesis of a 45-nucleotide minus sense leader RNA, and also for synthesis of full-length progeny genomes. These new

ss(-)RNA progeny genomes can be then used as templates for secondary transcription, or assembled into infectious particles and exit the cell (Figure 8) (Barr et al., 2002).

The early events of the replication cycle (i.e. attachment, penetration, uncoating and primary transcription) take place within the first hours postinfection, whereas the remaining steps (i.e. genome replication, secondary transcription and assembly) occur within 12-18 hours (Reis et al., 2009). At the end of the cycle, transcription and replication are inhibited, the RNP becomes intensively condensed and is subsequently delivered to the cell membrane, where the virion acquires the glycoprotein and lipid envelope for self-assembly (Kuzmin et al., 2009).

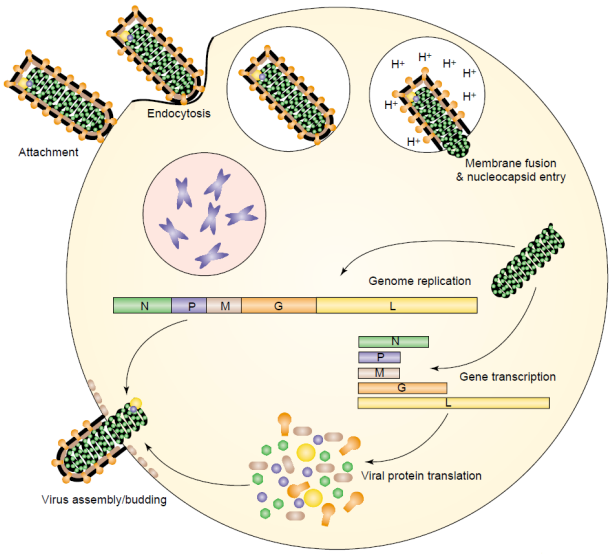


Figure 8. Vesicular stomatitis virus genome replication. Following attachment, penetration and uncoating, the genome is delivered into the cytoplasm of the infected cell where the RNA synthesis occurs. At the end of the replication cycle, the virion is delivered to the cell membrane for self-assembly and exits the cell by lysis. From Lichty et al. (2004).

4.1.5 VSV as a model for experimental evolution

Vesicular stomatitis virus is one of the most widely used systems for the experimental study of viral evolution because of its relatively simple and well characterized genome organization, fast replication, and its high adaptability and capacity to spread between infected cells (Letchworth et al., 1999; Novella, 2003). Also, since the virion attaches to some ubiquitous cell surface receptor it can productively infect mammalian cells from various origins (Whelan, 2008), as well as insect cells (Novella, 2003).

4.2 Calicivirus: Norwalk virus

4.2.1 History and generalities

Norwalk virus (NV) is a small, non-enveloped, non-segmented RNA virus of positive polarity, and is the prototype strain of the *Norovirus* genus in the family *Caliciviridae* (group IV in Baltimore's classification). The virion particle is about 38 nm in diameter (Figure 9).

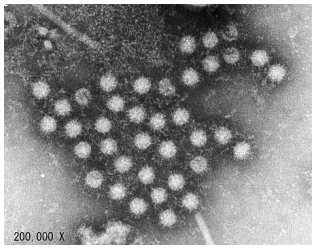


Figure 9. Electron microscope photograph of Norwalk virus from stool sample from an individual with gastroenteritis, F. P. Williams, U.S. Environmental Protection Agency (http://pathmicro.med.sc.edu/).

Noroviruses are human pathogens responsible for more than 90% of outbreaks of non-bacterial gastroenteritis in developed countries, and as many as 23 million cases of gastroenteritis in the United States each year are caused by these infectious agent (Fankhauser et al., 1998; Mead et al., 1999). Outbreaks are typically associated with winter seasonality (Guix et al., 2007; Allen et al., 2008) and are mainly reported in hospitals and elderly care homes, but have also been associated with other semi-closed environments such as schools, restaurants, cruise ships and hotels (Lopman et al., 2003; Mounts et al., 2000). The clinical features of norovirus infections consist of an incubation period of 12-48 h characterized by acute onset of nausea, vomiting, abdominal cramps and diarrhea that generally last for about 48 h (Nilsson et al., 2003). However, due to their short incubation time, low infectious dose, efficient transmission and the short duration of protective immunity, noroviruses cause important morbidity in the human population and impose a significant economic burden on national health services (Lopman et al., 2003; Lopman et al., 2004). The majority of human noroviruses are divided in two genogroups, GI and GII, the most frequently detected noroviruses associated with outbreaks belonging to genogroup GII-4 (Gallimore et al., 2007).

4.2.2 NV lifecycle

The transmission route of noroviruses is fecal-oral *via* aerosols, fomites, food, or water (Dingle, 2004). Although human noroviruses were originally identified more than 30 years ago, our understanding of their replication cycle and mechanisms of

pathogenecity has been limited because these viruses are noncultivable in established cell lines and an animal model to study viral infection is not available (Duizer et al., 2004). Indeed, the factors that block norovirus replication in standard cell cultures remain unknown. Therefore, features of their lifecycle have been essentially inferred from studies using other animal caliciviruses sustained by mammalian cell cultures (Guix et al., 2007). It has been only recently discovered a murine norovirus that is able to replicate in a murine macrophage-like cell line (Wobus et al., 2014). Also, transfection of a full-lenght cDNA clone of the NV genome into modified vaccinia Ankara (MVA)-T7 infected cells allowed the expression of viral proteins and subsequent RNA replication (Asanaka et al., 2005).

4.2.3 NV genome organization and replication

The single-stranded, positive sense RNA genome of approximately 7.7 kb in length encodes the viral structural and non-structural proteins in three open reading frames (ORF) (Figure 10). The genomic RNA is protein-capped at the 5' end and polyadenylated at the 3' end (Hardy, 2005). ORF1 encodes a polyprotein that is post-translationally cleaved by the viral protease into the non-structural proteins including VPg-like protein, viral protease and an RdRp. ORF2 encodes the major structural protein VP1, which is organized into distinct domains (Prasad et al., 1999), including a conserved shell (S) domain, which forms the contiguous shell of the virus, and a protruding (P) domain that extends away from the capsid (Cao et al., 2007). The P domain is organised into a highly variable P2 region, flanked on

either side by more conserved P1-1 and P1-2 domains. Whereas P1 is more interior, P2 is located at the outer surface, suggesting that receptor binding function and antigenic properties of the capsid are localised in this region (Tan and Jiang, 2005; Lochridge et al., 2005). Therefore, P2 might be responsible for host attachment and recognition by the host immune system (Cao et al., 2007; Prasad et al., 1999). Finally, ORF3 encodes a small basic protein believed to be involved in the assembly of progeny particles (Jiang et al., 1993; Glass et al., 2000; Bertolotti et al., 2003).

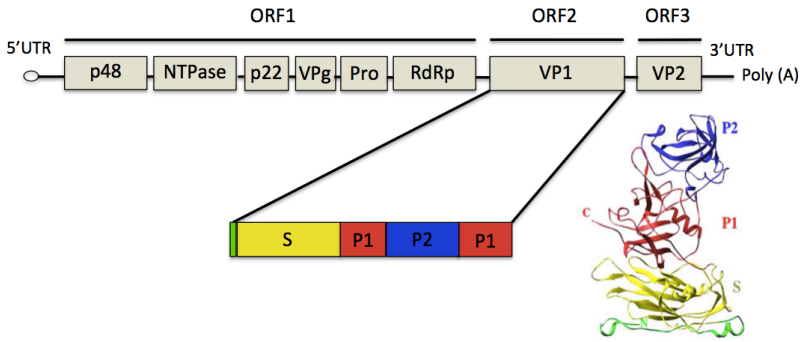


Figure 10. Norwalk virus genome organization. Are indicated, the location of the nonstructural proteins within the ORF1 polyprotein, the viral structural proteins VP1 and VP2 encoded by ORF2 and ORF3, respectively, and the 5' and 3' untranslated regions. Within VP1, conserved (P1) and hypervariable (P2) domains are represented and the colors correspond to the ribbon structure. The small N-terminal (in green) faces the interior of the particle. The shell (S) domain is colored yellow and extends from amino acid 50-225. The P1 subdomains are colored red and comprise amino acids 226-278 and residues 406-520. The P2 domain (blue) is an insertion in the P1 domain and consists of amino acids 279-405. Adapted from Asanaka et al. (2005), Hardy et al. (2005) and Prasad et al. (1999).

Upon infection of cells, nonstructural proteins are expressed from genomic RNA and form an RNA replication complex, which generates new genomic RNA molecules as well as subgenomic RNAs encoding VP1 and VP2. After expression of the structural proteins from these subgenomic RNA molecules, the capsid is assembled and viral RNA encapsidates prior to progeny release (Guix et al., 2007).

5. Objectives

The objective of this PhD thesis is to investigate some host factors underlying the variability in RNA virus mutation rates. First, we studied the effect of the cell type being infected, and the variation depending on the host species in which the virus replicates. Then, we focused on the variability occuring at the single-cell level, by characterizing the genetic diversity of viruses released from individual cells. Finally, we investigated the potential effect of the RNA editing of host enzymes, such as ADAR, by determining the type of spontaneous mutations occuring in noroviruses.

MATERIALS AND METHODS

1. Viruses

1.1 Vesicular stomatitis virus

Viruses were obtained from an infectious cDNA clone by transfecting baby hamster kidney (BHK-21) cells, purified by filtration (0.22 μm), and stored at -70ºC in aliquots until use. Briefly, as described in Sanjuán et al. (2004), confluent (approx. 10^4 cells/mm^2) BHK-21 cells were infected with a recombinant vaccinia virus vTF7-3 (American Type Culture Collection) that expressed bacteriophage T7 RNA polymerase. After 1 h incubation, BHK-21 cells were transfected with a full-length cDNA clone (provided by G. T. W. Wertz, University of Alabama at Birmingham, Birmingham) and three helper plasmids encoding the N, P, and L proteins of VSV (Whelan et al., 1995). Transfections were performed by using Lipofectamine supplemented with Plus Reagent (Invitrogen) and, to inhibit vaccinia replication, 25 μg/mL 1-β-D-arabinofuranosylcytosine (AraC) were added 6 h post-inoculation (hpi). The supernatant was harvested 96 hpi, and residual vaccinia was removed by filtering the supernatant through 0.22 μm membranes (Millipore). Dilutions of the supernatant (10^2- to 10^4-fold) were used for plaque assays that allowed to score for the presence of plaque-forming units (PFU) 24 hpi, indicating the successful recovery of infectious VSV particles. The sequence of this virus was identical to GenBank

accession number AM690336, except for six point substitutions (A3351G, C6190A, A6523G, A7421C, U9532C, U11136A).

1.2 Norwalk virus

The cDNA clone of Norwalk virus (NV) was provided by Dr. M. K. Estes (Departments of Molecular Virology and Microbiology and Medicine, Baylor College of Medicine, Houston). As described by Asanaka et al. (2005), the construction of subgenomic NV cDNA was based on the nucleotide sequence of subgenomic RNA found in animal caliciviruses (Green et al., 2001). The infectious cDNA was cloned in our laboratory by bacterial transformation (XL10-Gold ultracompetent cells, Agilent Technologies), under heat shock conditions (42°C for 45 sec followed by 2 min in ice). A colony-forming-unit (CFU) was picked and amplified overnight in Luria Broth (LB) media supplemented with Ampicilin (Amp). The cDNA was purified using midiprep (PureLink® HiPure Plasmid Midiprep kit, Invitrogen™) and then stored in aliquots at -20°C until use.

Tranfections were performed using Lipofectamine® LTX Reagent (Invitrogen™). The day before transfection, 1×10^5 HEK293T cells were split in a 24-well microplate in 500 µL of media in order to reach a cell density of 50-80% the day of transfection. The NV cDNA was transfected into HEK293T cells previously infected with the recombinant vaccinia vTF7-3 virus (see section 1.1) at a MOI 10 for 1 h, which initiated the primary transcription of NV cDNA to RNA. Once RNA was synthesized, it replicated by itself into the cytoplasm of the cells, and the activity of T7 RNA polymerase was blocked by adding AraC (1-β-D-

arabinofuranosylcytosine) to the culture. To do so, for each well to be transfected, 0.5 µg of plasmid were diluted in 100 µL of DMEM, to which 1.75 µL of Lipofectamine® LTX Reagent were added. The solution was mixed gently and incubated for 30 min at room temperature to form the DNA-Lipofectamine® LTX Reagent complexes. After 1 h incubation at 37°C (shaking at 30 min) cells were washed two times with 0.2 mL DMEM and a third time with 0.15 mL DMEM. After 30 min incubation, 100 µL of the DNA-Lipofectamine® LTX Reagent complexes were added to each well to be transfected and mixed gently by rocking the plate. Then, after 5 h incubation at 37°C, 250 µL of AraC (1 mg/mL) diluted in DMEM supplemented with 2% FBS were added to each well and the plate was incubated for 48 h. As control for transfection efficiency, a VSV plasmid harboring a green-fluorescent-protein (GFP) was also transfected under the same conditions.

2. Cells

2.1 Mammalian cells

2.1.1 BHK-21 cells

Baby hamster kidney (BHK-21) cells are originally derived from fibroblasts of a 1-day-old hamster, *Mesocricetus auratus*. BHK-21 cells were obtained from the American Tissue Culture Collection (ATCC® CCL-10™) and were cultured in DMEM (Dulbecco's modified Eagle medium) supplemented with 10% fetal bovine serum (FBS), 0.02 mM L-Glutamine, a mix of non-essential amino-acids, 100 µg/mL streptomycin, 60 µg/mL

penicillin, and 2 µg/mL fungizone. Since BHK-21 cells are adherent, confluent cells. Confluent cultures were washed with phosphate-buffered saline (PBS), detached with trypsin, split 1:6 every two days and incubated at 37ºC under 5% CO_2 (Figure 11A).

2.1.2 HEK293T cells

Human embryonic kidney (HEK293T) cells are originally derived from 293 cells that were immortalized with adenovirus 5 (strain F2853-5b) and transformed with the temperature sensitive simian virus (SV40) large T antigen. HEK293T cells were obtained from the American Tissue Culture Collection (ATCC® CRL-11268™) and were cultured in DMEM F12 (Dulbecco's modified Eagle medium) supplemented with 10% FBS and antibiotics. Adherent HEK293T cells were washed with PBS, detached with trypsin and split 1:10 twice a week. Cells were incubated at 37°C under 5% CO_2 (Figure 11B).

2.1.3 MEFs and MEF p53-/- cells

Mouse embryonic fibroblast (MEF) cells and their p53-/- derivative were obtained from Dr. Carmen Rivas (Centro Nacional de Biotecnología, Madrid) and cultured in DMEM supplemented with 12% FBS, 0.02 mM L-Glutamine, a mix of non-essential amino-acids, 100 µg/mL streptomycin, 60 µg/mL penicillin, and 2 µg/mL fungizone. Adherent cells were washed with PBS, detached with trypsin, split 1:3 every three days and incubated at 37ºC under 5% CO_2 (Figure 11C).

The p53 tumor suppressor gene is a regulator of normal cell proliferation and is mutated in several human cancers,

including tumors of colon, breast, lung and brain (Jacks et al., 1994). The MEF p53-/- cell line was also provided by Dr. Rivas.

2.1.4 Neuro-2a cells

Mouse neuroblastoma (Neuro-2a) cell line was originally established from a spontaneous tumor of a strain A albino mouse, *Mus musculus*. Neuro-2a cells were obtained from Prof. José M. García-Verdugo (Department of Cell Biology, University of Valencia) and cultured in MEM (Eagle's minimum essential medium) supplemented with 2 mM L-Glutamine, 1 mM sodium pyruvate, 10% FBS, non-essential amino acids and antibiotics. Adherent Neuro-2a cells were washed with PBS, detached with trypsin and split in a ratio 1:3 every three days. Cells were incubated at 37ºC under 5% CO_2 (Figure 11D).

2.1.5 CT26 cells

Colon cancer CT26 cells originate from an undifferentiated grade IV colon adenocarcinoma, induced by intrarectal injections of N-nitrous N-methylurethamine into a BALB/c female mouse (Corbett et al., 1975). CT26 cells were obtained from the American Tissue Culture Collection (ATCC® CRL-2638™) and were cultured in DMEM supplemented with 10% FBS, 2 mM L-Glutamine, 10 mM HEPES and antibiotics. Adherent cells were washed with PBS, detached with trypsin, split 1:8 twice a week, and incubated at 37ºC under 5% CO_2 (Figure 11E).

2.2 Insect cells

2.2.1 S2 cells

Drosophila S2 cells were originally derived from a primary culture of late-stage (20-24 hours old) *Drosophila melanogaster* embryos (Schneider, 1972). S2 cells were obtained from Dr. Rubén Artero (Department of Genetics, University of Valencia) and cultured in Schneider's medium (SDM) supplemented with 10% FBS and antibiotics. S2 cells are semi-adherent and were therefore detached by slapping the flask by hand. Cells were split 1:5 once a week, and incubated at 25ºC in the absence of CO_2. Cells were infected at 28ºC (Figure 11F).

2.2.2 Sf21 cells

The sf21 cell line, cloned from ovaries of the Fall Army worm *Spodoptera frugiperda*, was originally developed at the Henry A. Wallace Beltsville Agricultural Research Center (USA). Sf21 cells were obtained from Dr. Salvador Herrero (Department of Genetics, University of Valencia) and were cultured in Grace's insect medium supplemented with 10% FBS and antibiotics. Adherent sf21 cells were detached by hand slapping, split 1:6 once a week, and incubated at 28ºC in the absence of CO_2 (Figure 11G).

2.3 Mosquito cells

2.3.1 C6/36 cells

The C6/36 cell line was originally derived from *Aedes albopictus* whole larvae cells. C6/36 cells were obtained from the American Tissue Culture Collection (ATCC® CRL-1660™) and were cultured in DMEM supplemented with 10% FBS, 2 mM L-glutamine, non-essential amino acids, 1 mM sodium pyruvate and

antibiotics. C6/36 cells are non-adherent and were split 1:8 twice per week, and incubated at 28°C under 5% CO_2 (Figure 11H).

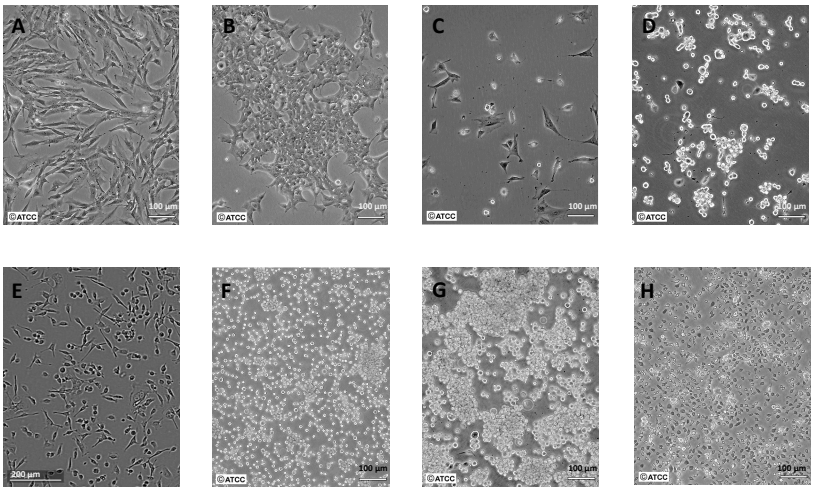


Figure 11. Cell lines. Mammalian cells: A, BHK-21; B, HEK293T; C, MEFs and MEFs p53-/-; D, Neuro-2a; E, CT26. Insect cells: F, S2; G, sf21. Mosquito cells: H, C6/36 (http://www.lgcstandards-atcc.org).

2.4 Cell culture under hypoxia

Hypoxia (1% O_2) was achieved by displacing oxygen with nitrogen, using a Galaxy 170R incubator (Eppendorf).

3. Plaque assays

Plaque assays were performed to determine the proportion of infected cells, the burst size (BS, number of

infectious particles produced per cell), for plaque titration and also to isolate individual plaques for sequencing. Viruses were conveniently diluted and used to inoculate confluent monolayers of BHK-21 cells in 60 mm dishes, which were incubated for 24 h in DMEM semi-solidified supplemented with 0.4% agarose containing 2% FBS. Monolayers were either used to aspirate indidivual plaques, or fixed with 10% formaldehyde and stained with 2% crystal violet to count PFUs (particle-forming-units).

4. Mutation rate estimation using Luria–Delbrück fluctuation tests

4.1 Rationale and design of the test

The Luria-Delbrück fluctuation test was originally proposed by Luria and Delbrück (1943) in order to determine whether phage-resistant bacteria arose from spontaneous mutations during the growth of the bacterial culture, or through mutations induced by interactions with the phages present in the environment (Sarkar, 1991; Sarkar et al., 1992). The principle of this test was based on the growth of several independent cultures of bacteria that mutated with cell division and gave rise to clones of mutants. After the growth period, the cells were plated in the presence of phage to score resistant cells (Stewart, 1994). Luria and Delbrück focused their attention on the first generation of resistant bacteria (i.e. bacteria that survived immediately after the addition of the virus) and hypothesized that they were resistant before the addition of the virus, suggesting that mutations

occurred independently of the virus. Under this hypothesis, mutants that arose early in the culture would increase exponentially through the cell division whereas many cultures would show no mutations at all or only one in the last few generation. Therefore, there should be a large variation in the number of resistant bacteria among cultures. Their results confirmed that the number of phage resistant colonies arose through spontaneous mutations before the bacteria were exposed to the phage, and that mutants increased in numbers as the cultures grew (Stewart et al., 1989). Then, the Luria-Delbrück fluctuation test was applied to viruses and became a standard estimation method used for several viruses, including poliovirus (Sedivy et al., 1987), vesicular stomatitis virus (Furió et al., 2005), influenza A virus (Suárez et al., 1992), measles virus (Schrag et al., 1999), turnip mosaic virus (de la Iglesia et al., 2012), and bacteriophages φ6 (Chao et al., 2002) and Qβ (García-Villada and Drake, 2012). The method has the advantage of being relatively insensitive to selection and the viral replication mode, provided that the null-class method is used (see below). Here, fluctuation tests were used to measure the mutation rate of VSV in mammalian, insect and mosquito cells, and were performed under different experimental conditions, such as BHK-21, MEF, MEF p53-/-, Neuro-2a, CT26, BHK-21 under hypoxia (1% O_2), BHK-21 at 28ºC, and S2, sf21, and C6/36 cell cultures.

In each fluctuation test, we inoculated 32 identical cultures each containing 10^4 confluent cells with approximately 300 pfu/well (N_i) each and incubated them until approximately 3×10^4

pfu/well were produced (N_f) (Table 1). After a round of freeze-thawing to release intracellular particles, we used eight cultures for titration (N_f) and 24 for plating the entire undiluted volume (100 µL) in the presence of a monoclonal antibody against the surface glycoprotein G at a concentration that neutralizes completely the wild-type virus and selects for MAR (monoclonal antibody resistant) mutants (Figure 12). The antibody, in the form of a hybridoma supernatant, was added to the plating medium (25% v:v) to avoid phenotypic masking (Holland et al., 1989).

Table 1. Incubation time for each cell type tested**.** Cultures were infected with approximately 300 pfu/well (N_i) and incubated until approximately 3×10^4 pfu/well were produced (N_f).

Cell type	Incubation time (hpi)
BHK-21 at 37°C	7h 40 min
BHK-21 at 28°C	7h
BHK-21 under 1% O_2	9h
MEF	10h 20min
MEF p53-/-	10h 40 min
CT26	15h
Neuro-2a	8h
S2	24h
Sf21	15h
C6/36	20h

Assuming that mutation is a rare and random event, the number of mutations per culture is expected to follow a Poisson distribution of parameter $\lambda = m(N_f - N_i)$ and therefore the probability of observing no mutants in a culture is $P_0 = e^{-m(N_f - N_i)}$, where m is the mutation rate from the wild-type

to the MAR phenotype (null-class method). However, if there is incomplete plating, some cultures may contain undetected MAR mutants. If we define z as the plating efficiency (relative to BHK-21 cells), the probability of observing no mutants can be expressed as $P_0 = \sum_k Q_k(1-z)^k$, where Q_k is the probability of k actual mutants in a culture. Using a Poisson distribution of parameter $\lambda = m(N_f - N_i)$ for k, we numerically solved Q_0 given P_0, N_i, N_f, and z and calculated the mutation rate as $m = -\ln(Q_0)/(N_f - N_i)$. This gave a mutation rate (m) to the MAR phenotype per round of copying, which was converted to per nucleotide units as $\mu = 3m/T$, where T is the set of observable mutations leading to the phenotype (i.e. mutation target) and three stands for the number of possible nucleotide substitutions per site (Sanjuán et al., 2010). Therefore, mutation rates were expressed as substitution per nucleotide per round of copying (s/n/r).

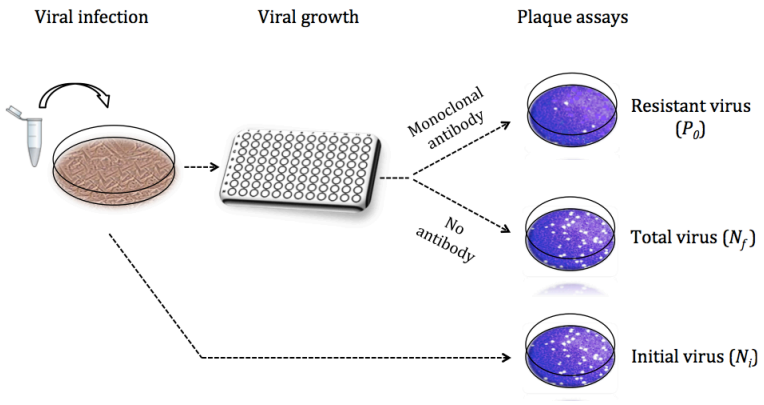


Figure 12. Luria-Delbrück fluctuation test. Confluent cells are infected with a small viral inoculum (N_i) under standard growth conditions, and, after a final number of viral particles (N_f) produced, the entire supernatant is plated in the presence of a monoclonal antibody to score MARMs. The mutation rate from antibody sensitivity to the MAR phenotype (m) can be calculated using the null-class method, which consists of counting the fraction of cultures with zero mutants (P_0).

4.2 Plating efficiency in fluctuation tests

For each cell type tested, the plaque assay for scoring MAR mutants was done on confluent monolayers of BHK-21 cells for technical feasibility and also to control for differences in plating efficiency among cells. However, since plaque assays to score MAR mutants were done with undiluted supernatants, antiviral cytokines or other compounds released from the cells in which the virus was grown could modify plating efficiency. In contrast, plaque assays for determining the number of viruses produced (N_f) were done at a roughly 1/100 dilution and thus were much less affected by this problem. For instance, BHK-21 cells are partially responsive to interferon (Lin et al., 2004), potentially

inhibiting growth of MAR mutants and biasing mutant counts down. To calibrate this effect, we titrated a MAR clone obtained by site-directed mutagenesis (substitution D259A in the surface glycoprotein G) in the presence of undiluted supernatants harvested from cells previously infected with the wild-type virus ($N_i \approx 300$ pfu and $N_f \approx 10^4$ pfu, similar to fluctuation tests), and adding monoclonal antibody to the plates to observe MAR plaques only. The wild-type infections were performed under each of the experimental conditions (BHK-21, MEF, MEF p53-/-, CT26, Neuro-2a, BHK-21 under 1% O_2, BHK-21 at 28°C, S2, Sf21 and C6/36 cells). Based on this, the plating efficiency for each of the other conditions, based on at least six independent plaque assays, was adjusted relative to BHK-21 cells (Supplementary 1).

4.3 Mutational target size in fluctuation tests

To ascertain the number of possible mutations conferring the MAR phenotype (T), we sequenced individual MAR plaques. Viral RNA was purified (Macherey-Nagel NucleoSpin® Viral RNA isolation) and reverse-transcribed using AccuScript High Fidelity Reverse Transcriptase (Agilent Technologies) and specific primers (3'-ACATTCCATCCGATCCTTCAC-5'). The retrotranscription conditions used were 25°C 10 min, 42°C 30 min, 70°C 15 min. Then, the cDNA was PCR-amplified using Phusion High Fidelity DNA polymerase (New England Biolabs), which presents an error rate of 4.4×10^{-7} errors per nucleotide per round of replication, as provided by the manufacturer. Specific primers (forward 3'-ACATTCCATCCGATCCTTCAC-5' and reverse 5'-CGGAGAACCAAGAATAGTCCAATG-3') were used to amplify and

sequence a region of the G protein (genome sites 3361 to 4501 in GenBank accession number EF197793) controlling the MAR phenotype (Holland et al., 1991). The PCR conditions used were 98°C 2 min, 35 cycles of 98°C 5 sec, 66°C 30 sec, 72°C 30 sec, and a final extension step of 72°C 5 min.

PCR products were purified (Macherey-Nagel NucleoFast® 96 PCR plates) and sequenced by Sanger (Genomic Core Facility S.C.S.I.E., University of Valencia). Sequences were cleaned, assembled and then analyzed for SNPs using the Staden software (http://staden.sourceforge.net).

5. Mutation rate estimation using molecular clone sequencing

5.1 Limiting dilution

In order to ascertain the reliability of the estimates obtained by the Luria-Delbrück fluctuation test, we used a molecular clone sequencing approach. To do so, a 96-well plate containing 10^4 BHK-21 cells per well was inoculated with a limiting dilution of the viral VSV stock, such that approximately 10% of wells were infected with a single infectious particle (i.e. plaque forming units, pfu). Plates were incubated at 37°C for 24 h, inspected under the microscope to determine wells showing cytopathic effects (i.e. positive wells), and freeze-thawed to allow release of intracellular viruses. The resulting viral bursts (N_f) were purified (Macherey-Nagel NucleoSpin® Viral RNA isolation) from the supernatant of each of five positive wells, reverse-transcribed and PCR-amplified (same conditions as in section 4.3)

using specific primers located in the P (genome sites 1339-1899), G (genome sites 3858-4347) and L (genome sites 6974-7462) genes (Supplementary 2). To obtain the mutation frequency (*f*), the number of observed mutations was divided by the total number of bases sequenced. Then, the per-generation increase in mutations frequency was calculated as f / c, where c is the number of infection cycles (i.e. viral generations) elapsed. This number was calculated $c = \frac{\ln Nf}{\ln B}$, and where B is the per-cell burst size (BS).

5.2 Molecular cloning

PCR products were gel-purified (Zymoclean™ Gel DNA Recovery Kit) and cloned with CloneJET™ PCR cloning kit (Thermo Scientific), using the linearized cloning vector pJET1.2/blunt, and T4 DNA ligase. Purified PCR products were ligated with the vector in 5 min and the mixture was directly used for bacterial transformation under heat shock conditions (42°C for 45 sec followed by 2 min in ice). Colonies were picked and PCR-amplified using Taq DNA polymerase (VWR) and CloneJET™ primers (forward primer 5'-CGACTCACTATAGGGAGAGCGGC-3' and reverse primer 5'-AAGAACATCGATTTTCCATGGCAG-3'). The PCR conditions used were 95°C 5 min, 35 cycles of 95°C 30 sec, 60°C 30 sec, 72°C 2 min, and a final extension step of 72°C 5 min.

5.3 Effect of selection on mutation frequency

To correct for the effect of selection on mutation frequency, we used the empirically characterized distribution of mutational fitness effects of random single-nucleotide

substitutions in VSV. We did so numerically by simulating the combined effects of mutation and selection. The statistical distribution of fitness effects (s) for viable substitutions can be roughly captured using an exponential distribution truncated at $s = 1$ (lethality) plus a class of lethals occurring with probability p_L: $P(s) = (1 - p_L)\frac{\lambda e^{-\lambda s}}{1-e^{-\lambda}}$ if $0 < s < 1$, $P(s) = p_L$ if $s = 1$, and $P(s) = 0$ otherwise. In a previous work using the same VSV strain as here, it was estimated that $p_L = 0.40$ and that $\mathrm{E}(s) = \frac{1}{\lambda} = 0.13$ (Sanjuán et al., 2004; Sanjuán, 2010). Fitness effects were measured as growth rate ratios, $s_i = 1 - r_i/r_0$, where r is the exponential growth rate and subscripts *i* and *0* refer to the mutant and wild-type, respectively (Sanjuán et al., 2004). These *s*-values were transformed to per cell infection units as $s_i' = 1 - \frac{B^{1-s_i}-1}{B-1}$, where *B* is the burst size. After simulating fitness effects using the truncated exponential plus lethal distribution and applying the per cell infection transformation, selection was applied by picking individuals for the next cell infection cycle with weighted probability 1 – s′, and the process was iterated. This provided an expected mutation frequency f and therefore a relationship between μ and *f*. Genetic drift was ignored since it should not modify the expected value of *f*. Also, for simplicity, mutations were assumed to have independent fitness effects (no epistasis) and back mutations were ignored, which seems reasonable in the short-term, because single forward mutations will greatly outnumber secondary and back mutations. Simulations were performed using Wolfram Mathematica and Excel. A graphical

representation of this correction can be found in the paper of Sanjuán et al. (2010).

6. Analysis of published molecular evolutionary rates

In a previous meta-analysis, our group collected evolutionary rate estimates for RNA viruses, that were originally inferred from field isolates using Bayesian analysis of dated sequences after validation of the molecular clock (Sanjuán, 2012). Here, we used 170 of these estimates, which corresponded to 62 different riboviruses. We sought to compare viruses transmitted directly through respiratory secretions, blood, sexual contact, feces, or animal bites (n = 113) against arboviruses (n = 57). We used a two-way ANOVA in which the following factors were included: transmission mode (fixed), viral family (random) to account for phylogenetic relatedness, and sampling timespan (covariate) to account for the known time-dependency of evolution rate estimates (Jenkins et al., 2002; Ho et al., 2011). Since rates ranged several orders of magnitude log-transformed data were used.

7. Genetic diversity released by single infected cells

7.1 Viral infection and single cell isolation

BHK-21 cells were grown to a confluent adherent monolayer (approx. 10^6 cells/mL), and were then washed with PBS and detached with trypsin. Cells were resuspended in infection media (i.e. culture media with 2% FBS), infected with

VSV at an expected multiplicity of infection (MOI) 0.5 (approx. 5.10^5 pfu/mL), and incubated for 45 min at 37ºC to allow viral adsorption. Infected cells were then diluted 1/20 in a 60 mm dish to isolate individual cells. The physical retrieval of single cells was performed with a glass capillary (NARISHIGE, Glass Capillary G-1), previously home made to fit the diameter of BHK-21 cells and set onto a micromanipulator (NARISHIGE, MN-151), under an inverted microscope. Individual cell were focused at × 100 magnification, aspirated in the glass capillary, and immediately re-injected in a drop of 10 µL of DMEM using an O_2-supplied microinjector (NARISHIGE, IM-300). Drops containing a single cell was pipetted and released into 100 µL of infection media in an assigned well of a 96-well microplate, and cells were incubated for 24 h at 37ºC to allow for viral replication and progeny release. Aliquots of the supernatant were then stored at -70ºC.

7.2 Matrix design and library construction

For each single cell, 10 individual plaques were picked from distinct 60 mm dishes to avoid cross-contamination by viral diffusion in the agar monolayer. Plaques were inoculated in DMEM in 1.5 mL Eppendorf tubes, homogeneized by vortexing, stored in multiple aliquots at -70ºC, and titrated by plaque assays in triplicates. Individual plaques were equalized to 10^5 pfu/mL. Given the large number of plaques (881), pools of 7-10 plaques were made for reasons of tractability before RNA extraction, such that each plaque should be represented at roughly 10% frequency in each pool. To do so, matrices were constructed by allocating each plaque to a unique pair of "Mix" and "Pool" (Table 2). A "Mix"

contained all plaques from the same cell, and a "Pools" contained plaques from different cells. This design allowed for plaque calling after sequencing.

Table 2. Matrix design. For each single infected cell, 10 individual plaques were pooled in equal ratio in a "Mix" and in a "Pool". "Mix 1" to "Mix 10" represent the within cell plaque pooling, from cell 1 to cell 10. "Pool 1" to "Pool 10" represent the between cell plaque pooling, from plaque 1 (1.1, 2.1 … 10.1) to plaque 10 (1.10, 2.10 … 10.10).

Mix 1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10
Mix 2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	2.10
Mix 3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10
Mix 4	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10
Mix 5	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	5.10
Mix 6	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10
Mix 7	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
Mix 8	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	8.10
Mix 9	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	9.10
Mix 10	10.1	10.2	10.3	10.4	10.5	10.6	10.7	10.8	10.9	10.10
	Pool 1	**Pool 2**	**Pool 3**	**Pool 4**	**Pool 5**	**Pool 6**	**Pool 7**	**Pool 8**	**Pool 9**	**Pool 10**

Viral RNA from each "Mix" and "Pool" was purified (Macherey-Nagel NucleoSpin®Viral RNA isolation kit) without freeze thawing of the supernatant, and was reverse-transcribed using Accuscript High Fidelity Reverse Transcriptase (Agilent Technologies). The reverse transcription conditions used were 42°C 60 min, 70°C 15 min. The cDNA was PCR-amplified with Phusion High Fidelity DNA polymerase (New England Biolabs) by using 4 PCR fragments of approx. 3 kb length (Supplementary 3). The PCR conditions used were 98°C 2 min, 35 cycles of 98°C 5 sec, 58-70°C 30 sec, 72°C 30 sec, and a final extension step of 72°C 5 min. PCR products were purified (Macherey-Nagel NucleoFast® 96 PCR plate) and quantified (Qubit® dsDNA HS Assay Kit,

Invitrogen), and the four PCR fragments corresponding to the same “Mix” and “Pool” were mixed in equimolar ratios, leading to a PCR pool. Then, each PCR pool was combined to another one to constitute pairs of PCR used to build 90 tagged libraries using the Library Builder™ Fragment Core Kit (Applied Biosystems®). Consequently, the final frequency of each plaque within a library was expected to be 5%, with each individual plaque being allocated to a unique pair of libraries. Sequencing was performed using the SOLID 5500 XL machine (Life Technologies, Genomic Core Facility S.C.S.I.E., University of Valencia), which is as efficient as Illumina or 454 in detecting single nucleotide polymorphism (SNP) (Ratan et al., 2013). The schematic representation of the full experimental protocol is provided in Figure 13.

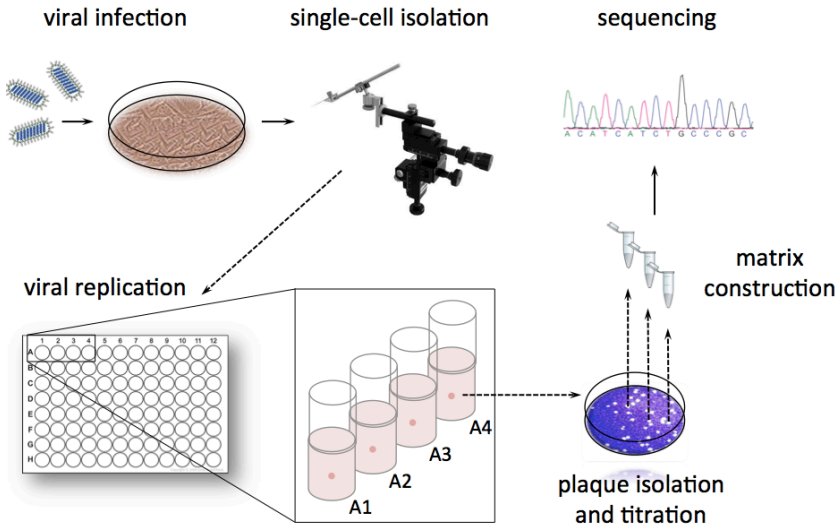


Figure 13. Experimental protocol of the single-cell approach to estimate viral mutation rates. BHK-21 cell cultures were infected with VSV at a MOI 0.5, and after incubation to allow for viral adsorption, single cells were retrieved with a micromanipulator and transferred into a multi-well plate. After incubation (24 h) plaque assays allowed us to determine which wells produced single-cell viral bursts. Individual viral plaques were isolated and their titer was determined. Plaques were pooled in equal titer ratios to construct matrices for sequening.

7.3 Analysis of SOLID data and SNPs calling

Raw sequence data were converted to FASTQ format using using the SOLID™ System XSQ Tools, where 5'-end adapters were removed using CutAdapt (v1.6) (Martin, 2012) and sequences were trimmed 75 base length reads using the FASTXTookit (v0.014) (http://hannonlab.cshl.edu/fastx-toolkit/index.html). Then, FASTQ files were cleaned to remove PCR primer sequences using CutAdapt, with the specific settings of trimming the anchored 5'-end primer sequence and keeping reads ≥ 25

nucleotide length after the clip. PRINSEQ (v0.20.4) was used for dereplication (exact duplicates and exact reverse complement duplicates, derep 14), selection of reads with 25-50 bases length, trimming sequences from the 5'-end with base quality score < 20, and removal of sequences with ambiguous bases (Schmieder and Edwards, 2011; http://prinseq.sourceforge.net/manual.html). Filtered reads were mapped onto the reference VSV sequence (GenBank accession number AM690336) using the local aligner Bowtie 2 (input Phred quality "Phred+33"; Langmead et al., 2009; http://bowtie-bio.sourceforge.net/). Mapped files were then converted into binary format (BAM), sorted and indexed in SAMtools (Kuiken et al., 2005). Base coverage across the genome was computed with BEDTools (Quinlan and Hall, 2010). The average per-library coverage was (1791 ± 1020), with a minimum coverage > 100 in 99.65% of the genome (Supplementary 4). Then, by converting post-alignment BAM files into mpileup format in SAMtools, SNPs were called in VarScan (v2.3.7) using the "pileup2snp" and "pileup2indel" commands. To stringently rule out sequencing errors, only SNPs with ≥ 10 supporting reads at a position (Koboldt et al., 2009; DePristo et al., 2011) (http://varscan.sourceforge.net/), > 2% frequency in the library, and equal ratios in plus and minus orientation (Fisher test: $p > 0.005$) were considered. The entire pipeline was repeated for each library.

This protocol provided a conservative list of SNPs, since although their expected frequency in each library was 5% (for those appearing in a single plaque) or higher (for those appearing

in several plaques), their actual frequency may be lower than 2% as a result of experimental errors. Moreover, some SNPs may be real despite the plus/minus orientation bias. These SNPs were assigned to libraries according to their tagging, and since each individual plaque was allocated to a unique pair of libraries, the presence of a given SNP in a specific pair allowed for its unambiguous assignation to an individual plaque. Some SNPs appeared in only one plaque (scSNPs), whereas others appeared in multiple plaques within a cell, or in multiple plaques in various cells (preSNPs). Therefore, we restricted the maximum copies of a SNP such that the maximal allowed number was 1 for SNPs with < 5% frequency, 2 for SNPs < 10% frequency, and so on. Finally, additional copies of the identified SNPs within a given cell were probed by lowering the frequency cut to 1% and relaxing the requirement for equal ratios in plus and minus orientation reads.

8. Norwalk virus virion recovery and sequence analysis

8.1 Virus recovery

After transfection of NV cDNA into HEK293T cells (see section 1.2 in "Materials and Methods"), viruses were recovered from supernatants and from cells. The supernatant of each well was harvested and centrifuged 15 min at 16,000 × g and at 4°C. Then, the supernatant was transferred in a new 1.5 mL Eppendorf tube and viral RNA was extracted (Macherey-Nagel NucleoSpin® Viral RNA isolation). In parallel, virions were recovered from cells using TRIzol® Reagent (Invitrogen). Briefly, adherent cells were

lysed with 0.2 mL of TRIzol® Reagent, and RNA were extracted with chloroform and isopropanol (100% v/v) and were then washed with ethanol (75% v/v). To eliminate any remaining template DNA that could contaminate RNA, samples were treated with DNase I, RNase-free endonuclease (Thermo Scientific) that digests single- and double stranded DNA. Briefly, 1U of DNase I was mixed to each µg of RNA with reaction buffer containing $MgCl_2$ and samples were incubated for 30 min at 37°C. Then, to inactivate DNase I, samples were incubated for 10 min at 65°C with EDTA, as recommended by the manufacturer and RNA was column-purified using NucleoSpin® RNA Clean-up XS kit (MACHEREY-NAGEL).

8.2 RT-PCR amplification and Sanger sequencing

RNA was reverse-transcribed with Accuscript High Fidelity Reverse Transcriptase (Agilent Technologies) and strand-specific primers; the RT was performed with the forward primer (5'-ATTACTCTCTGTGCACTGTCTG-3') in order to selectively amplify (-)-strands RNA, which might be synthesized by the RdRp of the virus. In order to correct for T7-induced mutations during the primary transcription of the cDNA to RNA, a RT was also performed with the reverse primer (3'-CAGTGTAGAAGAGGCTGTTGAA-5') allowing the selective amplification of (+)-strands RNA. The reverse transcription conditions used were 42°C 60 min, 70°C 15 min. The cDNA was PCR-amplified with Phusion High Fidelity DNA polymerase (New England Biolabs) and specific primers (forward primer 5'-AAATGTATGTCCCAGGATGGC-3' and reverse primer 3'-

CCCAATCATTTCCCTGTCAT-5') to amplify the full VP1 gene of Norwalk virus (GenBank accession number NC001959). The PCR conditions used were 98°C 30 sec, 35 cycles of 98°C 10 sec, 59°C 30 sec, 72°C 1 min, and a final extension step of 72°C 10 min. Purified PCR products were cloned and colony-PCR were used for sequencing (see sections 4.3 and 5.2 in "Materials and Methods").

8.3. Amplification of hypermutated sequences and 454 high-throughput sequencing

The primary PCR of each transfection samples was used to perform a secondary PCR that increased sensitivity and specificity using degenerated primers. In the forward primer, T nucleotides are degenerated for Y nucleotides to amplify A-to-G hypermutations, whereas in the reverse primer A nucleotides are degenerated for R nucleotides and amplify specifically T-to-C hypermutations (Suspène et al., 2008). Two distinct regions within the VP1 gene were amplified: region 1 composed of 266 bp in length, using forward 5'-GACGCyACAyCAAGCGyGG-3' and reverse 3'-CTCrTGTTrCCrrCCCrrCC-5'primers, and region 2 composed of 405 bp in length, with forward 5'-CyCyGyCyAACyCACGyGCC-3' and reverse 3-CrrCTTGGGrGCCrGrCGG-5' primers. Purified PCR products were used for molecular cloning and massive sequencing was performed with the 454 GS Junior machine (Genomic Core Facility S.C.S.I.E., University of Valencia) that generates reads of approx. 400 bp in length.

Reads were cleaned by removing adapters as well as primer sequences and low quality regions and reads shorter than

100 bp were removed with the PRINSEQ software. Pair-wise alignments between libraries and the NV reference sequence were performed using 454 Sequencing System Software (Roche). Only reads showing at least one SNP (excluding indels) compared to the reference sequence were used for the mapping with BWA software (http://bio-bwa.sourceforge.net/). SAM files were converted into aligned FASTA files. For each read, the type and the number of SNPs were reported using bioinformatic tools and home-made scripts.

9. Statistics

Statistical analyses were performed using SPSS Statistics (V.20) (http://www-01.ibm.com) and R (v3.1.2) softwares (http://www.r-project.org).

RESULTS AND DISCUSSION

Chapter 1. Variation in RNA virus mutation rates across host cells

Although host factors might affect viral mutation rates, nearly all estimates for animal viruses have been obtained in standard laboratory cell lines, which are usually immortalized or cancerous and thus show aberrant mitotic/metabolic and gene expression patterns. Furthermore, all viral mutation rate studies have been conducted under atmospheric oxygen levels that are substantially higer than those found in most tissues (Carreau et al., 2011), but the impact of this type of environmental stress on the estimates is unknown.

Here, we investigated the impact of the cell type and host species on the variability in the mutation rate of VSV. Since in nature this virus infects a very large number of mammals, wild animals but also insects, it replicates in very different cellular environments and thus represented a good experimental model to investigate the impact of this cellular heterogeneity.

1.1 Fluctuation tests in BHK-21 cells and validation by molecular clone sequencing

To study the variability in the mutation rate of VSV across host cells, we performed Luria-Delbrück fluctuation tests. To score mutants, we used a monoclonal antibody against the envelope glycoprotein G and determined the probability of

appearance of monoclonal antibody resistant (MAR) mutants in independent cultures (null-class method). To set up the method, we first performed six independent tests in baby hamster kidney cells (BHK-21), for which we had previous estimates (Furió et al., 2005). The results of the fluctuation test gave an average mutation rate to the MAR phenotype of $m = (1.64 \pm 0.27) \times 10^{-5}$ per round of copying (Table 3). This rate was then converted to per-nucleotide units as $\mu = 3m/T$, where the mutational target T was determined by sequencing the glycoprotein G gene from 15 MAR plaques (see section 4.3 in "Materials and Methods"). We identified four different nucleotide substitutions conferring the MAR phenotype, which led to amino acid changes D257N, D259A, D259N, and S273T. Since previous work reported the same substitutions at position 259 of the G glycoprotein, in addition to substitutions D257G, D257V, D257Y and A263E (Holland et al., 1991), we used $T = 8$ and found a mutation rate of VSV in BHK-21 cells of $\mu = 6.15 \times 10^{-6}$ substitutions per nucleotide per round of copying (s/n/r).

Table 3. Fluctuation tests of VSV in BHK-21 cells

	Test 1	**Test 2**	**Test 3**	**Test 4**	**Test 5**	**Test 6**
N_i (PFU)	160 ± 11	267 ± 18	358 ± 114	355 ± 35	293 ± 36	290 ± 10
N_f (PFU)	24562 ± 1021	10875 ± 956	20375 ± 849	38150 ± 1590	46157 ± 1044	17200 ± 934
Total cultures	24	24	24	24	24	24
With no MAR	16	20	17	16	16	15
With 1 MAR	6	4	6	7	7	4
With 2 MARs	0	0	1	1	0	2
With >2 MARs	2	0	0	0	1	3
P_0[a]	0.667	0.833	0.708	0.667	0.667	0.625
m[b]	1.66×10^{-5}	1.71×10^{-5}	1.72×10^{-5}	1.07×10^{-5}	0.88×10^{-5}	2.78×10^{-5}

[a]: Fraction with no MAR
[b]: Mutation rate

In order to ascertain the reliability of the estimate obtained by the Luria-Delbrück fluctuation test, we used a molecular clone sequencing approach. This allowed us to score mutations more directly than in fluctuation tests and to analyze a wider genome region, although the interpretation of the data is complicated by the fact that the observed mutation frequency is dependent on selection or the viral replication mode (i.e. number of rounds of genome copying). Using limiting dilutions (see section 5.1 in "Materials and Methods"), BHK-21 cells were infected with a single infectious unit and the resulting viral bursts (i.e. 1.55×10^7 final PFU on average) were used to sequence three genome regions mapping to genes P, G, and L. Among the 77,500 bases analyzed, we found four single-nucleotide substitutions, giving a mutation frequency of $f = 5.16 \times 10^{-5}$ (Table 4). Assuming a per-cell burst size of $B = 1250$ (Furió et al., 2005), the number of infection cycles (i.e. viral generations) elapsed were estimated to be $c = \frac{\ln 1.55 \times 10^7}{\ln 1250} = 2.3$. Therefore, the per-generation increase in mutation frequency was $f / c = 2.24 \times 10^{-5}$. Then, to account for the effect of selection, we used the previously characterized distribution of mutational fitness effects (see section 5.3 in "Materials and Methods"). Based on this distribution, the expected fraction of observable mutations after 2.3 generations was 53% and, thus, the estimated per-cell mutation rate was $\mu_c = \frac{f}{c\alpha} = 4.23 \times 10^{-5}$. Although the exact number of round of genome copying per cell was unknown, a previous work suggested $r_C = 5.8$ rounds/cell, implying that $\mu = 7.30 \times 10^{-6}$ s/n/r. Therefore, since this estimate was fully consistent with the above results of the

Luria-Delbrück fluctuation test, subsequent experiments in mammalian and insect cell lines were done using fluctuation tests only because they provided a faster and simpler approach.

Table 4. Molecular clone sequencing of VSV from BHK-21 cells.

Gene	**P**	**G**	**L**
Genome sites	1339-1899	3858-4347	6974-7462
Clones	50	50	50
Total bases	28050	25000	24450
Mutations	A1821C (Lys →Thr) G1640A (Gly→Arg)	A3983G (Glu →Glu) T3937A (Leu→His)	None

1.2 Constant mutation rate of VSV in different mammalian cells

Altough VSV replicates in widely different intra-cellular and host environments, mutation rates published to data have been measured in BHK-21 cells only (Sanjuán et al., 2010; Furió et al., 2005; Holland et al, 1989), which are immortalized or tumoral as opposed to those typically encountered by the virus *in vivo*. Furthermore, VSV has a tropism for neural cells, and kidney fibroblasts are thus not a natural target of this virus.

To address the potential effect of cell immortalization on the viral mutation rate, we performed fluctuation tests in primary mouse embryonic fibroblasts (MEFs) and isogenic, p53 knock-out,

MEFs. The average rates were similar in both cell types, since we found $m = 1.27 \times 10^{-5}$ in normal MEFs and $m = 0.82 \times 10^{-5}$ in p53 knock-out MEFs, thus revealing no significant effect of cellular immortalization (Figure 14; Student's t-test: $p = 0.232$, $n = 6$; Supplementary 1). However, many tumoral cell lines show other genetic and metabolic alterations in addition to p53 inactivation. To check the potential effects of these changes, we performed fluctuation tests in CT26 cells from an undifferentiated grade IV colon adenocarcinoma of a BALB/c mouse (Corbett et al., 1975). We found $m = 1.18 \times 10^{-5}$, a value not significantly different from that of primary MEFs (Student's t-test: $p = 0.885$, $n = 6$). Of note, BHK-21 are also tumor-forming cells, and the mutation rate in these cells was similar to the rate observed in MEFs or CT26 cells (one-way ANOVA: $p = 0.293$, $n = 12$). Because metabolic and mitotic activity should alter the availability of intra-cellular dNTPs (Bray and Brent, 1972) and could potentially impact RNA replication fidelity, this homogeneity in the VSV mutation rate was not an obvious expectation (although this virus replicates in the cytoplasm and may not be strongly affected by these alterations). This finding has implications for the field of oncolytic virotherapy (Russell et al., 2012), since it is critical to assess the genetic stability of these therapeutic viruses during large-scale manufacturing and clinical use. In particular, CT26 cells have been used in mice as a model for testing the oncolytic activity of VSV (Breitbach et al., 2011).

Although these results suggested that VSV replicated with similar fidelity in different cell types, we sought to test whether

this would also hold for neural cells. We therefore performed fluctuation tests in Neuro-2a cells from a mouse neuroblastoma (Klebe and Ruddle, 1969). Again, we found an average mutation rate of $m = 1.06 \times 10^{-5}$, which did not significantly differ from the rate obtained in BHK-21 cells (Student's t-test: p = 0.461, n = 9).

All viral mutation rate estimates available to date have been conducted *in vitro* under atmospheric oxygen levels (i.e. 20% O_2). However, these levels are substantially higher than those found *in vivo* in most tissues, which range between 11% and 1% O_2 (Carreau et al., 2011). For instance, oxygen delivery is dependent on the metabolic requirements and functional status of each organ, and hypoxia (1% oxygen) is marked during cancer formation. Therefore, to test for the impact of this type of environmental stress on viral mutation rates, we performed fluctuation tests in BHK-21 cells varying oxygen levels. The VSV mutation rate in BHK-21 cells cultured under hypoxic conditions was $m = 2.71 \times 10^{-5}$, thus slightly higher but not significantly different from the rate obtained under standard conditions (Student's t-test: p = 0.122, n = 9). Although oxidative stress should lead to the release of reactive oxygen species (ROS), which have been previously shown to be mutagenic for hepatitis C virus (Seronello et al., 2011), our results suggested that VSV was not sensitive to oxidation levels. We proposed that this might be related to the fact that the nucleocapsid of mononegavirales forms a tunnel-like structure which wraps the viral genomic RNA and remains assembled during the entire infection cycle (Ge et al.,

2010; Green et al., 2006), effectively isolating the viral RNA (Ostertag et al., 2007).

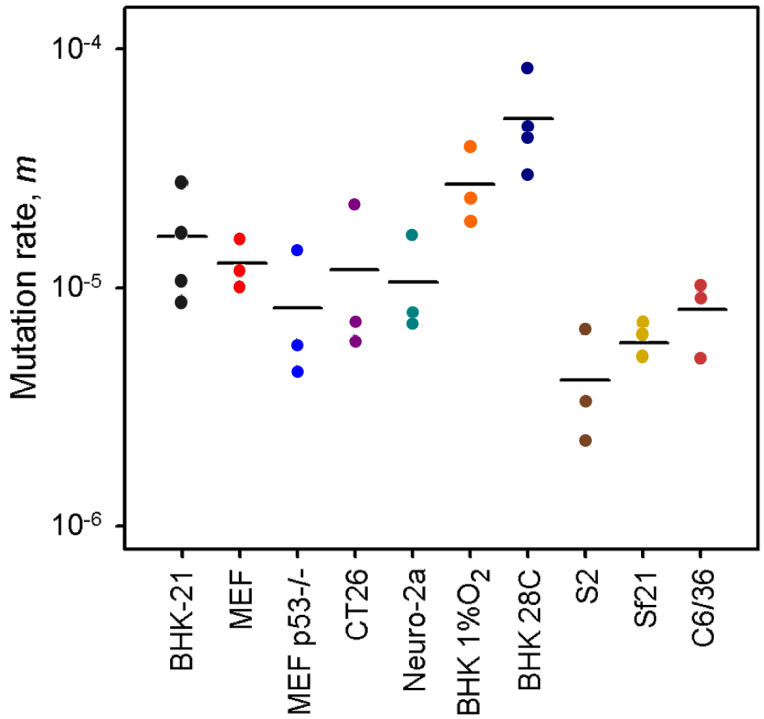


Figure 14. VSV mutation rate to the MAR phenotype estimated by the Luria-Delbrück fluctuation test in different cellular environments. Each dot represents an independent estimate (n = 3 for all except n = 6 for BHK-21 and n = 4 for BHK-21 at 28°C). Horizontal bars indicate the mean rate. Detailed information for each test is provided in Supplementary 1. Adapted from Combe and Sanjuán (2014b).

1.3 Lower mutation rate in insect cells

Since the natural lifecycle of VSV alternates between mammalian and insect hosts, we sought to estimate the viral

mutation rate in insect cells. First, we measured the mutation rate of VSV in S2 cells from *Drosophila melanogaster* embryos. The average estimate from three independent fluctuation tests was $m = 4.08 \times 10^{-6}$, representing a fourfold decrease compared with BHK-21 cells (Figure 14, Student's t-test: $p = 0.009$, $n = 9$). Then, to further investigate this potential decrease in viral mutation rate in insect cells, we selected two additional insect cell lines, such as sf21 ovarian cells from the moth *Spodoptera frugiperda*, and C6/36 cells from *Aedes albopictus* mosquito larvae and found mutation rates of $m = 5.9 \times 10^{-6}$ and $m = 8.14 \times 10^{-6}$, respectively (Supplementary 1). Also, since insect cells were infected at 28°C and mammalian cells at 37°C, we performed four additional tests in BHK-21 at 28°C to control for the effect of the incubation temperature. In order to test for the effects of host type and temperature (fixed factors), we used estimates obtained in mammalian (BHK-21, BHK-21 at 28°C, MEF, MEF p53-/-, CT26, and Neuro-2a) and insect/mosquito cells (S2, Sf-21, and C6/36) and performed a two-way ANOVA in which the specific cell line was treated as a random factor nested within host type. The results confirmed that VSV showed a lower mutation rate in insect cells compared to mammalian cells (ANOVA: $p < 0.001$), and also that temperature did not account for this result because the estimates in BHK-21 cells were actually higher at 28°C than at 37°C ($p = 0.001$). In $\log_{10}$-transformed data (used for all statistical comparisons), the estimated effect size of the host type in the above model was 0.590 ± 0.205, which corresponds to a 3.9-fold mutation rate decrease in insect cells.

We also observed lower viral yields in C6/36 mosquito cells (Supplementary 1), confirming previous observations (Schloemer and Wagner, 1975). In a similar study, chikungunya (CHIKV) and West Nile (WNV) virus variants with lower mutation frequencies showed replication defects (reduced viral yield) in mosquito cells compared to mammalian cells (Rozen-Gagnon et al., 2014). However, in agreement with our results, all studies discarded the effect of the incubation temperature as a possible explanation for the differences in virus production between mammalian and insect cells. However, we wondered whether the observed differences may result from differences in the sensitivity to detect MAR mutants in mammalian and insect cells. To address this potential bias, we first verified that MAR plating efficiency was similar in regardless of viruses were grown in BHK-21, S2, Sf21, and C6/36 (Supplementary 1). Then, we tested for differences in the mutation target size (*T*) by sampling 15 individual MAR plaques from fluctuation tests performed in S2 cells and sequenced the region of the G protein controlling this phenotype (see section 4.3 in "Materials and Methods"). We found the same amino acid replacements as in fluctuation tests performed in BHK-21 cells (D257N, D259N, S273T, see section 1.1 above) except for D259A. However, because the D259 mutant is viable in insect cells (Novella et al., 1999), failure to detect it was probably just due to insufficient sampling depth. We also found substitution A263E, which was reported previously in BHK-21 cells (Holland et al., 1991). Since viruses grown in S2 and BHK-21 cells shared similar mutational repertoires and plating efficiencies

our results supported the reliability of the observed mutation rate differences between insect and mammalian cells.

1.4 Comparison of evolutionary rates between arboviruses and directly transmitted viruses

Interestingly, since arboviruses such as VSV have to replicate in disparate hosts, they are subjected to different selective pressures (Coffey et al., 2011) and tend to evolve more slowly than directly transmitted viruses due to these evolutionary constraints (Hanada et al., 2004; Jenkins et al., 2002; Rodriguez et al., 1996). For instance, for alphaviruses, it has been shown that viral diversity is more restricted in the insect host because of more stringent population bottlenecks and selective pressures (Rozen-Cagnon et al., 2014). Here we performed a meta-analysis using 170 previously published evolutionary rates which confirmed that, after accounting for phylogenetic relatedness and the timespan of sequence sampling, arboviruses showed a significantly lower evolution rate than directly transmitted viruses (Figure 15; two-way ANOVA: $p = 0.006$), the geometric mean rates being 5.7×10^{-4} substitutions per site per year (s/s/y) and 1.3×10^{-3} s/s/y, respectively. Lower evolution rates in arboviruses have been often interpreted in terms of fitness trade-offs, whereby neutral or beneficial mutations in mammals can be deleterious in insects, and *vice versa*, thus restricting viral evolution. Indeed, previous studies where one host was artificially removed from the alternate cycling revealed that host alternation constrains the adapatibility of arboviruses (Coffey et al., 2011). Our results provide an alternative explanation for the slower

evolution of arboviruses, based on lower mutation rates in the insect host compared to mammalian host.

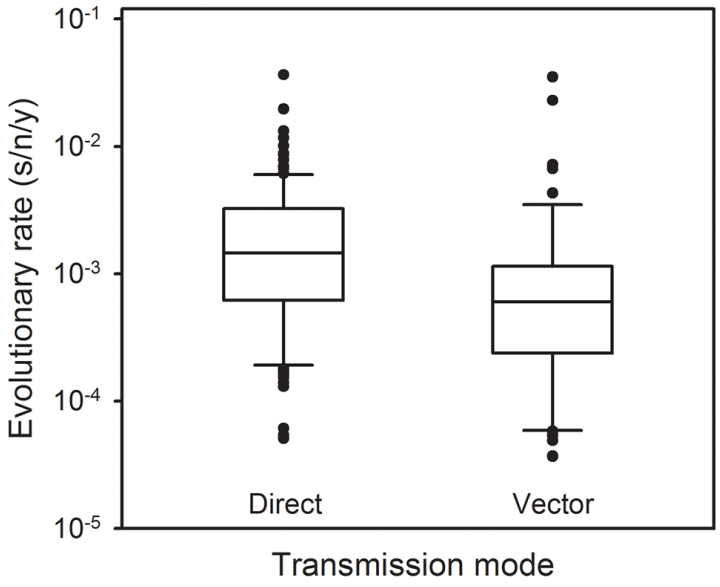


Figure 15. Molecular evolution of directly transmitted and arthropod-transmitted riboviruses. Data were collected from the supplementary information of a previous meta-analysis (Sanjuán, 2012) and include 170 evolutionary rates, 113 for directly transmitted viruses and 57 for arboviruses. The box plot indicates the median (central lines), percentiles 25/75 (box) and percentiles 10/90 (bars), and outliers (dots). From Combe and Sanjuán (2014b).

Chapter 2. Genetic diversity of an RNA virus at the single-cell level

It has been traditionally assumed that cell cultures represent homogeneous environments. However, it has been recently shown that cells exhibit important variability in their metabolism and physiological status, as well as in gene expression patterns (Yoshimoto et al., 2013; Kurimoto et al., 2007; Wang and Bodovitz, 2010; Narsinh et al., 2011). These cell-to-cell variations might have major implications for the outcome of viral infection (Pathak and Temin, 1990; Monk et al., 1992). For instance, analysis of individual cells helped to clarify how viruses antagonize the interferon response, which has a stochastic nature and varies from cell to cell (Rand et al., 2014). For animal viruses, the study of the differences in viral yield between single cells has provided important clues to viral lifecycle aspects such as the intra-cellular determinants of viral fitness, or the viral replication mode (Schulte and Andino, 2014; Zhu et al., 2009; Schulte et al., 2015; Timm and Yin, 2012). In addition to determining host-to-host transmission, viral replication also governs the production of genetic diversity through the appearance of spontaneous mutations. However, the rate at which mutations accumulate in viral populations should also be variable (Schubert et al. 1984, Ninio, 1991, Drake et al., 2005), as demonstrated for Spleen Necrosis Virus (SNV) where a single provirus had a mutation rate 1,000-fold higher than the rate determined for the virus population because of the presence of multiple mutations in its genome (Pathak and Temin, 1990). Although there are

fundamental gaps in our knowledge of the mutational process, characterizing such processes underlying the genetic diversity of viruses is critical in order to understand their evolution, as well as immune escape, vaccine failure, drug resistance, or disease emergence (Duffy et al., 2008; Elde et al., 2012; Holmes, 2009; Lauring et al., 2013).

Recent advances in deep sequencing have improved our knowledge of viral genetic diversity, but these studies did not reach the single-cell level (Acevedo et al., 2014). However, single-cell analyses are necessary to clarify whether the genetic diversity is produced homogeneously among cells, or tends to be allocated to a subset of cells, as previously proposed (García-Villada and Drake, 2012). Here, by combining single-cell micromanipulation with deep sequencing, we measured the accumulation of SNPs for hundreds of full-length VSV plaques released from individual cells.

2.1 Infected cells and MOI

Following infection of confluent BHK-21 cells, micromanipulation was used to isolate 350 single cells and transfer them to separate wells within 2 h post inoculation (hpi), such that no viral progeny could be released before cell isolation (Timm and Yin, 2012). After 24 h incubation, culture media were plated to measure the viral yield per cell by enumeration of PFUs. Since, theorically, approx. 5×10^5 PFU/ml were used to infect cell cultures that were then diluted 1/20, we estimated that the proportion of free particles that might have been aspirated with the cell (i.e. 2-4 µl of media aspirated) should be around 2.5-5

PFUs. This suggested that the fraction of free PFUs mixed to the progeny of an infected cell accounted for a very small proportion of the plaques that were sequenced. In addition, a pilot experiment allowed us to establish an arbitrarily threshold to discriminate the proportion of infected- and non-infected cells since, after plaque assay of the supernatant, wells contained either less than 7 PFUs (between 0-6) or more than 34 PFUs (Figure 16). Therefore, when less than 10 PFUs per well were observed, the cell was considered as non-infected and was discarded from the analysis, whereas when more than 10 PFUs per well were scored the cell was considered as infected.

Among the 350 single cells isolated we found 95 infected cells, which represents 27% of the total. Therefore, according to the Poisson distribution, we calculated the MOI as follows: – ln (1-95/350) = 0.32 PFU/cell, such that 85% of the cells should be infected by a single infectious particle, whereas 15% by two or more units. Although 95 cells were productively infected, we used plaques from 90 cells to construct libraries because plaques from the remaining 5 cells did not have enough titer to allow for their equimolar pooling.

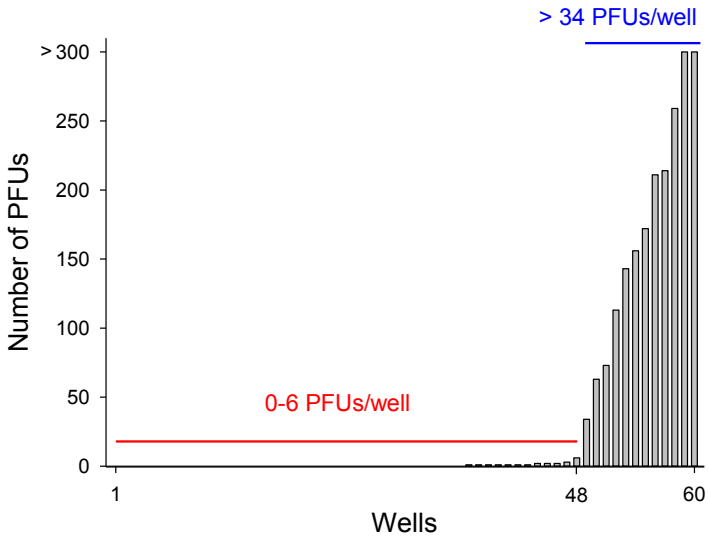


Figure 16. Number of PFUs per well. In a pilot experiment, 60 wells were inoculated with a single infected cell each. After incubation, the supernatant was used for plaque assays and we scored 48 wells showing between 0-6 PFUs, whereas 12 wells had > 34 PFUs.

2.2 VSV mutational spectrum

To analyse the genetic diversity of viruses released from single-cells, we picked 7-10 plaques for each of these 90 cells, such that a total of 881 full-length individual plaques (approx. 11.2 kb each), corresponding to 9 Mb sequenced by the SOLID technology. Since a previous study showed that viruses with up to 90% reduction in fitness, compared to the wild-type VSV used here, can form visible plaques (Sanjuán et al., 2004), a priori our experimental protocol allowed the sampling of most of the non-lethal mutations present in these 90 single-cells. Mapping of the sequences of these 881 plaques to a common reference sequence gave a total of 285 SNPs (Figure 17, Supplementary 5).

Of these 285 SNPs, 264 were coding mutations with 155 synonymous substitutions and 109 nonsynonymous substitutions, including 13 stop codons (Table 5).

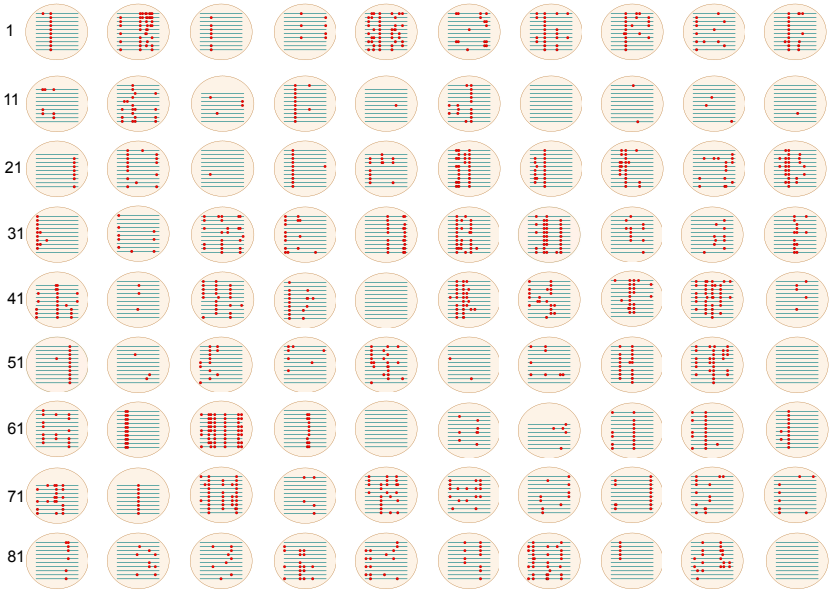


Figure 17. SNPs found in the 90 single-cells examined. Each circle represents a cell and green lines represent 7-10 individual plaques. Red dots indicate the nucleotide position of each SNP on the plaque where it occurred. When a SNP was found in several plaques, a dot is indicated on each. Among the 90 single-cells analyzed, 85 cells showed at least one SNP, whereas 5 cells have no SNP. Adapted from Combe et al. (submitted).

Table 5. VSV mutational spectrum at the single-cell level. For each type of mutation the counts are reported, as well as the nucleotide substitution matrix.

Mutation type	Count	SNPs type				
Total mutations	285		A	U	G	C
Base substitutions	285	A	-	5	62	25
Indels	0	U	11	-	20	44
Transitions	180	G	41	17	-	0
Transversions	105	C	26	33	1	-
Coding	264					
Non-coding	21					
Synonymous	155					
Nonsynonymous	109					
Stop-codon	13					

The mutational spectrum was dominated by transitions, representing 63.2% of all SNPs, against 36.8% of transversions. Finally, among transversions the nucleotide changes A -> C (23.8%), C -> A (24.8%) and their respective complementary U -> G (19%) and G -> U (16.2%) were significantly more frequent than A -> U (4.8%), U -> A (10.5%), C -> G (0.9%) and G -> C (0%) changes (Fisher's exact test, $p = 2.5 \times 10^{-9}$). These nucleotide changes were consistent with the mutational spectrum found in sequences from natural isolates of VSV (Figure 18, Spearman's correlation, $\rho = 0.918$, $P < 0.001$), suggesting thus their relevance to natural diversity.

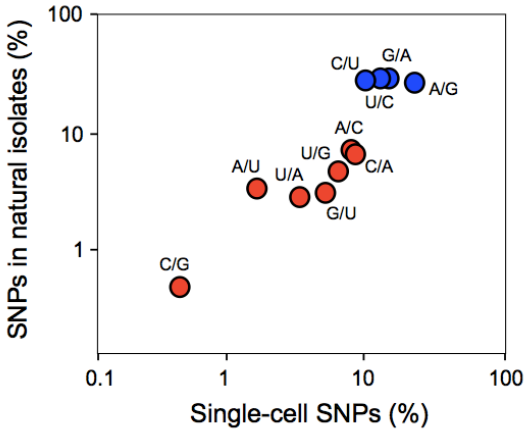


Figure 18. Correlation between SNPs spectra in viruses released from single-cells and in sequences from natural isolates of VSV. The frequency of each type of SNP in single-cells was taken from Table 5. Blue dots indicate transitions and red dots transversions. For natural isolates of VSV, 1033 SNPs were extracted from the following available genome sequences of the Indiana VSV serotype: EU849003 (Mudd-Summers strain) which was used as reference, AF473865 (isolated from cattle in Colombia, 1985), AF473866 (isolated from cattle in Guatemala, 1994), AF473864 (isolated from horse in Colorado, 1998), EF197793 (unknown host), J02428 (unknown host), and NC_001560 (unknown host). The dots indicate the percentage of total SNPs of each type (A -> G, U -> C, and so on), in log-scale. The G -> C data point does not appear because it showed 0% frequency in single cells; its abundance in natural isolates was 0.4%. Adapted from Combe et al. (submitted).

2.3 SNP distribution between and within single-cells

Of the 90 cells examined, 85 contained at least one SNP whereas 5 cells had no SNP. The number of SNPs varied between 0-11 depending on the cell (Figure 19A), with on average 4.3 SNPs per cell (including SNPs found in multiple-copies as a single mutation event). Despite this ample variation in the number of SNPs between cells, we could not discard the null hypothesis that

the number of mutations per cell followed a Poisson distibution (p-value after applying the Dunn-Sidâk correction for multiple tests, $p > 5.7 \times 10^{-4}$).

The copy number of any given SNP within a cell varied also. Whereas 90 SNPs occurred in only 1 plaque (scSNPs), 195 SNPs were scored in multiple-copy (Figure 19B). The latter scenario indicates intra-cellular amplification of SNPs as a result of viral replication. These multiple-copy SNPs probably occured early during viral replication and were then copied to more progeny molecules under a geometric replication model. Although this latter strategy of replication has been recently demonstrated for poliovirus (Schulte et al., 2015) we were not able to test this scenario, as shown below.

Finally, the number of SNPs per individual plaque ranged between 0-9, with 1404 SNP copies scored among the 881 plaques sequenced, leading to an average of 1.59 SNPs per plaque (Figure 19C).

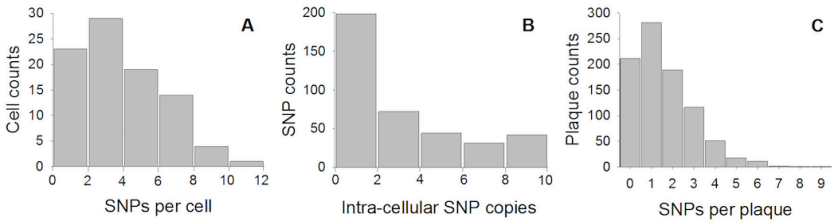


Figure 19. Distribution of the number of SNPs between and within single-cells. A, number of SNPs per cell; B, number of intra-cellular SNP copies for each SNP; C, number of SNPs per plaque. Adapted from Combe et al. (submitted).

2.4 Pre-existing variability

Interestingly, of the 285 SNPs 62 were observed in several (2-12) cells (Supplementary 5). Because it is unlikely that the exact same point mutation appeared *de novo* and simultaneously in multiple cells, we attributed these 62 SNPs to pre-existing variability (preSNPs). Consequently, these mutations were obviously present in the viral stock and thus delivered to cells during the infection process. Then, because 85% of the cells might have been infected by a single PFU (see section 2.1 above), one could thought that each cell should contained only one of these preSNPs. In contrast, among the 64 cells with preSNPs, 44 cells had more than 2 preSNPs, with an average of 2.56 preSNPs per cell. Because of this result, we hypothesized that single PFU could have delivered pools of genetically diverse infectious particles to each single-cell.

2.4.1 Spatial aggregation of SNPs within the same PFUs

To test the above hypothesis, we compared the observed distribution of preSNPs for all single-cells isolated (350 single-cells) under a Poisson and under a negative binomial model. The Poisson model assumes random spatial distribution of infectious particles, whereas the negative binomial model accommodates putative aggregation of virions within PFUs. The Poisson distribution was simulated as $Po(x|m) = \frac{e^{-m}\, m^x}{x!}$, where x is the observed number of SNPs per cell and m is the average number of preSNPs per cell. The negative binomial model was simulated as

$NBi(x|p,r) = \frac{\Gamma(x+r)}{\Gamma(r)x!}p^x(1-p)^r$, where Γ denotes the Gamma function, p is the probability of getting a preSNP if these were homogeneously distributed across cells and r is the average number of cells that have to be examined before obtaining a preSNP. Therefore r measures the degree of preSNP aggregation. For each distribution, the observed number of preSNPs per cell was used to fit the model by maximum likelihood. Similarly, we found that the number of preSNPs per isolated cell showed a significant over-dispersion compared to the Poisson distribution since the variance-to-mean ratio was 2.92 (Figure 20), thus rejecting a scenario of a Poisson distribution of preSNPs (Kolmogorov-Smirnov test, $p < 0.001$). Therefore, a negative binomial distribution fitted better our results, suggesting spatial aggregation of the number of preSNPs per isolated cells.

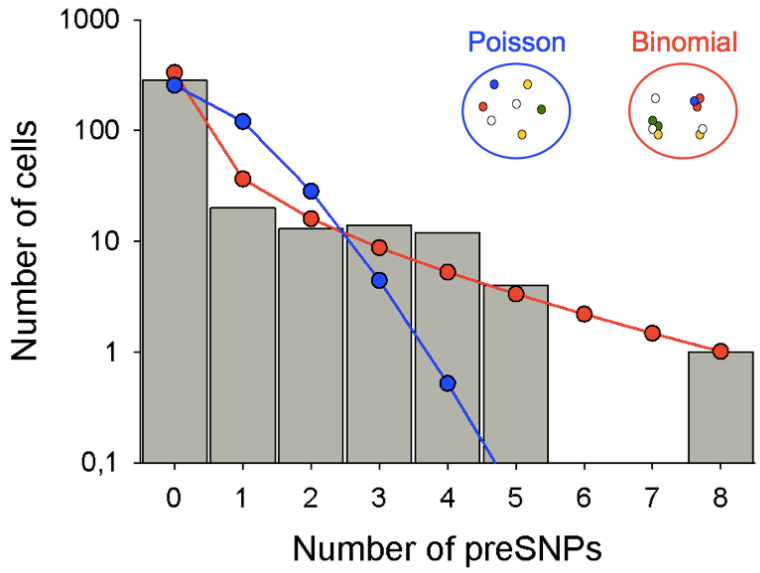


Figure 20. Distribution of the number of preSNPs per isolated single-cell. The distribution of the 62 preSNPs is represented for the 350 single-cells isolated (log-scale). The observed distribution is compated to two statistical models: a Poisson (blue) model that assumed random spatial distribution of infectious particles, and a negative binomial (red) model that accommodates putative spatial aggregation of infectious particles. Both circles above the histogram represent the assumption underlying each model. The Poisson model was rejected by a Kolmogorov-Smirnov test ($p < 0.001$), whereas the binomial model fitted better our results (log-likelihood: - log L = 389.84 for Poisson, -log L = 281.16 for negative binomial). Adapted from Combe et al. (submitted).

The spatial aggregation of viral particles within the same infectious unit is supported by microscopy studies showing that individual PFUs can be constituted by aggregates of virions (Bussereau and Flamand, 1978; Libersou et al., 2010). For Sindbis virus, it has been shown that virions are able to form aggregates at the cell surface within cytoplasmic vacuoles (Boehme et al., 2000). Also, plant viruses, such that Potato virus X (Cruz et al.,

1996), Tomato Golden Mosaic virus (Rushing et al., 1987) and Bamboo Mosaic virus (Lin and Chen, 1991), are able to form large and fibrillar aggregates of virions within infected cells. Additionally, a recent study showed that enteroviruses are transmitted from cell-to-cell inside lipid vesicles harboring multiple viral particles (Chen et al., 2015). Therefore, despite the low MOI used to infect the cultures, single cells were not infected by a single infectious particle, but rather by a pool of multiple, genetically non-identical infectious particles contained within the same PFU. Also, our results provided a quantitative estimation of the diversity of these multiple infectious particles since we found an average of 2.56 alleles delivered to each cell. However, this estimate was based on SNPs occurring in multiple cells only, and thus the actual number might be higher because we could not discard that some of the SNPs occurring in only one cell were also pre-existing.

2.4.2 Intra-PFU genetic complementation

During transmission between their hosts, viruses experience strong population bottlenecks where the genetic diversity is lost. Such bottlenecks have been thought to have negative consequences on viral fitness, but this view can be questioned if even the smallest infectious unit (i.e. PFU) can deliver viral genetic diversity to individual cells. Also, after co-infection of a cell with two or more viruses, viral products can be shared, thus allowing the genetic complementation of deleterious alleles (Novella et al., 2004). For Dengue virus it has been shown that this process can persist *in viv*o for multiple generations

(Aaskov et al., 2006). Here, we found that 9 of the 285 SNPs produced a premature stop codon in the matrix protein (M), the surface glycoprotein (G) and the polymerase protein (L), thus being obviously lethal for the virus (Table 6), suggesting intra-PFU genetic complementation. Moreover, since 6 of these stop codons were found in multiple plaques (between 2-10 copies) within a cell, and 3 were scored in several cells (between 2-3 cells), they could not be attributed to sequencing bias.

Table 6. Premature stop codons found in single-cell SNPs. For each stop codon, the cell number and the number of copies within each cell where it occurred are indicated.

Mutation	Cells	Plaques	Protein	Codon change
U2330A	36, 46	10, 9	M	TAT -> TAA
U2330G	24, 62, 63	10, 9, 8	M	TAT -> TAG
G3414U	83	2	G	GAA -> TAA
U4552A	4	2	G	TTA -> TAA
U4965G	19, 71	1, 3	L	TTA -> TGA
C4971A	47	1	L	TCA -> TAA
G5834U	49	1	L	GGA -> TGA
G9290U	57	1	L	GGA -> TGA
U9682A	59	5	L	TAT -> TAA

Site-directed mutagenesis studies using infectious molecular clones of VSV and other RNA viruses showed that viruses are extremely sensitive to the introduction of single-

nucleotide substitutions, with up to 40% being effectively lethal (Sanjuán et al., 2004). Therefore, the fact that these stop codons were detected despite their lethal effect for the virus suggested that genetic trans-complementation occurred within infected cells. As a consequence, we proposed that the genetic diversity delivered to individual cells by multiple, genetically non-identical, infectious particles within the same PFU may confer greater robustness against the effects of deleterious mutations than it was previously thought. Related to this, it has been shown using poliovirus that the fitness of a particular viral genotype is determined by cooperative interactions established among genotypes, suggesting thus that viral fitness is a property of the entire viral population, rather than an individual property (Vignuzzi et al., 2006). Therefore, in agreement with previous microscopy studies, our results showed that infectious particles might allow the co-transmission of pools of variants to individual cells, allowing intra-cellular genetic complementation and conferring greater robustness to viral populations.

2.4.3 Co-transmission of SNPs within the same PFU

Since the above results suggested that multiple infectious particles might have been delivered to cells as part of the same PFU, we thought to perform an extra and complementary experiment to check this finding. To do so, BHK-21 cells were co-infected with the wild-type (WT) VSV virus and a monoclonal antibody-resistant (MAR) mutant (input ratio 1:1) at a MOI of 20 PFU/cell. After 24 h of incubation, the supernatant was plated on confluent BHK monolayers, 20 individual plaques (PFUs) were

picked, amplified and then used for a second plaque assay in the presence/absence of monoclonal antibody. If these PFUs were initially constituted by a single infectious particle, they should show clones of either full WT or fully MAR phenotype. However if variants co-transmitted within the same PFU, these plaques would show both phenotypes. In agreement with our finding, this experiment confirmed the co-transmission of multiple infectious particles since 10/20 PFUs showed a WT phenotype, 4/20 PFUs behaved as MAR phenotype, whereas 6/20 PFUs showed both phenotypes (Figure 21).

By examinating the sequence of several clones (between 15-19 clones) from 5 WT PFUs, we observed that 9/15 and 2/19 clones had the MAR-conferring mutations (Combe and Sanjuán, 2014b). Whilst the MAR mutations were present in these clones, they did not allow the growth of the virus in the presence of monoclonal antibody, suggesting that they behaved as recessive alleles. One explanation could be that they were neutralized by the WT sequence located within the same PFU. Therefore, these results suggested that the phenotype of a given mutation is determined collectively by a pool of variants co-transmitted whithin the same multi-particle infectious units.

Finally, we investigated if SNPs could be co-transmitted over time. Two PFUs showing both phenotypes (i.e. WT and MAR) were passaged (ratio 1:1) at low MOI (0.02 PFU/cell) in BHK-21 cells. The plaque purification and plaque assays were repeated. Here again we observed that individual PFUs showed both MAR

and WT phenotypes, indicating that SNPs co-transmisison could last for several generations.

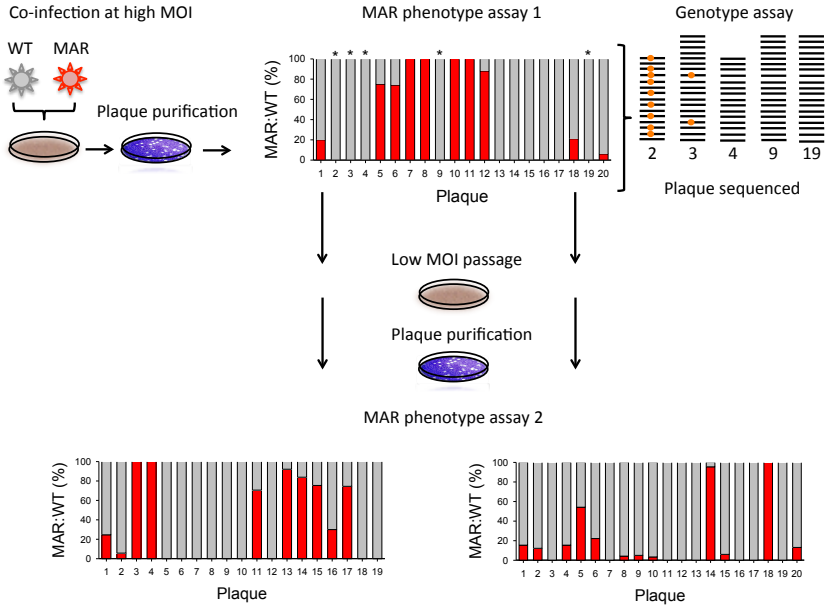


Figure 21. Co-transmission of MAR-mutant and wild-type VSV viruses within the same infectious particle. BHK-21 cells were infected with an input ratio 1:1 of MAR:WT viruses, at a MOI of 20 PFU/cell that allowed co-infection. After release of viral progeny, 20 individual plaques were picked and tested for the MAR phenotype. On the plot, the proportion of MAR viruses within each PFU is shown in red whereas the proportion of WT viruses is represented in dark grey. Then, 15-19 clones from 5 WT phenotype plaques were sequenced. Clones with the MAR-conferring mutation, although they behaved recessively in the phenotype assay, are represented by red dots. In parallel, two plaques with a mixed phenotype were passaged in fresh cells, and the plaque purification and MAR phenotype assay were repeated. In both cases mixing of MAR and WT phenotypes was still evident after this transfer. Adapted from Combe et al. (submitted).

As a result of co-transmission within the same infectious units, we hypothesized that some SNPs may be found in the same cells as other SNPs, and thus spatially co-localize. To test this scenario, we focused on the 62 preSNPs and we observed that 12 pairs co-localized in two cells, whereas 3 pairs co-localized in three cells (substitutions A6757G/A7573G, U3894A/A7573G, G1790A/U3894A), and 1 pair co-localized in four cells (substitutions A2925C/C6712U, Figure 22). However, although these SNPs co-localized within the same cell, they were not systematically found together in all plaques sequenced. This observation suggested that the coupling of these SNPs within the same cell was not the result of genetic linkage, but rather supported their co-transmission within the same PFU. This co-transmission might have allowed particle reassortment during the release of new PFUs by each cell, thus explaining the observation of some preSNPs within the same plaque after sequencing. Indeed, in PFUs released from each cell SNPs might belong to different virion sequences contained in each PFU (Figure 23). Despite the fact that virion reassortment or recombination within PFU should have resulted in the same number of SNPs per plaque, *Mononegavirales* such as VSV show extremely low recombination rates (Han and Worobey, 2011), and thus the possibility that virion recombination within PFUs occurred was excluded.

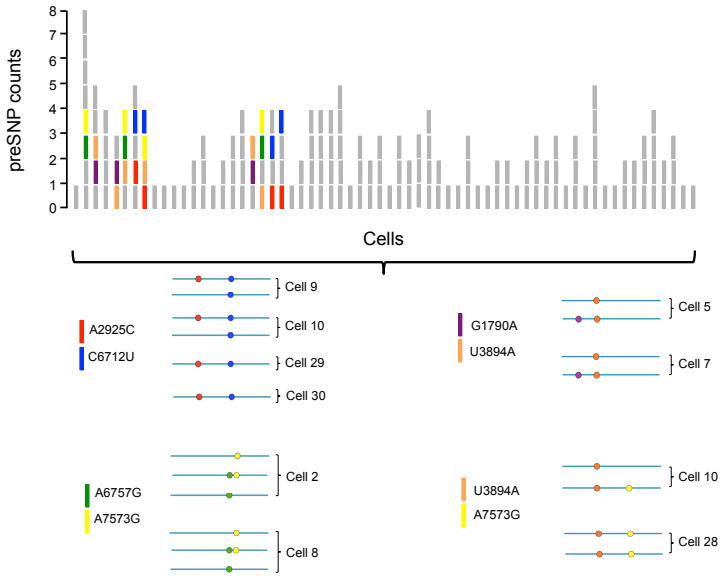


Figure 22. Number of preSNPs found in each individual cells. Panel above: each rectangle represents a different preSNP and each vertical pile of rectangles corresponds to a cell. The 64 cells showing at least one preSNP are graphically represented. Coloured rectangles represent preSNPs that co-locate in three (substitutions A6757G/A7573G, U3894A/A7573G, G1790A/U3894A) and four (substitutions A2925C/C6712U) cells. Panel below: representation of some plaques for cells containing co-locating preSNPs. Other SNPs found in these plaques are not represented here. Co-locating SNPs within a cell are not systematically found in the same individual plaque, ruling out their genetic linkage. Adapted from Combe et al. (submitted).

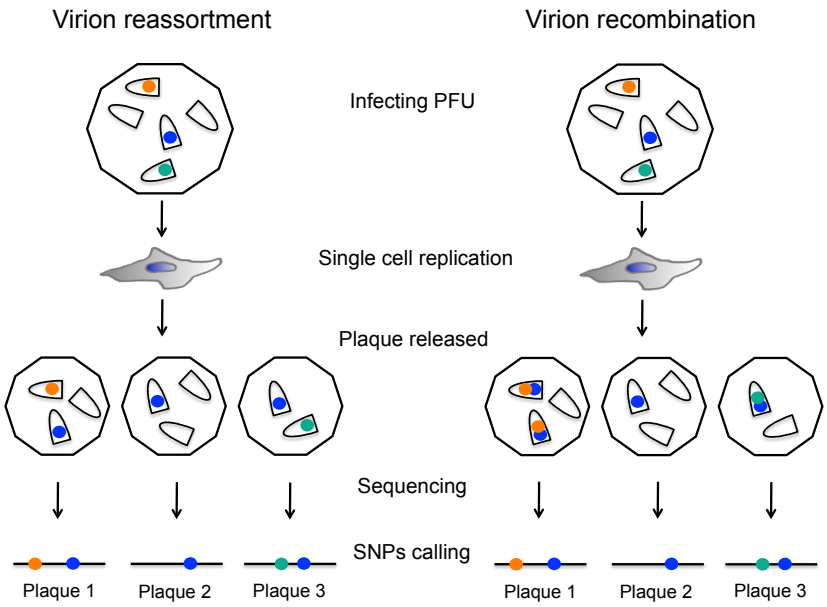


Figure 23. Virion reassortment *versus* recombination within PFUs. Within each infecting PFU, aggregates of infectious particles are represented. Coloured dots within each infectious particle represent variability, e.g. preSNPs. After replication, PFUs are released from each single cell as a new pool of variants able to infect another cell. A scenario of virion reassortment as well as virion recombination within PFUs is indicated. Then these PFUs, also called individual plaques, were sequenced and SNPs scored in each PFU were attributed to the corresponding plaque. Therefore, the final picture of the number of SNPs per plaque we got is identical if reassortment or recombination within PFUs occurred. However, since *Mononegavirales* such as VSV show very low recombination rates (Han and Worobey, 2011) we proposed a scenario of virion reassortment rather than recombination.

2.5 Production of spontaneous mutations in single-cells and per-cell burst size

Our approach allowed us to measure the *de novo* production of spontaneous mutations in 90 individual cells. To do so, we had to discard SNPs that were likely to pre-exist in the

inoculum, such as the 62 preSNPs found in multiple cells, but also SNPs found in single cells but as multiple intra-cellular copies. Indeed, since we were not able to determine if these multiple-copy SNPs came from *de novo* spontaneous mutation occurring during genome replication within a cell or were pre-existing and thus delivered to the cell by infectious particles, we removed both types of SNPs. Therefore, we focused on SNPs that occurred as single-copies only (scSNPs). Of the 90 cells analysed, we scored a total of 91 scSNPs, and, by computing the frequency of each scSNPs (between 7-10 plaques examined per cell), we found an average mutation frequency of 0.104 scSNPs per plaque per cell (Figure 24A). Then, by counting how many cells showed at least one scSNP, we determined the rate of spontaneous mutation per round of copying (U), which has already been demonstrated to be robust to selection acting on deleterious mutations and independent on the replication mode (Sanjuán et al., 2010). According to this method, U = - ln (P_0/N), where P_0 is the fraction of cells showing no scSNPs (40 out of 90 cells examined) and N the number of plaques sequenced per cell (between 7-10 depending on the cell). We found that the mutation rate ranged from U = 0.081 (10 plaques sequenced) to U = 0.116 (7 plaques sequenced). By considering the mutational target size (T = 11,161 bp), this gave $\mu = 7.26 \times 10^{-6}$ and $\mu = 1.04 \times 10^{-5}$ substitution per nucleotide per round of copying (s/n/r), respectively. This range is in good agreement with our previous estimates obtained by scoring the appearance of spontaneous MAR mutants using the Luria-Delbrück fluctuation test (see "Chapter 1"), therefore

confirming that scSNPs are convenient for the analysis of *de novo* spontaneous mutations.

The accumulation of spontaneous mutation within a cell depends on their probability of appearance, but also on the number of rounds of genome copying. It has been proposed that viruses grow exponentially within infected cells, leading to high viral yields (Cuevas et al., 2005). Also, although it has been suggested that most RNA viruses might replicate their genome *via* the "stamping-machine" model, in which progeny genomes are all synthesized from the same initital template (García-Villada and Drake, 2012; Chao et al., 2002; French and Stenger, 2003; Martinez et al., 2011; Safari and Roossinck, 2014), indirect evidence with VSV (Sanjuán et al., 2010) and poliovirus (Schulte et al., 2015) suggested multiple rounds of genome copying per cell, with stochastic variation from cell-to-cell. Therefore, although undergoing multiple rounds of genome copying per cell allows the production of higher viral yields, it comes at the cost of a greater load of mutations, suggesting thus that there might be a trade-off between the efficiency and the fidelity of viral replication at the single-cell level. To test for this trade-off, we determined the BS for each of the 90 cells analysed. We found that BS varied amply between cells and ranged from 195 to 5,840 with an average of 1,160 PFU/cell (Figure 24B). Moreover, BS was positively correlated to the frequency of scSNPs per genome per cell (Pearson correlation: r = 0.914, P = 0.011), suggesting that increased progeny production was accompanied by a greater load of mutations (Figure 24C). This result was supported by previous

studies that showed that increased polymerase fidelity reduced fitness (measured as the BS) in both T4 and VSV (Mansky and Cunningham, 2000; Furió et al., 2005). Under a scenario where the dynamics of intra-cellular viral particles positively correlate with the load of mutations, more mutations will accumulate in fewer infection cycles, thus accelerating adaptation. This has implications for the elaboration of attenuated live virus vaccines, where each additional passage of the clonal stock might generate an exponential increase of more robust and/or more virulent variants (Wimmer et al., 1993), but also for disease emergence, where more advantageous mutations will be generated in burst, accelerating thus host-to-host successful transmission.

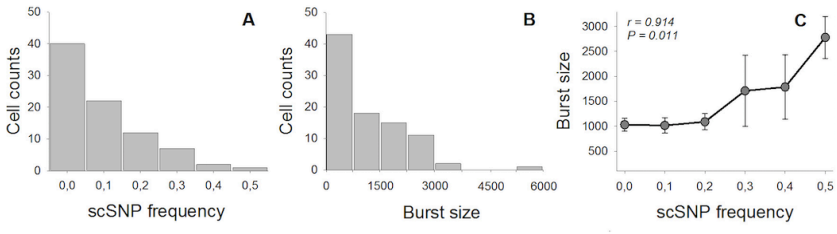


Figure 24. Viral production and frequency of SNPs in single-cells. A, distribution of scSNP frequencies in the viral progeny of each single-cell; B, burst size between single-cells; C, relationship between scSNP frequency and burst size (Pearson correlation: $r = 0.914$, $P = 0.011$). For each category of scSNP frequency the average burst size is represented.

2.6 Low-frequency variants with a mutator phenotype

Among the 881 plaques analyzed, 670 were mutated (76%) with 389 plaques (58%) accumulating more than one

mutation in burst. Previous work suggested that viral populations should be composed of subpopulations with different mutation frequencies, (Schubert et al., 1984; Ninio, 1991; Drake et al., 2005), as clearly demonstrated for Spleen Necrosis Virus (SNV) where a single provirus showed a mutation rate 1,000-fold higher than the rate determined for the virus population because of the accumulation of multiple mutations in its genome (Pathak and Temin, 1990). Moreover, the occurrence of low-frequency variants with multiple mutations has been proposed to be a common phenomenon in viral populations (Parthasarathi et al., 1995; Tromas and Elena, 2010; García-Villada and Drake, 2012), as recorded for poliovirus populations (de la Torre et al., 1992), for the bacteriophage T4 gene (Bebenek et al., 2001), but also for TMV (*Tobacco mosaic virus*), CMV (*Cucumber mosaic* virus) and CCMV (*Cowpea chlorotic mottle* virus) (Schneider and Roossinck, 2000). In bacteria, low-frequency mutators have been shown to significantly contribute to population adaptability and virulence (Gonzalez et al., 2008; Jolivet-Gougeon et al., 2011; LeClerc et al., 1996; Sniegowski et al., 1997). For viruses, one could hypothesize that whereas single SNPs might be neutral or deleterious, they could become advantageous when combined together within the same infectious particle (Drake, 2005), as exemplified above with the hijacking of the lethal effect of stop codons for the virus. Also, in terms of adaptability, the occurrence of multiple mutations in a single replication cycle might confer a selective advantage for successful cross-species transmission and adaptation to new environments (Stech et al., 1999; Kuiken et al., 2006).

To determine if low-frequency mutator variants could underly the variability between cells in the *de novo* production of genetic diversity, we measured the mutation rate of individual plaques picked from distinct cells using the Luria-Delbrück fluctuation test (see section 4.1 in "Materials and Methods"). We selected three plaques from a cell showing a high genetic diversity, such as cell 36 with 10 SNPs, and two plaques from a less diverse cell, such as cell 81 with 2 SNPs (Table 7). From cell 36, plaques 36.1 and 36.9 had replacements in residues 446 and 447 of the polymerase (L protein), but they did not map to any of the six functionally relevant conserved regions (CRs) of this protein. However, plaque 36.2 had an additional replacement in residue 957, thus mapping to CRIV, a concerved domain that is a key structural component of the replicase ring domain (Rahmeh et al., 2010). Finally, from cell 81, plaques 81.9 and 81.10 did not have any amino acid replacement in the genome. Therefore these plaques provided the wild-type mutation rate. We found differences in the mutation rate to the MAR phenotype between individual plaques (one-way ANOVA, $p = 0.002$), with plaque 36.2 showing a significant 3.6-fold increase compared to the other four plaques (Tukey's post-hoc test, $p < 0.05$).

Table 7. Mutation rate to the MAR phenotype determined by the Luria-Delbrück fluctuation test for viral plaques picked from single-cells. Following the amino acid (AA) change, the protein in which the change occurred is indicated, as the conserved region in the polymerase.

Cell	Plaque	Mutation	AA change	MAR[a] × 10^{-5}
36	1	C4636A	None	1.08 ± 0.28
		A6069G	E446G (L)	
		A6070G	E446G (L)	
		C6071U[b]	L447F (L)	
36	2	C2833U	S195F (M)	4.03 ± 2.02
		G3485U	None	
		A6069G	E446G (L)	
		A6070G	E446G (L)	
		C6071U	L447F (L)	
		A7601G[b]	M957V (L: CRIV)	
36	9	A3620C	None	1.27 ± 0.20
		C6071U	L447F	
81	9	None	None	1.20 ± 0.46
81	10	None	None	0.91 ± 0.15

[a] MAR refers to mutation rate (*m*) to the MAR phenotype, expressed as substitutions per nucleotide per round of copying (s/n/r).
[b] A non-expressed stop codon mutation (U2330A) was also found in these plaques.

Our results suggested that the variability in the mutation rate between individual plaques released from a cell could contribute to explaining differences in the viral genetic diversity observed between cells. Since progeny released from cells with

different genetic diversity showed similar mutation rates (plaques 36.1, 36.9 and plaques 81.9, 81.10), but that progeny within the same cell had different mutation rates (plaques 36.1, 36.9 and plaque 36.2), we proposed that the production of genetic diversity between cells might be the result of mixed stochastic cellular and viral effects, such that a small fraction of the progeny accumulated multiple mutations and showed a mutator phenotype, whereas other plaques had only few or no mutations.

Chapter 3. Norwalk virus mutation rate

Human noroviruses are highly genetically diverse, and viral populations co-circulating change between and within outbreaks (Siebenga et al., 2007). Variation at the genome level can result in silent mutations (i.e. homologous amino acid substitutions) or in amino acid substitutions that change the properties of virus binding or the immunogenicity of virus epitopes, and thus lead to the generation of variants with altered tropism or antibody-escape mutants (Gallimore et al., 2007). Because the success of these variants is either due to immune escape or changes in receptor binding, the high diversity and exposed structure of the P2 domain of the capsid (VP1) protein has lead to the asumption that receptor binding sites and key antigenic determinants should be localised in this region (Cao et al., 2007; Tan et al., 2003).

The cell tropism of human noroviruses and the development of an *in vitro* infection model remain elusive, restricting thus considerably their study. Recently a cell culture system has been developed with the discovery of B cells as a cellular target of human GII-4 norovirus strain and histo-blood group antigen (HBGA) expressed by enteric bacteria being necessary for norovirus susceptibility and infection (Jones et al., 2014). Human B cells possess three members of the ADAR family: ADAR1 and ADAR2 that are functional deaminases (Bass and Weintraub, 1988; Kim et al., 1994) and ADAR3, which do not have a known enzymatic function (Chen et al., 2000). Recently, DNA and mRNA sequences from B cell cultures were examined by deep

sequencing and the results showed extensive ADAR-mediated A-to-G editing sites in premature and mature mRNAs as well as in long noncoding RNAs (Wang et al., 2013). However, no studies have focused on the effect of ADAR on norovirus genome replication in human cells.

Noroviruses replicate quickly and are expected to exhibit high rates of mutation (Dingle, 2004), although there is no clear information yet about the rate at which spontaneous mutations accumulate in the norovirus genome. The only available information comes from studies where mutations were scored after several weeks of virus shedding, reflecting thus substitution rates rather than mutation rates. Nilsson et al. (2003) found 32 amino acid changes in the capsid protein of a norovirus shed chronically over one year by an immunocompromised patient, whereas Dingle (2004) sequenced 49 viruses from 22 outbreaks and observed a high mutation frequency (4 point mutations in 3,255 nt sequenced).

Because of the impact of norovirus-associated gastroenteritis outbreaks in the human population, it is necessary to have a better understanding about the type and frequency of spontaneous mutations in noroviruses. Here we measured the NV mutational spectrum after transfection of HEK293T cells. Also, because of its essential function in cell binding and neutralizing antibodies (Lochridge et al., 2005), we focused on spontaneous mutations arising in the VP1 gene. Finally, since the cellular expression of ADAR in the human HEK293T cells has been demonstrated before (Maas et al., 1996; Maas et al., 2001; Clerzius

et al., 2009; Li et al., 2010), this allowed us to estimate the potential effect of the ADAR editing on viral mutation rates.

3.1 Transfection of NV cDNA into HEK293T cells and amplification efficiency

Norwalk virus cDNA (pNV) was transfected into HEK293T cells previously infected with the recombinant vaccinia vTF7-3 virus to allow the primary transcription of the cDNA clone. After 5 h incubation, replication of any remaining vaccinia was blocked and transcribed NV RNA should replicate then by itself into the cytoplasm of the cells. Because NV is not able to infect HEK293T cells, we scored spontaneous mutation after only one cycle of cell infection. This allowed us to avoid the effect of selection or to score the accumulation of mutations over several generations. Three transfections were performed (NV1, NV2, NV3) and (-)-strands RNA from supernatants and from cells were RT-PCR amplified using strand-specific primers in the RT step. However, since we got a weak PCR amplification from supernatants, only RNA extracted from cells was used for subsequent experiments.

Positive and negative strands RNA were amplified from these three transfection samples using strand-specific primers (see section 8.2 in "Materials and Methods") and purified PCR products were used for molecular clone sequencing (see section 5.2 in "Materials and Methods") of the full VP1 gene (ORF2). In order to verify that the treatment of the samples with DNase I, RNase-free endonucleases have eliminated all remaining template DNA, and thus that positive RT-PCRs were not due to DNA amplification, PCRs of each transfection were performed without

reverse-transcription. The results showed no amplification, whilst the positive control (i.e. with cDNA) was efficiently amplified. Mutations scored in (+)-strands were used to correct for mutation frequency of (-)-strands RNA. Indeed, because of the large amount of cDNA template inoculated to transfect, most mutations found in (+)-strands RNA might have been introduced by T7 RNA polymerase, whereas mutations in (-)-strands RNA should have been generated by the NV polymerase during replication.

3.2 Norwalk virus mutational spectrum and mutation frequency

In (+)-strands RNA, a total of 10 mutations were observed among 296,412 nucleotides analysed, with 6 base substitutions and 4 indels (Table 8). Among these base substitutions, we found 3 transitions and 3 transversions. In parallel, clones of negative polarity synthesized during NV self-replication were also sequenced. A total of 35 mutations were observed among the 272,064 bases examined, with 30 base substitutions and 5 indels (Table 8). These results indicated that (-)-strands RNA accumulated significantly more mutations than (+)-strands RNA (Fisher's exact test, $p = 1.5 \times 10^{-5}$), confirming that mutations found in (+)-strands might have been generated by T7 RNA polymerase, whereas most mutations found in (-)-strands should have been introduced by the NV polymerase. Also, we observed a bias towards transition mutations, with a transition:transversion ratio of 26:4.

The above results allowed us to determine the mutation frequency f of (+)- and (-)-strands RNA. We found $f = 2.02 \times 10^{-5}$ for (+)-strands and $f = 1.1 \times 10^{-4}$ for (-)-strands RNA. Therefore,

after correction for T7-induced mutations in (-)-strands RNA, we found $f = 9 \times 10^{-5}$. Since only one cycle of cell infection elapsed mutation frequency and mutation rate were similar in this experiment ($\mu = 9 \times 10^{-5}$ s/n/c). These results provide a first quantification of the NV rate of spontaneous mutation, and shows that mutations accumulate quickly, similar to other RNA viruses (Sanjuán et al., 2010).

Table 8. Mutations found in (+)- and (-)-strands RNA.

Type of Mutation	(+)-strand RNA	(-)-strand RNA
Clones sequenced	138	128
Bases analyzed	296,412	272,064
Total mutations	10	35
Base substitutions	6	30
Indels	4	5
Transitions	3	26
Transversions	3	4
Synonymous	1	13
Nonsynonymous	5	17

3.3 Hypermutated NV sequences

Of the 30 base substitutions observed 23 were T -> C transitions (Figure 25). Close examination of the sequences of the 128 clones sequenced showed two clones with several T -> C mutations. Clone M1.8 had 6 T -> C changes in a region composed

of 452 bases, such that these hypermutations appeared at a frequency of 1.3×10^{-2}, whereas clone NV2.5 showed 16 T -> C substitutions in a region composed of 456 nucleotides, such that at a frequency of 3.5×10^{-2} (Table 9). These results suggested that, during genome replication, a small fraction (1.5%) of the NV sequences were hypermutated. Since sequences were compared against the plus-strand reference genome, these T -> C hypermutations actually correspond to A -> G mutations occurring in the (-)-strands RNA used for RT-PCR.

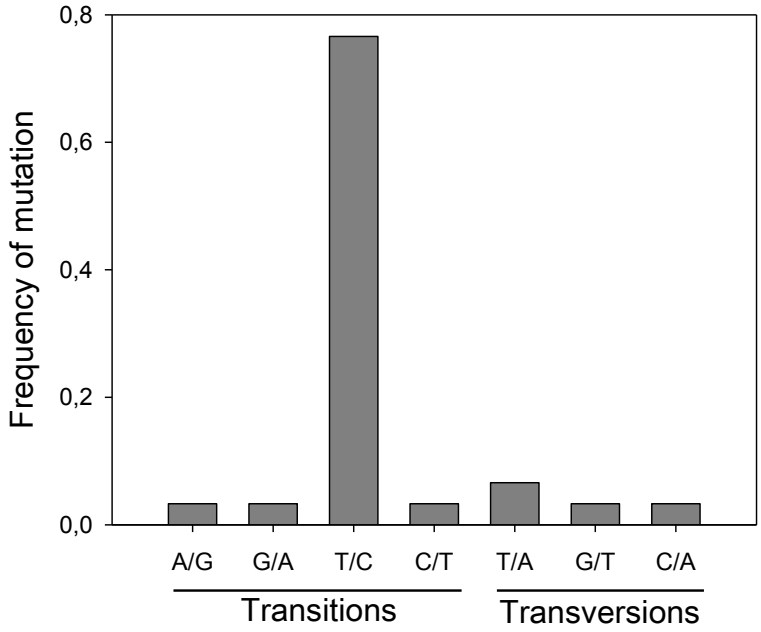


Figure 25. Frequency of each type of base substitution found in (-)-strands RNA. Among transitions T -> C changes were significantly more abundant than other types of mutation (Fisher's exact test, p = 0.0007).

Because of the small fraction of clones that were hypermutated, and the very high frequency at which these hypermutations occured, we hypothesized that they might be the results of the editing of cellular enzymes, such as adenosine deaminases of the ADAR family (Valente and Nishikura, 2005), rather than being generated by the NV polymerase. Since ADAR editing can occur on either the genome or antigenome (Samuel, 2011), this supported that adenosine deamination to guanosine occurred in (-)-strands RNA. Whilst ADAR editing has been shown to affect several RNA viruses, this cellular enzyme is also known to require dsRNA as template (see section 3.2.1 in "Introduction"). One hypothesis could be that ADAR-like mutations tend to accumulate in folded regions of the NV RNA. Alternatively, ADAR might have hit on intermediate dsRNAs produced by the NV RdRp during genome replication, as it is widely assumed for (+)-strand RNA viruses (Jacobs and Langland, 1996; Weber et al., 2006). However, under this latter scenario A -> G hypermutations should also be observed in (+)-strands RNA.

Table 9. Mutations in the VP1 gene of (-)-strands RNA clones. Insertions are symbolized by a "+" and deletions by a "-" symbol. "AA" means amino acid.

Clone	Mutation	AA change	Codon change
M6.2	G5406A	Gly -> Ser	GGC -> AGC
M6.6	C5868T	His -> Tyr	CAT -> TAT
M6.1	+6014T	-	-
M6.16	T6523C	Leu -> Pro	CTT -> CCT
M6.2	C6747A	Pro -> Thr	CCT -> ACT
M6.3	T6841A	Leu -> Gln	CTG -> CAG
NV3.5	-5696A	-	-
NV3.5	-5697G	-	-
NV3.5	-5698C	-	-
NV3.5	-5699T	-	-
NV3.14	G6091T	Arg -> Leu	CGT -> CTT
NV3.12	-6094C	-	-
M1.8	T5732C	Val -> Val	GTT -> GTC
M1.8	T5804C	Ile -> Ile	ATT -> ATC
M1.8	T5840C	Pro -> Pro	CCT -> CCC
M1.8	T5860C	Val -> Ala	GTT -> GCT
M1.29	-5874A	-	-
M1.29	-5875A	-	-
M1.29	-5876T	-	-
M1.8	T5911C	Met -> Thr	ATG -> ACG
M1.8	T6184C	Leu -> Pro	CTG -> CCG
M1.4	-6222A	-	-

M1.33	A6320G	Pro -> Pro	CCA -> CCG
NV2.5	T5956C	Phe -> Ser	TTT -> TCT
NV2.5	T6134C	Asn -> Asn	AAT -> AAC
NV2.5	T6150C	Phe -> Leu	TTC -> CTC
NV2.5	T6167C	Cys -> Cys	TGT -> TGC
NV2.5	T6250C	Val -> Ala	GTA -> GCA
NV2.5	T6284C	Phe -> Phe	TTT -> TTC
NV2.5	T6315C	Phe -> Leu	TTT -> CTT
NV2.5	T6332C	Gly -> Gly	GGT -> GGC
NV2.5	T6335C	Cys -> Cys	TGT -> TGC
NV2.5	T6339C	Trp -> Arg	TGG -> CGG
NV2.5	T6344C	His -> His	CAT -> CAC
NV2.5	T6368C	His -> His	CAT -> CAC
NV2.5	T6371C	Ser -> Ser	TCT -> TCC
NV2.5	T6384C	Tyr -> His	TAT -> CAT
NV2.5	T6391C	Val -> Ala	GTA -> GCA
NV2.5	T6412C	Phe -> Ser	TTT -> TCT
M2.35	T6248A	Thr -> Thr	ACT -> ACA

3.4 ADAR-like mutations revealed by deep sequencing

Although the mutations observed here might have been introduced by the editing of cellular enzymes such as ADAR, only few clones were analyzed by Sanger and among these only two sequences showed ADAR-like mutations. To further investigate this finding we set up a protocol based on a nested PCR amplification that allows us to amplify ADAR-edited RNA

sequences using degenerated primers (see section 8.3 in "Materials and Methods") and the 454 deep-sequencing technology.

We focused on ADAR-like mutations and thus when more than two T -> C or A -> G changes occurred in a specific read, we considered the read as possibly hypermutated. Interestingly, we observed both types (i.e. T -> C and A -> G) of ADAR-like mutations, in contrast to the above results from Sanger where only T -> C hypermutations occurred (Table 10). However, significantly less A -> G changes were scored by 454 sequencing, compared to the number of T -> C hypermutations (Fisher's exact test: $p < 0.001$). Therefore it might be that because of their lower frequency A -> G hypermutations were missed by Sanger. This finding suggested thus that: (i) the NV RNA might have been edited by the cellular enzyme ADAR during genome replication, thus confirming its expression in HEK293T cells (Maas et al., 1996; Maas et al., 2001; Clerzius et al., 2009; Li et al., 2010), (ii) A -> G hypermutations occurred in (-)-strands RNA RT-PCR amplified (remember that T -> C changes were scored in strands of positive polarity sequenced whilst strands of negative polarity were amplified), and (iii) some A -> G hypermutations also occurred in (+)-strands RNA.

Whilst 2 out of 128 clones analysed by Sanger (frequency of 1.5×10^{-2}) were hypermutated, the frequency of hypermutated reads ranged between 3×10^{-3} and 1.7×10^{-2} for T -> C changes, and between 2.4×10^{-4} and 2×10^{-3} for A -> G changes. These results confirmed that during genome replication within the host

cell, a fraction of the sequences synthesized by the RdRp showed several ADAR-like mutations. This finding suggests that the variability in RNA virus mutation rates might be explained in part by host factors that lead to the production of hypermutated viral genomes. For instance, one could extrapolate that host factors such as ADAR could also explain differences in mutation rates between mammalian and plant viruses that seem to lack the immune pathway responsible for the activation of ADAR (Wang and Garmichael, 2004).

Table 10. Hypermutated reads for each region of the VP1 gene analysed by 454 deep sequencing. For each transfection sample, region 1 and region 2 sequenced are indicated, as well as the total number of reads generated by the 454 technology. The number of reads potentially hypermutated (i.e. > 2 T -> C or A -> G) is indicated with the corresponding frequency of hypermutated reads.

Sample	Region	Total reads	Reads T -> C	Frequency	Reads A -> G	Frequency
NV1	1	11,703	35	3×10^{-3}	24	2×10^{-3}
	2	16,636	63	3.8×10^{-3}	4	2.4×10^{-4}
NV2	1	8,332	143	1.7×10^{-2}	3	3.6×10^{-4}
	2	5,384	56	1×10^{-2}	8	1.5×10^{-3}
NV3	1	9,986	139	1.4×10^{-2}	2	4×10^{-4}
	2	36,612	268	7.3×10^{-3}	13	3.5×10^{-4}

Independently of the type of changes observed, significantly more hypermutated reads were scored in region 1, which is composed of only 266 bp as opposed to 405 bases in region 2 (Fisher's exact test: $p < 0.001$ for T -> C and $p = 0.035$ for A -> G). It has been proposed that the structure of the RNA substrate, rather than its sequence, might determine selectivity of ADAR (defined as the fraction of modified adenosines in an RNA substrate) (Lehmann and Bass, 1999). For instance, RNAs that are completely double-stranded are deaminated non-selectively (i.e. approx. 50% of adenosines deaminated) whereas whose base-paired structures that are interrupted by mismatches, bulges or internal loops are deaminated selectively and thus show less deamination. To check for this finding as explaining the higher level of hypermutations in region 1 compared to region 2, ADAR-like mutations found in reads with more than 10 of these mutations were mapped to the predicted minimum free energy (MFE) RNA secondary structure of each region, which was obtained using the RNAfold program online (rna.tbi.univie.ac.at/cgi-bin/RNAfold/cgi) (Figure 26). After closely looking at both stuctures in terms of internal loops, bulges and hairpins, we observed that region 1 shows a long stem in which most of the ADAR-like mutations (26 out of 44 mutations) were scored. This finding suggests that region 1 might have been more promiscuously deaminated by ADAR, whereas region 2 could have been more selectively modified, thus explaining why region 1 shows more ADAR-like mutations.

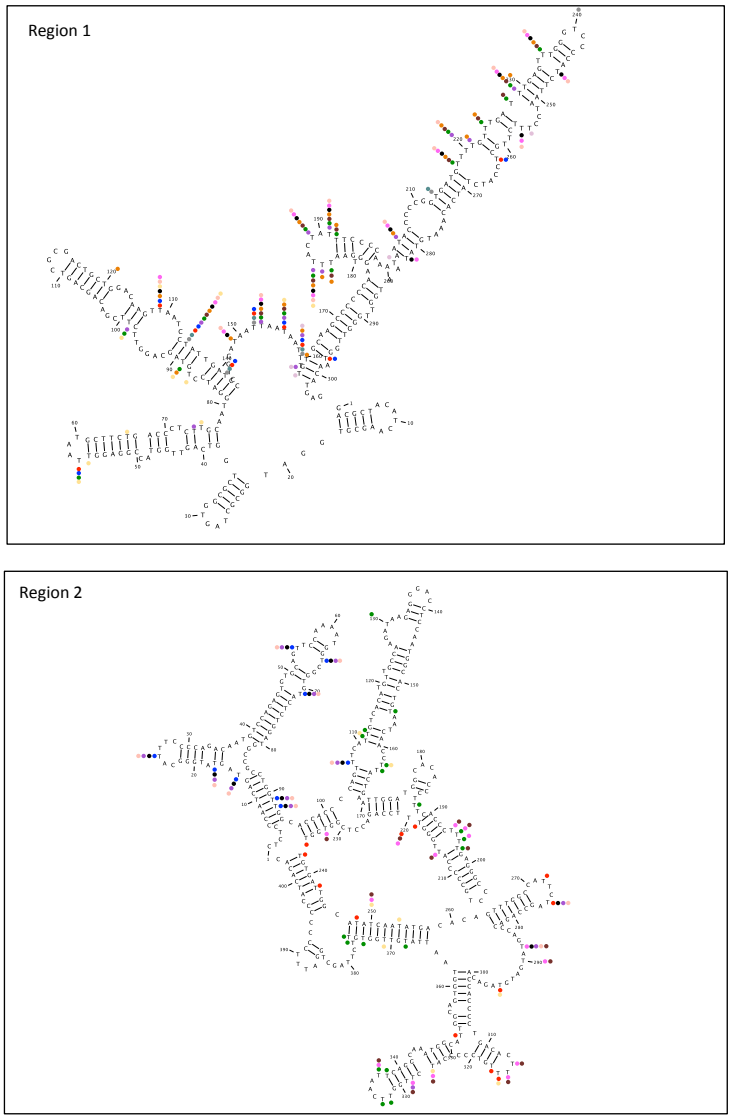


Figure 26. RNA secondary structure of region 1 and region 2 analyzed by 454 deep sequencing for ADAR-like mutations. For each region, ADAR-like mutations found in reads with > 10 of these mutations are mapped. For each mutation, each colored dot indicates a different read. The MFE structure was plotted using CLC workbench. Region 1 shows a long stem in which most of the ADAR-like mutations were scored (26 out of 44 mutations).

In addition to selectivity, it has been proposed that in an RNA substrate ADAR has also preferences, such that an adenosine is more efficienctly deaminated when it has as 5' neighbor As or Us, mainly because of their effect on base flipping (Nishikura et al., 1991; Kuttan and Bass, 2012). Indeed, AU or UA base pair (the underscore shows the 5' neighbor of the deaminated base) are optimal whereas GC or CG base pair should reduce binding to ADAR and thus lower the frequency of deamination at the adjacent adenosine (Polson and Bass, 1994). Irrespective of the RNA structure, we screened the sequences of these reads with > 10 ADAR-like mutations for the 5' neighbor of the As deaminated in the (-)-strand RNA (Table 11). For region 1 as well as for region 2 significantly more 5' neighbor As or Us were observed next to the deamination site, whereas in contrast more Gs or Cs preceded a non-deaminated base (Fisher's exact test, $p = 2.4 \times 10^{-4}$ and $p = 9.6 \times 10^{-7}$, respectively). In agreement with previous studies, our results show that ADAR has preferences for those adenosines that have as 5' neighbor As or Us.

Table 11. Preference for 5' neighbor As or Us next to ADAR deaminated site. For each region, the 5' neighbor A, T, G or C counts are indicated for deaminated As as well as unmodified As in (-)-strand RNA.

	Region 1				Region 2			
	A	T	G	C	A	T	G	C
Deaminated As	16	12	4	7	23	15	7	11
Unmodified As	7	1	11	12	10	1	14	28

Interestingly, region 2 falls in the P2 hypervariable domain of the VP1 gene, responsible for host interaction and antigen recognition (Cao et al., 2007; Prasad et al., 1999). Although previous studies from mammals and bacteria suggested that host-pathogen interaction loci showed a hypermutation response (Metzgar and Wills, 2000; Rando and Verstrepen, 2007; Teng and Papavasiliou, 2007; Moxon et al., 2006), the opposite scenario has been proposed for HIV-1, where fewer mutations accumulated in the region of the envelope protein that contains most antibody targets and determines cell tropism (personal communication from Geller et al., manuscript under submission). Whilst more sites of the NV genome have to be screened, our results suggest that different regions of the viral genome might exhibit different affinity for the antiviral cellular enzymes to act. Focusing on these particular regions could help determining key viral-host mechanisms underlying variability in mutation rates.

General Discussion

Riboviruses are the causative agents of important infectious diseases. Wilst the mortality rates and economic costs associated to these infectious pathogens are considerable in developing countries (Holmes, 2009), only few of these diseases can be efficiently controlled by vaccination or antiviral drugs. Indeed, these strategies are counterbalanced by the quick emergence of new riboviral pathogens or drug-resistant variants (Domingo, 2010; Pepin et al., 2010). A key feature that confers to riboviruses an extensive genetic diversity, allowing them to quickly adapt to changing environments, is their extremely high mutation rates (Duffy et al., 2008; Sanjuán et al., 2010; Combe and Sanjuán, 2014a). Whilst several virus-encoded factors have been shown to play an important role in the evolution of RNA viruses, only few studies have focused on host factors underlying the variability in the production of viral genetic diversity, neither on whether this variability could be heterogeneously distributed among individual cells.

When this PhD started, the effect of the cell metabolism on viral mutation rates was reported for retroviruses (Diamond et al., 2004; Bebenek et al., 1992; Julias and Pathak, 1998) and hepatitis C virus (Seronello et al., 2011) only. No studies have investigated the differences in the rate at which mutations accumulate in both hosts of arthropod-borne viruses as an alternative explanation for their slow evolution and thus fitness trade-offs only were assumed to drive this evolution (Jenkins et

al., 2002; Woelk and Holmes, 2002; Hicks and Duffy, 2014). Also, although recent advances in deep sequencing technologies allowed a better exploration of viral genetic diversity, no study have focused on the variability in the production of this diversity among single cells (Acevedo et al., 2014). However the extensive cell-to-cell variation in terms of metabolism, physiological status or gene expression patterns has obviously major implications for the outcome of viral infection (Pathak and Temin, 1990; Monk et al., 1992). Finally, although the expression of cellular genes such as APOBEC3G cytidine deaminases or RNA-dependent adenosine deaminase (ADAR) were known to edit hypermutation in riboviruses, the impact of these cellular enzymes on norovirus mutation rates has not been quantified before. Because viruses are obligate parasites and that the interactions with their host (cell) are finely adjusted to allow their efficient replication and transmission, the variability and host-dependency in viral mutation rates had to be investigated.

In chapter 1, we looked at the effects of the cell type and host species in which the virus replicates on the mutation rate using the Luria-Delbrück fluctuation test. Because previous estimates have been obtained in BHK-21 cells, an immortalized cell line, we investigated the effect of cellular immortalization and other changes leading to a tumoral cell by testing mammalian cells of various origins, but also under different oxygen levels. However, our results suggested that VSV was not sensitive to differences in metabolic and mitotic activity, but rather replicated

with similar fidelity in these different mammalian cell types. Moreover, the fact that this virus was not affected by the aberrant metabolism of cancerous cells, neither by the lower oxygen levels encountered in these cells (1% 0_2), suggested that VSV should be a good candidate for the field of oncolytic virotherapy, as previously proposed (Breitbach et al., 2011). Interestingly when comparing the host species involved in the lifecycle of VSV, we found a 3.9-fold decrease in mutation rate in insect cells compared to mammalian cells. Arboviruses such as VSV are known to evolve more slowly than directly transmitted viruses, and this is commonly explained in terms of fitness trade-offs. Since only few studies on arboviruses focused on both insect and mammal hosts, our results offer a new possible explanation for the relatively slow arboviral evolution, based on lower rates of mutation in the insect reservoir.

In chapter 2, we focused on the viral genetic diversity released by single-cells infected with VSV using micromanipulation and deep sequencing approaches. We showed that single-cells exhibit extensive viral genetic diversity, with most progeny plaques released by each cell being mutated, whilst this diversity is not homogeneously allocated by individual cells. Interestingly, this approach allowed us to demonstrate for the first time that single PFUs co-transmit several and genetically diverse variants, thus delivering genetic diversity to each cell. This co-transmission allows (i) the genetic complementation of lethal alleles (premature stop codons) between variants

belonging to the same PFU and (ii) the hitchhiking of beneficial alleles (MAR mutant which is beneficial in the presence of antibody) suggesting that the effect of a given mutation is determined collectively by other variants during their co-transmission, a process that could last for several generations. The co-transmission of genetic diversity within the same infectious unit has important implications for viral evolution: first, it promotes the maintenance of genetic diversity during strong population bottlenecks (such as during host-to-host transmission), and thus the negative effect of such bottlenecks on viral fitness has to be revised. Second, it creates a spatial association between alleles and thus questions the selection unit. Third, it confers a greater robustness against the effect of deleterious mutations. Finally, we showed that the variability in the amount of diversity produced per cell was positively correlated to the viral yield. These results suggested a trade-off between the efficiency and the fidelity of viral replication, which may be driven by geometric replication of viral genomic RNA within some infected cells (Safari and Roossinck, 2014). Although this type of replication fosters the appearance of new mutations, it also increases the viral genetic load. Counterbalancing, variants with a mutator phenotype might be encountered at a low frequency in viral populations, as supported by the similar mutation rate of the VSV viral population used to infect single-cells (see “Chapter 1”) and non-mutator plaques, whereas the rate of spontaneous mutations was higher for plaque 36.2 that showed a mutator phenotype.

In chapter 3, we scored NV spontaneous mutations after transfection of HEK293T cells and focused on the VP1 gene, which is composed of both conserved and more variable (P2) domains. The mutational spectrum provided a first estimate of the mutation rate for NV of $\mu = 9 \times 10^{-5}$ s/n/c, which is high and consistent with previously published estimates for RNA viruses (Sanjuán et al., 2010). Interestingly, we observed that, out of the 128 clones of negative polarity sequenced by Sanger two showed several A -> G hypermutations at a very high frequency (> 10^{-2}). By allowing the amplification of hypermutated NV sequences using degenerated primers and analyzing them with deep sequencing we confirmed the high frequency of A -> G hypermutated sequences, with the majority being located in region 1. Our results suggested that a fraction of NV RNA might have been edited by the cellular enzyme ADAR during genome replication within the cytoplasm of HEK293T cells. Moreover, different regions of the genome might be either promiscuously or selectively edited by ADAR, depending on the structure of the RNA substrate (such as long stems) as well as the 5' neighbor of the modified adenosine. These findings suggested thus that the variability in mutation rates might also be explained by the editing of cellular enzymes.

The experiments performed during this PhD present a unifying framework that helps to better appreciate and explain the variability in viral mutation rates as well as its host-dependency. Whilst mutation rates are similar in different mammalian cells in Chapter 1, we showed in Chapter 2 that the

genetic diversity is heterogeneously distributed between individual cells, suggesting thus that this constancy in mutation rates is likely due to the averaging effect of cell cultures. In Chapter 2 we demonstrate for the first time the variability in the level of genetic diversity during the course of a cell infection: first, infecting PFUs deliver a pool of variants to each cell. Then, in addition to this pre-existing variability, each cell releases different levels of genetic diversity as a new pool of variants able to infect another cell. Finally, within a cell each individual plaque show different mutation rates, also explaining the variability observed between cells. These findings allow us to better understand how RNA viruses deal with high rates of mutation and the resulting genetic load. We propose that RNA virus populations might cluster their genetic diversity in a subset of cells, such that they exist as a dynamic equilibrium between the fast appearance of low-frequency mutator variants accelerating adaptation and their elimination by out-competition of less diverse viruses preserving their genetic integrity. Moreover, Chapter 1 proposes an alternative explanation for the slow evolution rate of arboviruses, based on their lower mutation rates in the insect host. These differences between insects and mammals could be explained by a trade-off between replication efficiency and fidelity, as proposed in Chapter 2. Finally, the results obtained in Chapter 3 improve our knowledge on the effect of cellular editing enzymes such as ADAR on the production of genetic diversity and suggests that this host-mediated editing may play a major role in explaining variability in RNA virus mutations rates.

The results obtained through this PhD offer several possible future lines of research:

- Whilst Chapter 1 showed that VSV exhibits significantly lower mutation rates in the insect reservoir, more experiments with other arboviruses will be necessary to elucidate the generality of our findings. One possible explanation to higher mutation rates in mammals could be that arboviruses replicate more quickly in mammalian cells than in insect cells (Rozen-Gagnon et al., 2014). Then, if as suggested in Chapter 2, there is a trade-off between replication efficiency and fidelity, increased viral yield will also correlate with more mutations that accumulate in fewer replication cycles. Since the effect of differences in host temperature as well as in our sensitivity to detect MAR mutants on the lower viral yield in insect cells was discarded, future studies will have to clarify the cellular host factors implicated or missing in these host cell lines.

- The results of Chapter 2 showed that cell cultures are definitely not homogeneous environments and thus to better understand the virus-host mechanisms underlying high genetic diversity among RNA viruses, as well as their efficient transmission, this diversity has to be studied at the single-cell level. The micromanipulation approach developed here could be used for the study of other RNA viruses, such that for those viruses for which no

phenotypic markers are available. Moreover, since we found a trade-off between replication efficiency and fidelity, the mode by which RNA viruses replicate their genome should be investigated in more details. Finally, because of the major implications of virus co-transmission within single PFUs on viral evolution, it could be interesting to have a better quantification of the level of genetic diversity delivered to individual cells.

- Whilst in Chapter 3 ADAR-like hypermutations were observed from transfected samples, we propose that future experiments should attempt to generalize this finding by analysing the genetic diversity of norovirus isolates from stool samples. Since these hypermutations occurred at a high frequency they might play an important role in the production of genetic diversity during genome replication within the host and thus might strongly impact viral mutation rates. Therefore, future studies should focus on (i) the strength of this cellular editing on viral mutation rates and (ii) the variability in cellular editing between different regions of the viral genome, and most notably between those involved in virus-host reconition and receptor binding sites.

FINAL CONCLUSIONS

Looking at the effect of the cell type and the host species on viral mutation rates:

- The mutation rate of VSV was homogeneous between mammalian cells of various origins, such as BHK-21, MEFs and MEFs p53-/-, CT26 and Neuro-2a, suggesting that the fidelity of viral replication was not affected by differences in metabolic and mitotic activities between mammalian cells.

- VSV was not sensitive to different oxygen levels, since the mutation rate was similar under normoxia (20% 0_2) and hypoxia (1% 0_2).

- When comparing phylogenetically more distant hosts, VSV showed a mutation rate 3.9-fold lower in insect cells compared to mammalian cells. This finding suggests a new explanation for the lower evolutionary rates of arboviruses compared to directly transmitted viruses.

Focusing at the viral genetic diversity released by single-cells:

- Single-cells showed extensive viral genetic diversity. As a consequence of this high intra-cellular genetic diversity,

most progeny plaques released by each cell should, on average, be mutated.

- Infectious particles delivered viral genetic diversity to individual cells by co-transmission of pools of variants. This might promote the maintenance of genetic diversity during strong population bottlenecks, and allow the genetic complementation of alleles within infected cells. These findings also question the selection unit, and confer a greater robustness to deleterious mutations.

- Increased viral yield per cell was accompanied by a greater load of mutations, suggesting a trade-off between the efficiency and the fidelity of viral replication, which may be driven by the geometric model.

- Within some individual cells plaques with increased rates of spontaneous mutations contributed to the production of genetic diversity, thus explaining the variability observed between cells.

- To counterbalance the genetic load associated to geometric replication, these mutator variants should be maintained at low frequency in viral populations.

Determining norovirus mutation accumulation and the impact of cellular editing enzymes such as ADAR on mutation rates:

- We have provided an estimate of the mutation rate of NV in transfected HEK293T cells.

- The NV mutational spectrum showed a small fraction of A -> G hypermutated sequences, a finding that was further confirmed by deep sequencing. These hypermutations are reminiscent of ADAR edition, suggesting that this cellular enzyme might impact the production of genetic diversity and thus contribute to explaining the variability in mutation rates among RNA viruses.

REFERENCES

Aaskov, J., Buzacott, K., Thu, H.M., Lowry, K. and Holmes, E.C. (2006). Long-term transmission of defective RNA viruses in humans and Aedes mosquitoes. Science *311, 236-238*.

Abraham, G. and Banerjee, A.K. (1976). The nature of the RNA products synthesized in vitro by subviral components of vesicular stomatitis virus. Virology *71, 230-241*.

Acevedo, A., Brodsky, L. and Andino, R. (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature *505, 686-690*.

Akira, S. and Takeda, K. (2004). Toll-like receptor signalling. Nat. rev. Immunol. *4, 499-511.*

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*.

Allen, D.J., Gray, J.J., Gallimore, C.I., Xerry, J. and Iturriza-Gomara, M. (2008). Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. PLoS One *3, 1-9*.

Anderson, J.P., Daifuku, R. and Loeb, L.A. (2004). Viral error catastrophe by mutagenic nucleosides. Annu. Rev. Microbiol. 58, 183-205.

Arnold, J.J., Vignuzzi, M., Stone, J.K., Andino, R. and Cameron, C.E. (2005). Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. J. Biol. Chem. *280, 25706-25716.*

Asanaka, M., Atmar, R.L., Ruvolo, V., Crawford, S.E., Neill, F.H. and Estes, M.K. (2005). Replication and packaging of Norwalk virus RNA in cultured mammalian cells. Proc. Natl. Acad. Sci. U.S.A

102(29), 10327-10332.

Ball, L.A. and White, C.N. (1976). Order of transcription of genes of vesicular stomatitis virus. Proc. Natl. Acad. Sci. U.S.A *73, 442-446*.

Barr, J.N., Whelan, S.P. and Wertz, G.W. (2002). Transcriptional control of the RNA-dependent RNA polymerase of vesicular stomatitis virus. Biochim. Biophys. Acta. *1577, 337-353*.

Bass, B.L. and Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell *55, 1089-1098.*

Bass, B.L., Weintraub, H., Cattaneo, R. and Billeter, M.A. (1989). Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA. Cell *56, 331*.

Bebenek, A., Dressman, H.K., Carver, G.T., Ng, S-s., Petrov, V., Yang, G. et al. (2001). Interacting fidelity defects in the replicative DNA polymerase of bacteriophage RB69. J. Biol. Chem. *276, 10387-10397*.

Bebenek, K., Roberts, J.D. and Kunkel, T.A. (1992). The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. J. Biol. Chem. *267, 3589-3596*.

Bergmann, M., Garcia-Sastre, A., Carnero, E., Pehamberger, H., Wolff, K., Palese, P. et al. (2000). Influenza virus NS1 protein counteracts PKR-mediated inhibition of replication. J. Virol. *74, 6203-6206.*

Bertolotti-Ciarlet, A., Crawford, S.E., Hutson, A.M. and Estes, M.K. (2003). The 3' end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. J. Virol. *77, 11603-11615*.

Bertolotti-Ciarlet, A., White, L.J., Chen, R., Prasad, B.V. and Estes, M.K. (2002). Structural requirements for the assembly of Norwalk virus-like particles. J. Virol. *76, 4044-4055*.

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*.

Biek, R., Pybus, O.G., Lloyd-Smith, J.O. and Didelot, X. (2015). Measurably evolving pathogens in the genomic era. Trends Ecol. Evol. *Doi: 10.1016/j.tree.2015.03.009.*

Bienz, B., Zakut-Houri, R., Givol, D. and Oren, M. (1984). Analysis of the gene coding for the murine cellular tumour antigen p53. EMBO J. *3, 2179-2183*.

Boehme, K.W., Popov, V.J. and Heidner, H.W. (2000). The host range phenotype displayed by a sindbis virus glycoprotein variant results from virion aggregation and retention on the surface of mosquito cells. J. Virol. *74(23), 11398-11406*.

Bonnac, L.F., Mansky, L.M. and Patterson, S.E. (2013). Structure-activity relationships and design of viral mutagens and application to lethal mutagenesis. J. Med. Chem. *56, 9403-9414*.

Bradley, A. (1987). Production and analysis of chimaeric mice. In teratocarcinomas and embryonic stem cells: a practical approach. Edited by Robertson E.J. Oxford, England: IRL Press *113-152*.

Bradwell, K., Combe, M., Domingo-Calap, P. and Sanjuán, R. (2013). Correlation between mutation rate and genome size in riboviruses: mutation rate of Bacteriophage Qβ. Genetics *195, 243-251*.

Bray, G. and Brent, T.P. (1972) Deoxyribonucleoside 5'-triphosphate pool fluctuations during the mammalian cell cycle. Biochim. Biophys. Acta. *269, 184-191*.

Breitbach, C.J., De Silva, N.S., Falls, T.J., Aladl, U., Evgin, L., Paterson, J., et al. (2011). Targeting tumor vasculature with an oncolytic virus. Mol. Ther. *19, 886-894*.

Bull, J.C., Sanjuán, R. and Wilke, C.O. (2007). Theory of lethal mutagenesis for viruses. J. Virol. *81, 2930-2939.*

Bussereau, F. and Flamand, A. (1978). Coinfection with a rhabdovirus: vesicular stomatitis virus of Indiana and New-Jersey serotypes. Ann. Microbiol. *129B, 71-86*.

Cao, S., Lou, Z., Tan, M., Chen, Y., Liu, Y., Zhang, Z., et al. (2007). Structural basis for the recognition of blood group trisaccharides by norovirus. J. Virol. *81(11), 5949-5957*.

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*.

Carreau, A., El Hafny-Rahbi, B., Matejuk, A., Grillon, C. and Kieda, C. (2011). Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. J. Cell. Mol. Med. *15, 1239-1253*.

Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., Ter, M.V and Billeter, M.A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. Cell *55, 255-265*.

Chambers, P., Rima, B.K. and Duprex, W.P. (2009). Molecular differences between two Jeryl Lynn mumps virus vaccine component strains, JL5 and JL2. J. Gen. Virol. *90, 2973-2981*.

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

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*.

Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C. and Nishikura, K. (2000). A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. RNA *6, 755-767*.

Chen, Y.H., Du, W., Hagemeijer, M.C., Takvorian, P.M., Pau, C., Cali, A. et al. (2015). Phosphatidylserine vesicles enable efficient en bloc transmission of enteroviruses. Cell *160, 619-630*.

Clerzius, G., Gélinas, J.F., Daher, A., Bonnet, M., Meurs, E.F. and Gatignol, A. (2009). ADAR1 interacts with PKR during human immunodeficiency virus infection of lymphocytes and contributes to viral replication. J. Virol. *83, 10119-10128*.

Coffey, L.L., Beeharry, Y., Borderia, A.V., Blanc, H. and Vignuzzi, M. (2011). Arbovirus high fidelity variant loses fitness in mosquitoes and mice. Proc. Natl. Acad. Sci. USA *108(38), 16038-16043*.

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

Combe, M. and Sanjuán, R. (2014a). Variability in the mutation rates of RNA viruses. Future Virol. *9(6), 605-615*.

Combe, M. and Sanjuán, R. (2014b). Variation in RNA virus mutation rates across host cells. PLoS. Pathog. *10, e1003855*.

Comer, J.A., Tesh, R.B., Modi, G.B., Corn, J.L. and Nettles, V.F. (1990). Vesicular stomatitis virus, New Jersey serotype: replication in and transmission by Lutzomyia shannoni (Diptera: Psychodidae). Am. J. Trop. Med. Hyg. *42, 483-490*.

Corbett, T.H., Griswold, D.P., Roberts, B.J, Peckham, J.C. and Schabel, F.M. (1975). Tumor induction relationships in development of transplantable cancers of the colon in mice for chemotherapy assays, with a note on carcinogen structure. Cancer Res. *35, 2434-2439*.

Cruz, S.S., Chapman, S., Roberts, A.G., Prior, D.A.M. and Oparka, K.J. (1996). Assembly and movement of a plant virus carrying a Green fluorescent protein overcoat. Proc. Natl. Acad. Sci. USA *93, 6286-6290*.

Cuevas, J.M., Moya, A. and Sanjuán, R. (2005). Following the very initial growth of biological RNA viral clones. *J. Gen. Virol. 86, 435-*

43.

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

de la Iglesia, F., Martínez, F., Hillung, J., Cuevas, J.M., Gerrish, P.J., Daros, J.A. et al. (2012). Luria-Delbrück estimation of turnip mosaic virus mutation rate in vivo. J. Virol. *86, 3386-8*.

de la Torre, J.C., Giachetti, C., Semler, B. L. and Holland, J. J. (1992). High frequency of single-base transitions and extreme frequency of precise multiple-base reversion mutations in poliovirus. Proc. Natl. Acad. Sci. USA *89, 2531-5.*

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

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*.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Harti, C. et al., (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43, 491-498.*

Diamond, T.L., Roshal, M., Jamburuthugoda, V.K., Reynolds, H.M., Merriam, A.R., Lee, K.Y. et al. (2004). Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. J. Biol. Chem. *279, 51545-51553*.

Dingle, K.E. (2004). Mutation in a Lordsdale norovirus epidemic strain as a potential indicator of transmission routes. J. Clin. Microbiol. *42(9), 3950*.

Domingo, E. (2006). Quasispecies: concept and implications for virology. Springer *299, 1-31*.

Domingo, E. (2010). Mechanisms of viral emergence. Vet. Res. *41, 38*.

Domingo, E., Escarmís, C., Lázaro, E. and Manrubia, S.C. (2005). Quasispecies dynamics and RNA virus extinction. Vir. Res. *107, 129-139.*

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

Domingo, E., Sabo, D., Taniguchi, T. and Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. Cell *13, 735-744*.

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Pro. Natl. Aca. Sci. USA *88(16), 7160-4*.

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

Drake, J.W., Bebenek, A., Kissling, G.E. and Peddada, S. (2005). Clusters of mutations from transient hypermutability. Proc. Natl. Acad. Sci. USA *102, 12849-54*.

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*.

Duffy, S., Shackelton, L.A. and Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. *9, 267-76*.

Duizer, E., Schwab, K.J., Neill, F.H., Atmar, R.L., Koopmans, M.P. and Estes, M.K. (2004). Laboratory efforts to cultivate noroviruses. J. Gen. Virol. *85, 79-87*.

Eckerle, L.D, Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X. et al. (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. PLoS Pathog. *6, e1000896*.

Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L. and Denison, M.R. (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. J. Virol. *81, 12135-12144*.

Eigen, M. (1971). Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften *58, 465-523.*

Eigen, M. (1992). Steps towards life. Oxford University Press, New York.

Eigen, M., McCaskill, J. and Schuster, P. (1988). Molecular quasispecies. J. Phys. Chem. *92, 6881-6891*.

Eigen, M. and Schuster, P. (1977). The hypercycle, a principle of natural self-organization. Part A: Emergence of the hypercycle. Naturwissenschaften *64, 541-565*.

Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J. et al. (2012). Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell *150, 831-841*.

Elena, S.F., Bedhomme, S., Carrasco, P., Cuevas, J.M., de la Iglesia, F., Lafforge, G. et al. (2011). The evolutionary genetics of emerging plant RNA viruses. Mol. Plant Microbe Interact. *24, 287- 93*.

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

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*.

Fankhauser, R.L., Noel, J.S., Monroe, S.S., Ando, T. and Glass, R.I. (1998). Molecular epidemiology of "Norwalk-like viruses" in outbreaks of gastroenteritis in the United States. J. Infect. Dis. *178, 1571-1578*.

Follett, E.A., Pringle, C.R., Wunner, W.H. and Skehel, J.J. (1974). Virus replication in enucleate cells: vesicular stomatitis virus and

influenza virus. J. Virol. *13(2), 394-399*.

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*.

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*.

Galetto, R., Moumen, A., Giacomoni, V., Véron, M., Charneau, P. and Negroni, M. (2004). The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. J. Biol. Chem. *279, 36625-36632*.

Gallimore, C.I., Iturriza-Gomara, M., Xerry, J., Adigwe, J. and Gray, J.J. (2007). Inter-seasonal diversity of norovirus genotypes: emergence and selection of virus variants. Arch. Virol. *152, 1295-1303.*

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-Villada, L. and Drake, J.W. (2012). The three faces of riboviral spontaneous mutation: spectrum, mode of genome replication, and mutation rate. PLoS Genet*. 8, e1002832*.

Ge, P., Tsao, J., Schein, S., Green, T.J., Luo, M. and Zhou, Z.H. (2010). Cryo-EM model of the bullet-shaped vesicular stomatitis virus. Science *327: 689-693*.

Geiss, G., Jin, G., Guo, J., Bumgarner, R., Katze, M.G. and Sen, G.C.

(2001). A comprehensive view of regulation of gene expression by double-stranded RNA-mediated cell signaling. J. Biol. Chem. *276, 30178-30182.*

Gélinas, J.F., Clerzius, G., Shaw, E. and Gatignol, A. (2011). Enhancement of replication of RNA viruses by ADAR1 via RNA editing and inhibition of RNA-activated protein kinase. J. Virol. *85(17), 8460*.

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

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*.

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

Glass, R.I., Noel, J., Ando, T., Fankhauser, R., Belliot, G., Mounts, A., et al. (2000). The epidemiology of enteric caliciviruses from humans: a reassessment using new diagnostics. J. Infect. Dis. *181 (Suppl. 2), S254-S261*.

Gonzalez, C., Hadany, L., Ponder, R.G., Price, M. Hastings, P.J. and Rosenberg, S.M. (2008). Mutability and importance of a hypermutable cell subpopulation that produces stress-induced mutants in Escherichia coli. PLoS. Genet. *4, e1000208*.

Green, K.Y., Chanock, R.M. and Kapikian, A.Z. (2001). In Fields Virology, eds.Knipe, D.M. and Howley, P.M. (Lippincott Williams & Wilkins, Philadelphia), *6527-6532*.

Green, T.J., Zhang, X., Wertz, G.W. and Luo, M. (2006). Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science *313: 357-360*.

Guix, S., Asanaka, M., Katayama, K., Crawford, S.E., Neill, F.H., Atmar, R.L. et al. (2007). Norwalk virus RNA is infectious in

mammalian cells. J. Virol. *81(22), 12238*.

Hajjar, A.M. and Linial, M.L. (1995) Modification of retroviral RNA by double-stranded RNA adenosine deaminase. J. Virol. *69, 5878-5882*.

Hall, J.S., French, R., Morris, T.J. and Stenger, D.C. (2001). Structure and Temporal Dynamics of Populations within Wheat Streak Mosaic Virus Isolates. J. Virol. *75, 10231*.

Hanada, K., Suzuki, Y. and Gojobori, T. (2004). A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. Mol. Biol. Evol. *21: 1074-1080*.

Hardy, M.E. (2005). Norovirus protein structure and function. FEMS Microbiology Letters *253, 1-8*.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N. et al. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113, 803-809*.

Hartwig, D., Schüttle, C., Warnecke, J., Dorn, I., Hennig, H., Kirchner, H. et al. (2006). The large form of ADAR1 is responsible for enhanced hepatitis delta virus RNA editing in interferon-alpha-stimulated host cells. J. Viral Hepat. *13, 150-157*.

Hatada, E., Takizawa, T. and Fukuda, R. (1992). Specific binding of influenza A virus NS1 protein to the virus minus-sense RNA in vitro. J. Gen. Virol. *73, 17-25*.

Hicks, A.L. and Duffy, S. (2014). Cell tropism predicts long-term nucleotide substitution rates of mammalian RNA viruses. PLoS Pathog. *10, e1003838*.

Ho, S.Y.W., Lanfear, R., Bromham, L., Phillips, M.J., Soubrier, J., Rodrigo, A.G. and Cooper, A. (2011). Time-dependent rates of molecular evolution. Mol. Ecol. *Doi : 10.1111/j.1365-294X.2011.05178.x*

Holland, J.J., de la Torre, J.C., Clarke, D.K. and Duarte, E. (1991). Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. J. Virol. *65(6), 2960-7*.

Holland, J.J., de la Torre, J.C., Steinhauer, D.A., Clarke, D.K., Duarte, E.A. and Domingo, E. (1989). Virus mutation frequencies can be greatly understimated by monoclonal antibody neutralization of virions. J. Virol. *63, 5030-5036*.

Holmes, E.C. (2003). Patterns of intra- and inter-host nonsynonymous variation reveal strong purifying selection in dengue virus. J. Virol. *77, 296-298*.

Holmes, E.C. (2008). Evolutionary history and phylogeography of human viruses. Annu. Rev. Microbiol. *62, 307-328*.

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

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

Holtz, C.M. and Mansky, L.M. (2013). Variation of HIV-1 Mutation Spectra Among Cell Types. J. Virol. *87, 5296*.

Jacks, T., Remington, L., Williams, B.O., Schmitt, E.M., Halachmi, S., Bronson, R.T. et al. (1994). Tumor spectrum analysis in *p53*-mutant mice. Curr. Biol. *4, 1-7*.

Jacobs, B.L. and Langland, J.O. (1996). When two strands are better than one: the mediators and modulators of the cellular responses to double-stranded RNA. Virology *219, 339-249.*

Jayan, G.C. and Casey, J.L. (2002). Increased RNA editing and inhibition of hepatitis delta virus replication by high-level expression of ADAR1 and ADAR2. J. Virol. *76, 3819-3827*.

Jenkins, G.M., Rambaut, A., Pybus, O.G. and Holmes, E.C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. *54, 156-65*.

Jiang, X., Wang, M., Wang, K. and Estes, M.K. (1993). Sequence and genomic organization of Norwalk virus. Virology *195, 51-61*.

Johnson, T. and Barton, N.H. (2002). The effect of deleterious alleles on adaptation in asexual populations. Genetics *162, 395-411*.

Jolivet-Gougeon, A., Kovacs, B., Le Gall-David, S., Le Bars, H., Bousarghin, L., Bonnaure-Maller, M. et al. (2011). Bacterial hypermutation: clinical implications. J. Med. Microbiol. *60, 563-573*.

Jones, M.K., Watanabe, M., Zhu, S., Graves, C.L., Keyes, L.R., Grau, K.R., et al. (2014). Enteric bacteria promote human and mouse norovirus infection of B cells. Science *346, 755.*

Julias, J.G. and Pathak, V.K. (1998). Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. J. Virol. *72, 7941-7949*.

Kearney, C.M., Thomson, M.J. and Roland, K.E. (1999). Genome evolution of tobacco mosaic virus populations during long-term passaging in a diverse range of hosts. Arch. Virol. *144, 1513-26*.

Kew, O.M., Sutter, R.W., de Gourville, E.M., Dowdle, W.R. and Pallansch, M.A. (2005). Vaccine-derived polioviruses and the endgame strategy for global polio eradication. Annu. Rev. Microbiol. *59, 587-635.*

Kim, T., Mudry, R.A., Rexrode, C.A. and Pathak, V.K. (1996). Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. J. Virol. *70, 7594-7602*.

Kim, U., Wang, Y., Sanford, T., Zeng, Y. and Nishikura, K. (1994). Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. Proc. Natl. Acad. Sci. USA *91, 11457-11461*.

Kim, T., Youn, M.Y., Min, B.E., Choi, S.H., Kim, M. and Ryu, K.H. (2005). Molecular analysis of quasispecies of Kyuri green mottle mosaic virus. Virus Res. *110, 161-167*.

Klebe, R.J. and Ruddle, F.H. (1969). Neuroblastoma: cell culture analysis of a differentiating stem cell system. J. Cell. Biol. *43, 69*.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.E. et al. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics *25, 2283-2285.*

Korneeva, V.S. and Cameron, C.E. (2007). Structure-function relationships of the viral RNA-dependent RNA polymerase: fidelity, replication speed, and initiation mechanism determined by a residue in the ribose-binding pocket. J. Biol. Chem. *282, 16135-16145*.

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

Kuiken, T., Holmes, E.C., McCauley, J., Rimmelzwaan, G.F., Williams, C.S. and Grenfell, B.T. (2006). Host species barriers to influenza virus infections. Science *312, 394-397*.

Kuiken, C., Yusim, K., Boykin, L. and Richardson, R. (2005). The Los Alamos HCV Sequence Database. Bioinformatics *21, 379-384.*

Kumar, M. and Carmichael, G.G. (1998). Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. Microbiol. Mol. Biol. Rev. *62, 1415-1434*.

Kurimoto, K., Yabuta, Y., Ohinata, Y. and Saitou, M. (2007). Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. Nat. Protoc. *2, 739-752*.

Kuttan, A. and Bass, B.L. (2012). Mechanistic insights into editing-site specificity of ADARs. PNAS *E3295-E3304.*

Kuzmin, I.V., Novella, I.S., Dietzgen, R.G., Padhi, A. and Rupprecht, C.E. (2009). The rhabdoviruses: biodiversity, phylogenetics, and evolution. Infect. Genet. Evol. *9, 541-553*.

Land, A.M., Ball, T.B., Luo, M., Pilon, R., Sandstrom, P., Embree, J.E. et al. (2008). Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. J. Virol. *82, 8172-8182*.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Gen. Biol. *10.*

Lauring, A.S., Frydman, J. and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. Nat. Rev. Microbiol. *11, 327-336*.

LeClerc, J.E., Li, B., Payne, W.L. and Cebula, T.A. (1996). High mutation frequencies among Escherichia coli and Salmonella pathognes. Science *274, 1208-1211*.

Lecossier, D., Bouchonnet, F., Clavel, F. and Hance, A.J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science *300, 1112*.

Lehmann, K.A. and Bass, B.L. (1999). The importance of internal loops within RNA substrates of ADAR1. J. Mol. Biol. *291, 1-13.*

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

Letchworth, G.J., Rodriguez, L.L. and Del Cbarrera, J. (1999). Vesicular stomatitis. Vet. J. *157, 239-260*.

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

Li, Z., Wolff, K.C. and Samuel, C.E. (2010). RNA adenosine deaminase ADAR1 deficiency leads to increased activation of

protein kinase PKR and reduced vesicular stomatitis virus growth following interferon treatment. Virology *396, 316-322*.

Libersou, S., Albertini, A.A.V, Ouldali, M., Maury, V., Maheu, C., Raux, H. et al. (2010). Distinct structural rearrangements of the VSV glycoprotein drive membrane fusion. J. Cell Biol. *191, 199-210.*

Lichty, B.D., Power, A.T., Stojdl, D.F. and Bell J.C. (2004). Vesicular stomatitis virus: re-inventing the bullet. Trends Mol. Med. *10(5), 210-216*.

Lin, N.S. and Chen, C.C. (1991). Association of bamboo mosaic virus (BoMV) and BoMV-specific electron-dense crystalline bodies with chloroplasts. Phytoplathology *81, 1551-1555*.

Lin, R.J., Liao, C.L., Lin, E. and Lin, Y.L. (2004). Blocking of the alpha interferon-induced Jak-Stat signaling pathway by Japanese encephalitis virus infection. J. Virol. *78, 9285-9294*.

Lochridge, V.P., Jutila, K.L., Graff, J.W. and Hardy, M.E. (2005). Epitopes in the P2 domain of norovirus VP1 recognized by monoclonal antibodies that block cell interactions. J. Gen. Virol. *86, 2799-2806*.

Loeffler, F. and Frosch, P. (1898). Berichte der Kommission zur Erforschung der Maul- und Klauenseuche be idem Institut für Infectionskrankheiten. Part I *23, 371-391.*

Lopman, B.A., Adak, G.K., Reacher, M.H. and Brown, D.W. (2003). Two epidemiologic patterns of norovirus outbreaks: surveillance in England and Wales, 1992-2000. Emerg. Infect. Dis*. 9, 71-7*.

Lopman, B., Vennema, H., Kohli, E., Pothier, P., Sanchez, A. Negredo, A. et al. (2004). Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. Lancet *363, 682-688*.

Lu, Y., Wambach, M., Katze, M.G. and Krug, R.M. (1995). Binding of the influenza virus NS1 protein to double-stranded RNA inhibits the activation of the protein kinase that phosphorylates the elF-2

translation initiation factor. Virology *214, 222-228*.

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

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

Lynch, M. (2010). Evolution of the mutation rate. Trends Genet. 26, 345-352.

Maas, S., Melcher, T., Herb, A., Seeburg, P.H., Keller, W., Krause, S., Higuchi, M. and O'Connell, M.A. (1996). Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. J. Biol. Chem. *271, 12221-12226.*

Maas, S., Patt, S., Schrey, M. and Rich, A. (2001). Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. PNAS *98, 14787-14692.*

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.

Mannion, N.M., Greenwood, S.M., Young, R., Cox, S., Brindle, J., Read, D., et al. (2014). The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. Cell Reports *9, 1482-1494.*

Manrubia, S.C., Domingo, E. and Lazaro, E. (2010). Pathways to extinction: beyond the error threshold. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *365, 1943-1952*.

Manrubia, S.C., and Lázaro E. (2006). Viral evolution. Phys. Life Rev. *3, 65-92.*

Mansky, L.M., and Bernard, L.C. (2000). 3'-Azido-3'-deoxythymidine (AZT) and AZT- resistant reverse transcriptase

can increase in vivo mutation rate of human immunodeficiency virus type 1. J. Virol. *74, 9532-9539.*

Mansky, L.M. and Cunningham, K.S. (2000). Virus mutators and antimutators : roles in evolution, pathogenesis and emergence. Trends Genet. *16, 512-517.*

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, M. (2012). Cutadapt removes adapter sequences from high-throughput sequencing. EMBnet.JournalNorthAmerica, 17. Accessed 2013 September 18.

Martínez, I., Dopazo, J. and Melero, J.A. (1997). Antigenic structure of the human respiratory syncytial virus G glycoprotein and relevance of hypermutation events for the generation of antigenic variants. J. Gen. Virol. *78, 2419-2429*.

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.

Maynard Smith, J. and Szathmáry, E. (1995). The major transitions of evolution. W.H. Freeman and Co., Oxford.

Mead, D.G., Gray, E.W., Noblet, R., Murphy, M.D., Howerth, E.W. and Stallknecht, D.E. (2004). Biological transmission of vesicular stomatitis virus (New Jersey serotype) by Simulium vittatum (Diptera: Simuliidae) to domestic swine (Sus scrofa). J. Med. Entomol. *41, 78-82*.

Mead, D.G., Mare, C.J., and Ramberg, F.B. (1999). Bite transmission of vesicular stomatitis virus (New Jersey serotype) to laboratory mice by Simulium vittatum (Diptera: Simuliidae). J. Med. Entomol. *36, 410-413*.

Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresee, J.S., Shapiro, C., et al. (1999). Food-related illness and death in the United

States. Emerg. Infect. Dis. *5, 607-625*.

Menéndez-Arias, L. (2009). Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. Viruses *1, 1137-1165*.

Metzgar, D. and Wills, C. (2000). Evidence for the adaptive evolution of mutation rates. Cell *101, 581-584*.

Minskaia, E., Hertzig, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B. et al. (2006). Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc. Natl. Acad. Sci. USA *103, 5108-5113*.

Monk, R. J., Malik, F. G., Stokesberry, D. and Evans, L. H. (1992). Direct determination of the point mutation rate of a murine retrovirus. *J. Virol. 66, 3683-9*.

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.

Mounts, A.W., Ando, T., Koopmans, M., Bresee, J.S., Noel, J. and Glass, R.I. (2000). Cold weather seasonality of gastroenteritis associated with Norwalk-viruses. J. Infect. Dis. *181 (Suppl 2), S284-287*.

Moxon, R., Bayliss, C. and Hood, D. (2006). Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu. Rev. Genet. *40, 307-33*.

Murphy, D.F., Dimok, K. and Kang, C.Y. (1991). Numerous transitions in human parainfluenza virus 3 RNA recovered from persistently infected cells. Virology *181, 760-763*.

Narsinh, K.H., Sun, N., Sanchez-Freire, V., Lee, A.S., Almeida, P., Hu, S. et al. (2011). Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. J. Clin. Invest. *121, 1217-1221*.

Neumann, A.U., Lam, N.P., Dahari, H., Gretch, D.R., Wiley, T.E., Layden, T.J. et al. (1998). Hepatitis C viral dynamics in vivo and

the antiviral efficacy of interferon-alpha therapy. Science *282, 103-107.*

Nie, Y., Hammond, G.L. and Yang, J.H. (2007). Double-stranded RNA deaminase ADAR1 increases host susceptibility to virus infection. J. Virol. *81, 917-923*.

Nilsson, M., Hedlund, K.O, Thorhagen, M., Larson, G., Johansen, K., Ekspong A. et al. (2003). Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. J. Virol. *77(24), 13117.*

Ninio, J. (1991). Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. Genetics *129(3), 957-962*.

Nishikura, K., Yoo, C., Kim, U., Murray, J.M., Estes, P.A., Cash, F.E. and Liebhaber, S.A. (1991). Substrate specificity of the dsRNA unwinding/modifying activity. EMBO *10, 3523-3532.*

Novella, I. S. (2003). Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. Curr. Opin. Microbiol. *6(4), 399-405*.

Novella, I.S., Reissig, D.D. and Wilke, C.O. (2004). Density-dependent selection in vesicular stomatitis virus. J. Virol. *78, 5799-5804*.

O'Hara, P.J., Nichol, S.T., Horodyski, F.M., Holland, J.J. (1984). Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. Cell *36, 915-924*.

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

Orr, H.A. (2000). The rate of adaptation in asexuals. Genetics *155, 961-968*.

Ostertag, D., Hoblitzell-Ostertag, T.M. and Perrault, J. (2007).

Overproduction of double-stranded RNA in vesicular stomatitis virus-infected cells activates a constitutive cell-type-specific antiviral response. J. Virol. *81, 503-513*.

Parthasarathi, S., Varela-Echavarría, A., Ron, Y., Preston, B. D. and Dougherty, J. P. (1995). Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. J. Virol. *69, 7991-8000.*

Pathak, V. K. and Temin, H. M. (1990). Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl. Acad. Sci. USA 87, 6019-6023*.

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

Patterson, J.B., Thomis, D.C., Hans, S.L. and Samuel, C.E. (1995). Mechanism of interferon action – double-stranded RNA-specific adenosine-deaminase from human-cells is inducible by alpha-interferon and gamma-interferon. Virology *210, 508-511*.

Peluso, R.W. and Moyer, S.A. (1983). Initiation and replication of vesicular stomatitis virus genome RNA in a cell-free system. Proc. Natl. Acad. Sci. USA *80, 3198-3202*.

Pepin, K.M., Lass, S., Pulliam, J.R.C, Read, A.F. and Lloyd-Smith, J.O. (2010). Identifying genetic markers of adaptation for surveillance of viral host jump. Nat. Rev. Microbiol. *8, 802-813.*

Pereira-Gómez, M. and Sanjuán, R. (2014). Delayed lysis confers resistance to the nucleoside analog 5- fluorouracil and alleviates mutation accumulation in the single-standed DNA bacteriophage ϕX174. J. Virol. *Doi: 10.1128/JVI.02147-13*.

Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M. and Ho, D.D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cells life-span, and viral generation time. Science *271,*

1582-1586.

Pfeiffer, J.K. and Kirkegaard, K. (2003). A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. Proc. Natl. Acad. Sci. USA *100, 7289-7294*.

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., De Miranda, J.R., Schneider, W.L. and Roossinck, M.J. (2007). Environment determines fidelity for an RNA virus replicase. J. Virol. *81, 9072-7.*

Pita, J.S. and Roossinck, M.J. (2013). Mapping viral functional domains for genetic diversity in plants. J. Virol. *87, 790-797*.

Poch, O., Blumberg, B. M., Bougueleret, L. and Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. *71, 1153-1162*.

Polson, A.G. and Bass, B.L. (1994). Preferential selection of adenosine for modification by double-stranded RNA adenosine deaminase. EMBO *13, 5701-5711.*

Prasad, B.V.V., Hardy, M.E., Dokland, T., Bella, J., Rossmann, M.G. and Estes, M.K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. Science *286, 287-290*.

Pybus, O.G., Rambaut, A., Belshaw, R., Freckleton, R.P., Drummond A.J. and Holmes, E.C. (2007). Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. Mol. Biol. Evol. *24(3), 845-852.*

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26, 841-842*.

Rabinovici, R., Kabir, K., Chen, M., Su, Y., Zhang, D., Luo, X. et al.

(2001). ADAR1 is involved in the development of microvascular lung injury. Circ. Res. *88, 1066-1071*.

Rahmeh, A.A., Schenk, A.D., Danek, E.I., Kranzusch, P.J., Liang, B., Walz, T. et al. (2010). Molecular architecture of the vesicular stomatitis virus RNA polymerase. Proc. Natl. Acad. Sci. USA *107, 20075-20080*.

Rand, U., Hillebrand, U., Sievers, S., Willenberg, S. Köster, M., Hauser, H. et al. (2014). Uncoupling of the dynamics of host-pathogen interaction uncovers new mechanisms of viral interferon antagonism at the single-cell level. Nucleic Acids Res. *42, e109*.

Rando, O.J. and Verstrepen, K.J. (2007). Timescales of genetic and epigenetic inheritance. Cell *128, 655-68*.

Ratan, A., Miller, W., Guillory, J., Stinson, J., Seshagiri, S. and Schuster, S.C. (2013). Comparison of sequencing platforms for single nucleotide variant calls in a human sample. PLoS ONE *8, e55089*.

Reis, Jr.J.L., Mead, D., Rodriguez, L.L. and Brown, C.C. (2009). Transmission and pathogenesis of vesicular stomatitis viruses. Braz. J. Vet. Pathol. *2(1), 49-58.*

Roberts, J.D., Bebenek, K. and Kunkel, T.A. (1988). The accuracy of reverse transcriptase from HIV-1. Science *242, 1171-1173*.

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.*

Rodríguez, L.L., Fitch, W.M. and Nichol, S.T. (1996). Ecological factors rather than temporal factors dominate the evolution of vesicular stomatitis virus. Proc. Natl. Acad. Sci. USA *93, 13030-13035*.

Rozen-Gagnon, K., Stapleford, K.A., Mongelli, V., Blanc, H., Failloux, A.B., Saleh, M.C. et al. (2014). Alphavirus mutator variants present

host-specific defects and attenuation in mammalian and insect models. PLoS Pathog. *10, e1003877*.

Rueda, P., Garcia-Barreno, B. and Melero, J.A. (1994). Loss of conserved cysteine residues in the attachment (G) glycoprotein of two human respiratory syncytial virus escape mutants that contain multiple A-G substitutions (hypermutations). Virology, *198, 653-662*.

Rushing, A.E., Sunter, G., Gardiner, W.E., Dute, R.R. and Bisaro, D.M. (1987). Ultrastructural aspects of tomato golden mosaic virus infection in tobacco. Phytopathology *77, 1231-1236*.

Russell, S.J., Peng, K.W. and Bell, J.C. (2012). Oncolytic virotherapy. Nat. Biotechnol. *30, 658-670*.

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*.

Safari, M. and Roossinck, M. (2014). How does the genome structure and lifestyle of a virus affect its population variation? Curr. Opin. Virol. *9, 39-44*.

Samuel, C.E. (2011). Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. Virology *411, 180-193*.

Samuel, C.E. (2001). Antiviral actions of interferon. Clin. Microbiol. Rev. *14, 778-809.*

Sanjuán, R. (2010). Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Phil. Trans. R. Soc. B. *365, 1975-1982*.

Sanjuán, R. (2012). From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. PLoS Pathog. *8, e1002685*.

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., Lázaro, E. and Vignuzzi, M. (2012). Biomedical implications of viral mutation and evolution. Future Virol*. 7, 391–402*.

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

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

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

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*.

Sarkar, S. (1991). Haldane's solution of the Luria-Delbrück distribution. Genetics *127, 257-261*.

Sarkar, S., Ma, W.T. and Sandri, G.V.H (1992). On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. Genetica *85, 173-179*.

Schloemer, R. and Wagner, R. (1975). Mosquito cells infected with vesicular stomatitis virus yield unsialylated virions of low infectivity. J. Virol. *15(4), 1029-1032*.

Schmid, A., Spielhofer, P., Cattaneo, R., Baczko, K., ter Meulen, V. and Billeter, M.A. (1992). Subacute sclerosing panencephalitis is typically characterized by alterations in the fusion protein cytoplasmic domain of the persisting measles virus. Virology *188, 910-915*.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27, 863-864.*

Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morphol. *27, 353-365.*

Schrag, S.J., Rota, P.A and Bellini, W.J. (1999). Spontaneous mutation rate of measles virus: direct estimation based on mutations conferring monoclonal antibody resistance. J. Virol. *73, 51-54*.

Schubert, M., Harmison, G.G. and Meier, E. (1984). Primary structure of the vesicular stomatitis virus polymerase (L) gene : evidence for a high frequency of mutations . J. Virol. *51(2), 505-514.*

Schulte, M.B. and Andino, R. (2014). Single-cell analysis uncovers extensive biological noise in poliovirus replication. J. Virol. *88, 6205-6212*.

Schulte, M.B., Draghi, J.A., Plotkin, J.B. and Andino, R. (2015). Experimentally guided models reveal replication principles that shape the mutation distribution of RNA viruses. eLife *4, e03753*.

Schneider, W.L. and Roossinck, M.J. (2000). Evolutionary related sindbis-like plant viruses maintain different levels of population diversity in a common host. J. Virol. *74(7), 3130*.

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

Sedivy, J.M., Capone, J.P., RajBhandary, U.L. and Sharp, P.A. (1987). An inducible mammalian amber suppressor: propagation of a poliovirus mutant. Cell *50, 379-389*.

Seronello, S., Montanez, J., Presleigh, K., Barlow, M., Park, S.B. and Choi, J. (2011). Ethanol and reactive species increase basal sequence heterogeneity of hepatitis C virus and produce variants with reduced susceptibility to antivirals. PLoS ONE *6, e27436*.

Sharmeen, L., Bass, B., Sonenberg, N., Weintraub, H. and Groudine,

M. (1991). Tat-dependent adenosine-to-inosine modification of wild-type transactivation response RNA. Proc. Natl. Acad. Sci. USA *88, 8096-8100*.

Siebenga, J.J., Vennema, H., Zheng, D.P., Vinjé, J., Lee, B.E., Pang, X.L. et al. (2009). Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001-2007. J. Infect. Dis. *200(5), 802-812*.

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.

Simon-Loriere, E., Galetto, R., Hamoudi, M., Archer, J., Lefeuvre, P., Martin, D.P. et al. (2009). Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. PLoS Pathog. *5, e1000418*.

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

Sniegowski, P.D., Gerrish, P.J. and Lenski, R.E. (1997). Evolution of high mutation rates in experimental populations of E. coli. Nature *387, 703-705*.

Stech, J., Xiong, X., Scholtissek, C. and Webster, R. G. (1999). Independence of evolutionary and mutational rates after transmission of avian influenza viruses to swine. *J. Virol. 73, 1878-84*.

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*.

Stern, A., Bianco, S., Te Yeh, M., Wrigth, C., Butcher, K., Tang, C. et al. (2014). Costs and benefits of mutational robustness in RNA viruses. Cell Rep. *8(4), 1026-1036.*

Stewart, F.M. (1994). Fluctuation tests: how reliable are the

estimates of mutation rates? Genetics *137, 1139-1146*.

Stewart, F.M., Gordon, D.M. and Levin, B.R. (1989). Fluctuation analysis: the probability distribution of the number of mutants under different conditions. Genetic *124, 175-185*.

Stojdl, D.F., Lichty, B., Knowles, S., Marius, R., Atkins, H., Sonenberg, N. et al. (2000). Exploiting tumor-specific defects in the interferon pathway with a previously unknown oncolytic virus. Nat. Med. *6, 821-825*.

Streitenfeld, H., Boyd, A., Fazakerley, J.K., Bridgen, A., Elliott, R.M. and Weber, F. (2003). Activation of PKR by Bunyamwera virus is independent of the viral interferon antagonist NSs. J. Virol. *77, 5507-5511*.

Suárez, P., Valcárcel, J. and Ortín, J. (1992). Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator mutants. J. Virol. *66, 2491-2494*.

Suspène, R., Renard, M., Henry, M., Guétard, D., Puyraimond-Zemmour, D., Billecocq, A. et al. (2008). Inversing the natural hydrogen bonding rule to selectively amplify GC-rich ADAR edited RNAs. Nucl. Acids Res. *36(12), e72*.

Sutter, G. and Moss, B. (1992). Nonreplicating vaccinia vector efficiency expresses recombinant genes. Proc. Natl. Acad. Sci. USA *89, 10847-10851*.

Takeuchi, O. and Akira, S. (2010). Pattern recognition receptors and inflammation. Cell *140, 805-820.*

Tan, M., Huang, P., Meller, J., Zhong, W., Farkas, T. and Jiang, X. (2003). Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. J. Virol. *77(23), 12562-12571*.

Tan, M. and Jiang, X. (2005). Norovirus and its histo-blood group antigen receptors: an answer to a historical puzzle. Trends Microbiol. *13, 285-293*.

Taylor, D.R., Puig, M., Darnell, M.E.R., Mihalik, K. and Feinstone,

S.M. (2005). New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. J. Virol. *79, 6291-6298*.

Teng, G. and Papavasiliou, F.N. (2007). Immunoglobulin somatic hypermutation. Annu. Rev. Genet. *41, 107-120*.

tenOever, B.R., Ng, S.L., Chua, M.A., McWhirter, S.M., Garcia-Sastre, A. and Maniatis, T. (2007). Multiple functions of the IKK-related kinase IKKepsilon in interferon-mediated antiviral immunity. Science *315, 1274-1278*.

tenOever, B. R., Servant, M. J., Grandvaux, N., Lin, R. and Hiscott, J. (2002). Recognition of the measles virus nucleocapsid as a mechanism of IRF-3 activation. J. Virol. *76, 3659-3669*.

tenOever, B. R., Sharma S., Zou, W., Sun Q., Grandvaux, N., Julkunen, I., et al. (2004). Activation of TBK1 and IKKε kinases by vesicular stomatitis virus infection and the role of viral ribonucleoprotein in the development of interferon antiviral immunity. J. Virol. *78, 10636-10649*.

Tesh, R.B., Chaniotis, B.N. and Johnson, K.M. (1972). Vesicular stomatitis virus (Indiana serotype): transovarial transmission by phlebotomine sandlies. Science *175, 1477-1479*.

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*.

Timm, A. and Yin, J. (2012). Kinetics of virus production from single cells. Virology *424, 11-17*.

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

Tybulewicz, V.L.J, Crawford, C.E., Jackson, P.K., Bronson, R.T. and Mulligan, R.C. (1991). Neonatal lethality and lymphopenia in mice with a homozygous disruption of the *c-abl* proto-oncogene. Cell *65, 1153-1163*.

Ulferts, R. and Ziebuhr, J. (2011). Nidovirus ribonucleases: Structures and functions in viral replication. RNA. Biol. 8, 295-304.

Valente, L. and Nishikura, K. (2005). ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. Prog. Nucleic Acid Res. Mol. Biol. *79, 299-338*.

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*.

Vignuzzi, M., Wendt, E. and Andino, R. (2008). Engineering attenuated virus vaccines by controlling replication fidelity. Nat. Med. *14, 154-161*.

Wahl, L.M., Gerrish, P.J. and Saika-Voivod, I. (2002). Evaluating the impact of population bottlenecks in experimental evolution. Genetics *162, 961-971*.

Wang, D. and Bodovitz, S. (2010). Single cell analysis: the new frontier in "omics". Trends Biotechnol. *28, 281-290*.

Wang, I.X., So, E., Devlin, J.L., Wu, M. and Cheung, V. (2013). ADAR regulates RNA editing, transcript stability, and gene expression. Cell Rep. *5, 849-860.*

Wang, Q. and Garmichael, G.G. (2004). Effects of length and location on the cellular response to double-stranded RNA. Microbiol. Mol. Biol. Rev. *68, 432-452.*

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

Weber, F., Kochs, G. and Haller, O. (2004). Inverse interference: how viruses fight the interferon system. Viral Immunol. *17, 498-515*.

Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R. and Paludan, S.R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J. Virol. *80, 5059-5064.*

Weeks, S.A., Lee, C.A., Zhao, Y., Smidansky, E.D., August, A., Arnold, J.J. et al. (2012). A Polymerase mechanism-based strategy for viral attenuation and vaccine development. J. Biol. Chem. *287, 31618-31622*.

Whelan, S.P.J. (2008). Vesicular stomatitis virus. In: Mahy BWJ, van Regenmortel MHV, editors. Encyclopedia of Virology. Elsevier. *pp. 291-299*.

Whelan, S.P., Ball, L.A., Barr, J.N. and Wertz, G.T. (1995). Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. Proc. Natl. Acad. Sci. USA *92, 8388-8392*.

Wimmer, E., Hellen, C.U.T and Cao, X. (1993). Genetics of poliovirus. Annu. Rev. Genet. *27, 353-436*.

Wobus, C.E., Karst, S.M., Thackray, L.B., Chang, K.O., Sosnovtsev, S.V., Belliot, G. et al. (2004). Replication of norovirus in cell culture reveals a tropism for dendritic cells and macrophages. PLoS Biol. *2(12), e432*.

Woelk, C.H. and Holmes, E.C. (2002). Reduced positive selection in vector-borne RNA viruses. Mol. Biol. Evol. *19, 233-2336*.

Woolhouse, M. and Gaunt, E. (2007). Ecological origins of novel human pathogens. Crit. Rev. Microbiol. *33, 231-242*.

World Health Statistics (2013).
www.who.int/gho/publications/world_health_statistics/2013

Wyatt, L.S, Moss, B. and Rozenblatt, S. (1995). Replication-deficient vaccinia virus encoding bacteriophage T7 RNA polymerase for transient gene expression in mammalian cells. Virology *210(1), 202-205.*

Yoshimoto, N., Kida, A., Jie, X., Kurokawa, M., Lijima, M., Niimi, T. et al. (2013). An automated system for high-throughput single cell-based breeding. *Sci. Rep. 3, 1191*.

Zahn, R.C., Schelp, I., Utermöhlen, O. and von Laer, D. (2007). A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. J. Virol. *81, 457-464*.

Zárate, S., and Novella, I. S. (2004). Vesicular Stomatitis Virus Evolution during Alternation between Persistent Infection in Insect Cells and Acute Infection in Mammalian Cells Is Dominated by the Persistence Phase. J. Virol. *78(22), 12236-12242*.

Zheng, Q. (1999). Progress of a half century in the study of the Luria-Delbrück distribution. Math. Biosci. *162, 1-32*.

Zhu, Y., Yongky, A. and Yin, J. (2009). Growth of an RNA virus in single cells reveals a broad fitness distribution. Virology *385, 39-46.*

SUPPLEMENTARY INFORMATION

Supplementary 1. Fluctuation tests of VSV in mammalian, insect and mosquito cells.

Fluctuation tests in wild-type MEFs

	Test 1	**Test 2**	**Test 3**
N_i (pfu)	439 ± 94	124 ± 15	160 ± 11
N_f (pfu)	49812 ± 5985	12000 ± 2021	36075 ± 3162
Total cultures	24	24	24
With no MAR	15	21	14
With 1 MAR	5	3	10
With 2 MARs	2	0	0
With >2 MARs	2	0	0
Fraction no MAR (P_0)	0.625	0.875	0.583
Plating efficiency (z)[a]	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.04
Corrected P_0 (Q_0)	0.607	0.868	0.564
Mutation rate (m)	1.01×10^{-5}	1.20×10^{-5}	1.60×10^{-5}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used)

Fluctuation tests in p53-/- MEFs

	Test 1	**Test 2**	**Test 3**
N_i (pfu)	33 ± 10	368 ± 38	160 ± 11
N_f (pfu)	40929 ± 6820	9725 ± 1619	40650 ± 2602
Total cultures	24	24	24
With no MAR	20	21	19
With 1 MAR	4	2	3
With 2 MARs	0	1	2
With >2 MARs	0	0	0
Fraction no MAR (P_0)	0.833	0.875	0.792
Plating efficiency (z)[a]	1.01 ± 0.02	1.01 ± 0.02	1.01 ± 0.02
Corrected P_0 (Q_0)	0.833	0.875	0.792
Mutation rate (m)	4.46×10^{-6}	1.43×10^{-5}	5.77×10^{-6}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used).

Fluctuation tests in CT26 colon cancer cells

	Test 1	**Test 2**	**Test 3**
N_i (pfu)	465 ± 52	33 ± 10	123 ± 15
N_f (pfu)	42375 ± 5214	6675 ± 1217	134700 ± 4435
Total cultures	24	24	24
With no MAR	22	23	19
With 1 MAR	2	1	3
With 2 MARs	0	0	1
With >2 MARs	0	0	1
Fraction no MAR (P_0)	0.917	0.958	0.792
Plating efficiency (z)[a]	0.29 ± 0.05	0.29 ± 0.05	0.29 ± 0.05
Corrected P_0 (Q_0)	0.741	0.863	0.447
Mutation rate (m)	7.16×10^{-6}	2.23×10^{-5}	5.99×10^{-6}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used).

Fluctuation tests in Neuro-2a neuroblastoma cells

	Test 1	**Test 2**	**Test 3**
N_i (pfu)	124 ± 15	160 ± 11	215 ± 36
N_f (pfu)	28913 ± 4339	61275 ± 3417	29513 ± 3244
Total cultures	24	24	24
With no MAR	16	16	20
With 1 MAR	3	2	4
With 2 MARs	1	6	0
With >2 MARs	4	0	0
Fraction no MAR (P_0)	0.667	0.667	0.833
Plating efficiency (z)[a]	0.85 ± 0.06	0.85 ± 0.06	0.85 ± 0.06
Corrected P_0 (Q_0)	0.621	0.621	0.807
Mutation rate (m)	1.65×10^{-5}	7.80×10^{-6}	7.34×10^{-6}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used).

Fluctuation tests in BHK-21 cells under hypoxia (1% O_2)

	Test 1	**Test 2**	**Test 3[b]**
N_i (pfu)	355 ± 35	267 ± 18	267 ± 18
N_f (pfu)	49875 ± 175	19625 ± 1735	20750 ± 1652
Total cultures	24	24	24
With no MAR	15	19	16
With 1 MAR	4	4	6
With 2 MARs	3	0	2
With >2 MARs	2	1	0
Fraction no MAR (P_0)	0.625	0.792	0.667
Plating efficiency (z)[a]	0.51± 0.04	0.51± 0.04	0.51± 0.04
Corrected P_0 (Q_0)	0.398	0.633	0.452
Mutation rate (m)	1.90×10^{-5}	2.36×10^{-5}	3.88×10^{-5}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used).

[b]: performed the same day and from the same stock as test 1.

Fluctuation tests in BHK-21 cells at 28°C

	Test 1	**Test 2[b]**	**Test 3[b]**	**Test 4**
N_i (pfu)	327 ± 11	327 ± 11	327 ± 11	295 ± 12
N_f (pfu)	17667 ± 3781	13967± 4151	4900 ± 1052	4600 ± 708
Total cultures	24	24	24	24
With no MAR	15	14	17	20
With 1 MAR	3	2	5	4
With 2 MARs	3	4	0	0
With >2 MARs	3	4	2	0
Fraction no MAR (P_0)	0.625	0.583	0.708	0.833
Plating efficiency (z)[a]	0.85 ± 0.07	0.85 ± 0.07	0.85 ± 0.07	0.91 ± 0.05
Corrected P_0 (Q_0)	0.575	0.530	0.667	0.815
Mutation rate (m)	3.19×10^{-5}	4.65×10^{-5}	8.87×10^{-5}	4.76×10^{-5}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used)

[b]: performed the same day and from the same stock as test 1.

Fluctuation tests in S2 cells

	Test 1	**Test 2**	**Test 3**
N_i (pfu)	358 ± 11	358 ± 11	322 ± 10
N_f (pfu)	19025 ± 898	6775 ± 724	13080 ± 593
Total cultures	24	24	24
With no MAR	23	23	23
With 1 MAR	1	1	0
With 2 MARs	0	0	0
With >2 MARs	0	0	1
Fraction no MAR (P_0)	0.96	0.96	0.96
Plating efficiency (z)[a]	1.07 ± 0.07	1.07 ± 0.07	1.07 ± 0.07
Corrected P_0 (Q_0)	0.958	0.958	0.958
Mutation rate (m)	2.28×10^{-6}	6.63×10^{-6}	3.34×10^{-6}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used).

Fluctuation tests in sf21 cells

	Test 1	**Test 2**	**Test 3**
N_i (pfu)	295 ± 12	223 ± 26	312 ± 32
N_f (pfu)	6970 ± 174	14400 ± 898	17033 ± 652
Total cultures	24	24	24
With no MAR	23	22	22
With 1 MAR	1	2	1
With 2 MARs	0	0	1
With >2 MARs	0	0	0
Fraction no MAR (P_0)	0.958	0.917	0.917
Plating efficiency (z)[a]	1.09 ± 0.04	1.09 ± 0.04	1.09 ± 0.04
Corrected P_0 (Q_0)	0.958	0.917	0.917
Mutation rate (m)	6.38×10^{-6}	6.14×10^{-6}	5.20×10^{-6}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used).

Fluctuation tests in C6/36 cells

	Test 1	**Test 2**	**Test 3**
N_i (pfu)	295 ± 12	223 ± 26	312 ± 32
N_f (pfu)	10100 ± 410	9800 ± 546	10833 ± 499
Total cultures	24	24	24
With no MAR	22	23	22
With 1 MAR	1	0	2
With 2 MARs	1	1	0
With >2 MARs	0	0	0
Fraction no MAR (P_0)	0.916	0.958	0.916
Plating efficiency (z)[a]	0.88 ± 0.02	0.88 ± 0.02	0.88 ± 0.02
Corrected P_0 (Q_0)	0.906	0.953	0.906
Mutation rate (m)	1.01×10^{-5}	5.01×10^{-6}	9.32×10^{-6}

[a]: relative to fluctuation tests performed in BHK-21 cells (z = 1 was used).

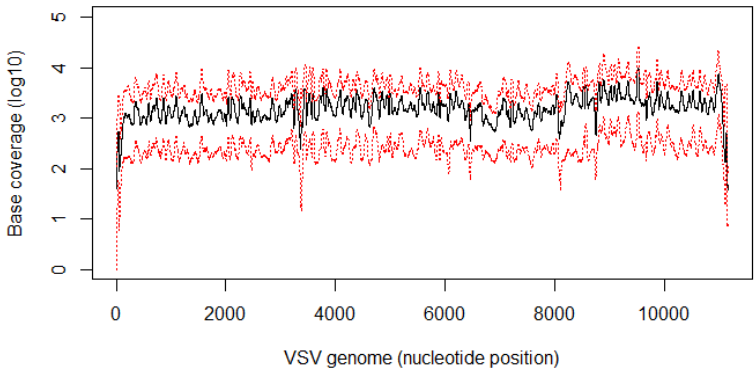
Supplementary 2. Specific primers used to PCR-amplify and sequence molecular clones of VSV from BHK-21 cells.

Gene	Primer sequence and orientation
P	3'-GGCAAGTATGCTAAGTCAGAG-5'
	5'-GAGACATTCGTCCGTTACC-3'
G	3'-ACATTCCATCCGATCCTTCAC-5'
	5'-CGGAGAACCAAGAATAGTCCAATG-3'
L	3'-GCTCTCAATCAAATGGTTTC-5'
	5'-GTCCTCACAATCAAATCATC-3'

Supplementary 3. Sequence and orientation of specific primers used to PCR-amplify VSV plaques sequenced by SOLID.

Primer name	Primer sequence and orientation
VSV-1f	3'-ACGAAGACAAACAAACCA-5'
VSV-3441r	5'-GGAAGCCTGGATTCAGCCAAGTTC-3'
VSV-3361f	3'-ACATTCCATCCGATCCTTCAC-5'
VSV-6444r	5'-GCAGTTAGATCGTCCGCCATT-3'
VSV-6052f	3'-TGGAGATAAATGGCATGAAC-5'
VSV-9683r	5'- TATAATGAGCGCCAGTTG-3'
VSV-8079f	3'-AGAGACTCCTTGTGCACCATGTAAC-5'
VSV-11140r	5'-CGAAGACCACAAAACCAGAT-3'

Supplementary 4. SOLID sequencing base coverage throughout the VSV genome. The average base coverage was computed after read quality control of the 90 libraries. The black line shows the mean base coverage whereas the red lines represent the upper and lower base coverage at each position. Since the 3'- and 5'-end primer sequences were removed during the read quality control, the base coverage at the extreme sites of the genome is lowest.



Supplementary 5. List of all SNPs found in the 90 cells. The type of mutation is indicated for each cell, as the number of plaques sharing the mutation. Non-coding mutations are indicated by a "-" symbol. Stop codons are symbolized by "*". Cells 17, 45, 60, 65 and 90 with no mutations are not indicated in this table.

Cell	Mutation	Plaque	Gene	Codon change	AA change
1	T1832G	4	P	GTT -> GGT	V146G
	T3894A	10	G	TCA -> ACA	S273T
2	T930C	6	N	TAT -> TAC	Y289Y
	A6150C	9	L	AAT -> ACT	N473T
	A6757G	2	L	GGA -> GGG	G675G
	A7573G	2	L	GTA -> GTG	V947V
	C7794T	6	L	ACT -> ATT	T1021I
	T8038C	4	L	ACT -> ACC	T1102T
	A8608G	2	L	ACA -> ACG	T1292T
	C8932T	1	L	TTC -> TTT	F1400F
	A8962C	2	L	AGA -> AGC	R1410S
	A8965C	2	L	GCA -> GCC	A1411A
	G9284T	7	L	GTC -> TTC	V1518F
3	T2952C	10	-	-	-
4	T4552A	2	G	TTA -> TAA	L492*
	A4553T	2	G	TTA -> TTT	L492F
	A4567C	2	G	AAA -> ACA	K497T
	G11041A	4	L	GAG -> GAA	E2103E
5	T930C	6	N	TAT -> TAC	Y289Y
	G1790A	4	P	GGA -> GAA	G132E
	A2243G	4	M	-	-
	T2327C	2	M	CCT -> CCC	P26P
	T3894A	10	G	TCA -> ACA	S273T
	T7846C	1	L	ACT -> ACC	T1038T
	A7978G	6	L	TCA -> TCG	S1082S
	T8038C	4	L	ACT -> ACC	T1102T
	C9074T	3	L	CCA -> TCA	P1448S
	T9926C	4	L	TAT -> CAT	Y1732H
	T2184C	2	P	TAT -> TAC	Y263Y
	A2711G	2	M	GTA -> GTG	V154V

6	A5161G	5	L	AAA -> AAG	K143K
	C5621T	1	L	CTG -> TTG	L297L
	A8608G	2	L	ACA -> ACG	T1292T
	A9695G	3	L	ATA -> GTA	I1655V
	A10184G	5	L	ACG -> GCG	T1818A
7	G1790A	4	P	GGA -> GAA	G132E
	T3894A	10	G	TCA -> ACA	S273T
	A7978G	6	L	TCA -> TCG	S1082S
	G9984A	2	L	GGC -> GAC	G1751D
8	C3506A	1	G	GTC -> GTA	V143V
	T3894A	10	G	TCA -> ACA	S273T
	G5557A	1	L	GTG -> GTA	V275V
	A6757G	2	L	GGA -> GGG	G675G
	A7573G	2	L	GTA -> GTG	V947V
	T10000C	4	L	ATT -> ATC	I1756I
9	T930C	6	N	TAT -> TAC	Y289Y
	A2925C	1	M	AGC -> CGC	S226R
	C5621T	1	L	CTG -> TTG	L297L
	C6712T	2	L	AAC -> AAT	N660N
	C8308T	2	L	CCC -> CCT	P1192P
10	A2925C	1	M	AGC -> CGC	S226R
	T3894A	10	G	TCA -> ACA	S273T
	G6523A	3	L	GCG -> GCA	A597A
	C6712T	2	L	AAC -> AAT	N660N
	A7573G	2	L	GTA -> GTG	V947V
11	C1743T	2	P	GAC -> GAT	D116D
	G1790A	3	P	GGA -> GAA	G132E
	T2384C	5	M	TAT -> TAC	Y45Y
	T4701G	1	G	-	-
12	C1368T	3	N	-	-
	T1927G	3	P	TAT -> GAT	Y178D
	A2736C	1	M	AAT -> CAT	N163H
	G4151A	1	G	AGG -> AGA	R358R
	C4804T	1	L	TTC -> TTT	F24F
	G10240A	4	L	ATG -> ATA	M1836I
13	C2405T	1	M	GAC -> GAT	D52D
	A4152C	2	G	ATG -> CTG	M359L
	A10813G	2	L	TCA -> TCG	S2027S

14	C2966A	1	M	-	-
	T6728C	2	L	TCA -> CCA	S666P
15	T8218C	1	L	TCT -> TCC	S1162S
16	G297A	2	N	AAG -> AAA	K78K
	T2580C	1	M	TTG -> CTG	L111L
	C2758T	10	M	ACG -> ATG	T170M
	A4752G	1	L	GAG -> GGG	E7G
	A4967C	1	L	ACA -> CCA	T79P
	A6150C	9	L	AAT -> ACT	N473T
18	T5901C	9	L	CTT -> CCT	L390P
	T7081G	1	L	CGT -> CGG	R783R
19	C3506A	1	G	GTC -> GTA	V143V
	T4965G	2	L	TTA -> TGA	L78*
	A4966C	1	L	TTA -> TTC	L78F
	A10234G	1	L	GTA -> GTG	V1834V
20	A6251G	1	L	AAA -> GAA	K507E
21	T10131A	7	L	TTT -> TAT	F1800Y
22	G2812T	1	M	AGG -> ATG	R188M
	G5129A	1	L	GGC -> AGC	G133S
	A6757G	1	L	GGA -> GGG	G675G
	A10499G	5	L	AGA -> GGA	R1923G
23	A2353G	4	M	TAT -> TGT	Y35C
24	A2329G	10	M	TAT -> TGT	Y27C
	T2330G	10	M	TAT -> TAG	Y27*
	A10869G	1	L	CAA -> CGA	Q2046R
25	T1338C	6	N	-	-
	T4485G	3	G	TTT -> GTT	F470V
	A4714G	1	-	-	-
	T7423G	3	L	CTT -> CTG	L897L
26	T2184C	5	P	TAT -> TAC	Y263Y
	A2711G	5	M	GTA -> GTG	V154V
	T2866C	6	M	TTA -> TCA	L206S
	A2857G	2	M	GAG -> GGG	E203G
	T4485G	3	G	TTT -> GTT	F470V
	T5671A	10	L	CAT -> CAA	H313Q
27	T1338C	6	N	-	-
	G1790A	2	P	GGA -> GAA	G132E
	T3894A	10	G	TCA -> ACA	S273T

28	A2370G	8	M	ATT -> GTT	I41V
	A2771C	1	M	ACA -> ACC	T174T
	T3894A	10	G	TCA -> ACA	S273T
	A4714G	1	-	-	-
	A6757G	1	L	GGA -> GGG	G675G
	A7573G	2	L	GTA -> GTG	V947V
29	C1747A	1	P	CAT -> AAT	H118N
	A2925C	1	M	AGC -> CGC	S226R
	A5549G	1	L	AAA -> GAA	K273E
	C6712T	1	L	AAC -> AAT	N660N
	T8426C	2	L	TCT -> CCT	S1232P
	G8737A	4	L	TCG -> TCA	S1335S
	T10429C	4	L	ATT -> ATC	I1899I
30	T1510G	1	P	TTG -> GTG	L39V
	A2925C	1	M	AGC -> CGC	S226R
	T3071C	7	G	-	-
	T3894A	10	G	TCA -> ACA	S273T
	T5996C	4	L	TTT -> CTT	F422L
	C6712T	1	L	AAC -> AAT	N660N
	T6847A	2	L	ACT -> ACA	T705T
	A8023G	2	L	ACA -> ACG	T1097T
31	G359A	9	N	GGG -> GAG	G99E
	C1132T	4	N	CCA -> TCA	P357S
	T2831C	1	M	AAT -> AAC	N194N
32	C447T	5	N	AGC -> AGT	S128S
	T3878C	1	G	CCT -> CCC	P267P
	T9811C	3	L	AGT -> AGC	S1693S
33	T786C	1	N	TCT -> TCC	S241S
	T4542G	1	G	TGC -> GGC	C489G
	A5390C	7	L	AAG -> CAG	K220Q
	T5401C	1	L	GAT -> GAC	D223D
	A6931C	2	L	TTA -> TTC	L733F
	T9815G	2	L	TTA -> GTA	L1695V
	T10210C	5	L	TTT -> TTC	F1826F
	A10979C	1	L	ATC -> CTC	I2083L
34	G360A	7	N	GGG -> GGA	G99G
	G361A	8	N	GAT -> AAT	D100N
	T1044A	1	N	CTT -> CTA	L327L

	C3683A	1	G	ATC -> ATA	I202I
	C3740A	2	G	TTC -> TTA	F221L
	C4245A	4	G	CTG -> ATG	L390M
	C8349A	1	L	TCT -> TAT	S1206Y
35	C5984T	2	L	CAT -> TAT	H418Y
	A6069C	7	L	GAA -> GCA	E446A
	A6070C	5	L	GAA -> GAC	E446D
	T10210C	5	L	TTT -> TTC	F1826F
	T10532A	9	L	TAT -> AAT	Y1934N
36	T2330A	10	M	TAT -> TAA	Y27*
	C2833T	8	M	TCT -> TTT	S195F
	G3485T	3	G	ACG -> ACT	T136T
	A3620C	1	G	ACA -> ACC	T181T
	C4636A	1	G	-	-
	C5738A	1	L	CAT -> AAT	H336N
	A6069G	4	L	GAA -> GGA	E446G
	A6070G	5	L	GAA -> GAG	E446E
	C6071T	8	L	CTT -> TTT	L447F
	A7601G	1	L	ATG -> GTG	M957V
37	T2184C	5	P	TAT -> TAC	Y263Y
	G4249A	9	G	AGG -> AAG	R391K
	C4968A	9	L	ACA -> AAA	T79K
	A7595G	1	L	ATC -> GTC	I955V
	A8671G	9	L	TCA -> TCG	S1313S
	G8737A	4	L	TCG -> TCA	S1335S
38	G4025A	10	G	GTG -> GTA	V316V
	T5223C	1	L	TTA -> TCA	L164S
	C8547T	4	L	ACG -> ATG	T1272M
39	C3506A	1	G	GTC -> GTA	V143V
	A6190T	2	L	AAA -> AAT	K486N
	T6507C	1	L	TTG -> TCG	L592S
	T6695G	1	L	TTG -> GTG	L655V
	C8837T	5	L	CTA -> TTA	L1369L
40	A5390C	7	L	AAG -> CAG	K220Q
	C6071T	5	L	CTT -> TTT	L447F
	C8547A	2	L	ACG -> AAG	T1272K
	A123G	3	N	GCA -> GCC	A20A
	G599A	3	N	CGT -> CAT	R179H

41	A5303G	3	L	AGA -> GGA	R191G
	T5671A	9	L	CAT -> CAA	H313Q
	T8038C	1	L	ACT -> ACC	T1102T
	G9371A	3	L	GAG -> AAG	E1547K
	A10945G	2	L	TCA -> TCG	S2071S
42	C5495A	1	L	CTG ATG	L255M
	G5806T	4	L	TGG -> TGT	W358C
43	T3894A	8	G	TCA -> ACA	S273T
	T4505G	1	G	ATT -> ATG	I476M
	A7639T	7	L	GTA -> GTT	V969V
	G10591A	1	L	CCG -> CCA	P1953P
44	G1422A	1	P	GAG -> GAA	E9E
	C5119A	1	L	ACC -> ACA	T129T
	G6280T	2	L	AAG -> AAT	K513N
	A7700C	1	L	ATA -> CTA	I990L
46	T576C	1	N	TTT -> TTC	F171F
	T2330A	9	M	TAT -> TAA	Y27*
	C4070T	1	G	GTC -> GTT	V331V
	C4137A	7	G	CCA -> ACA	P354T
	A4983C	4	L	AAT -> ACT	N84T
	C5473A	3	L	TCC -> TCA	S247S
	A6070C	1	L	GAA -> GAC	E446D
	A7109G	1	L	AGA -> GGA	R793G
47	G361A	5	N	GAT -> AAT	D100N
	A2552G	2	M	AAA -> AAG	K101K
	A3040G	1	-	-	-
	G4896T	3	L	AGG -> ATG	R55M
	C4971A	2	L	TCA -> TAA	S80*
	C5126A	1	L	CGC -> AGC	R132S
	A6069G	7	L	GAA -> GGA	E446G
	A7038C	2	L	CAA -> CCA	Q769P
48	G1051T	10	N	GCA -> TCA	A330S
	C2780T	8	M	ATC -> ATT	I177I
	C4668A	2	G	-	-
	C4968A	9	L	ACA -> AAA	T79K
	A4983C	4	L	AAT -> ACT	N84T
	A6070G	7	L	GAA -> GAG	E446E
	T7285C	1	L	GAT -> GAC	D851D

	A10615C	2	L	GCA -> GCC	A1961A
49	A1362G	6	N	-	-
	T3894A	8	G	TCA -> ACA	S273T
	G4249A	2	G	AGG -> AAG	R391K
	A5629G	9	L	AAA -> AAG	K299K
	G5834T	1	L	GGA -> TGA	G368*
	G6380A	1	L	GAA -> AAA	E550K
	T8762C	1	L	TAT -> CAT	Y1344H
	T10383C	2	L	TTG -> TCG	L1884S
50	C5476A	1	L	ATG -> ATA	M248I
51	T5446C	2	L	ATT -> ATC	I238I
	A8962C	10	L	AGA -> AGC	R1410S
	A8965C	10	L	GCA -> GCC	A1411A
52	T4803C	4	L	TTC -> TCC	F24S
	G7879A	1	L	AGG -> AGA	R1049R
	A8631G	1	L	AAG -> AGG	K1300R
53	C102T	3	N	GTC -> GTT	V13V
	G2618A	1	M	TTG -> TTA	L123L
	A4628C	3	G	-	-
54	A1149G	1	N	GGA -> GGG	G362G
	C2405T	1	M	GAC -> GAT	D52D
	G7420T	1	L	TTG -> TTT	L896F
	G10549T	1	L	TCG -> TCT	S1939S
55	T1694C	2	P	GTG -> GCG	V100A
	T2330C	1	M	TAT -> TAC	Y27Y
	A4966C	1	L	TTA -> TTC	L78F
	A4967C	3	L	ACA -> CCA	T79P
	G6003A	1	L	AGT -> AAT	S424N
	A6150C	4	L	AAT -> ACT	N473T
	A8754G	2	L	AAA -> AGA	K1341R
	A9950G	1	L	AGG -> GGG	R1740G
56	A522G	1	N	AGA -> AGG	R153R
	A6190T	1	L	AAA -> AAT	K486N
57	C752A	5	N	ACA -> AAA	T230K
	A4628C	3	G	-	-
	A8545G	1	L	GCA -> GCG	A1271A
	G9290T	1	L	GGA -> TGA	G1520*
	T2371A	1	M	ATT -> AAT	I41N

58	A4969C	1	L	ACA -> ACC	T79T
	A5629G	9	L	AAA -> AAG	K299K
59	A1362G	6	N	-	-
	A4983C	4	L	AAT -> ACT	N84T
	G5969A	2	L	GGA -> AGA	G413R
	T6223G	1	L	ACT -> ACG	T497T
	G8543A	3	L	GCA -> ACA	A1271T
	T9682A	5	L	TAT -> TAA	Y1650*
61	A1904G	1	P	AAG -> AGG	K170R
	C5313A	1	L	TCT -> TAT	S194Y
	G8737A	6	L	TCG -> TCA	S1335S
62	A2329G	8	M	TAT -> TGT	Y27C
	T2330G	9	M	TAT -> TAG	Y27*
	C2768T	4	M	CTC -> CTT	L173L
63	G360A	4	N	GGG -> GGA	G99G
	T2330G	8	M	TAT -> TAG	Y27*
	T2869G	1	M	ATG -> AGG	M207R
	T3894A	9	G	TCA -> ACA	S273T
	C6637T	9	L	ATC -> ATT	I635I
	A10182G	9	L	AAG -> AGG	K1817R
	C11006T	7	L	CGG -> TGG	R2092W
64	A6190T	5	L	AAA -> AAT	K486N
	C6616T	8	L	TTC -> TTT	F628F
66	C3029A	2	-	-	-
	T7639C	7	L	GTT -> GTA	V969V
	G7870A	3	L	AAG -> AAA	K1046K
67	A6775G	1	L	GAA -> GAG	E681E
	T8769C	1	L	TTA -> TCA	L1346S
	C9246T	1	L	TCA -> TTA	S1505L
	A10032G	2	L	GAT -> GGT	D1767G
68	G361A	5	N	GAT -> AAT	D100N
	A6069C	3	L	GAA -> GCA	E446A
	C6071T	9	L	CTT -> TTT	L447F
69	G359A	5	N	GGG -> GAG	G99E
	T3894A	10	G	TCA -> ACA	S273T
	A6989G	1	L	ATA -> GTA	I753V
70	C1353T	1	N	-	-
	G3350T	10	G	CCG -> CCT	P91P

71	G359A	4	N	GGG -> GAG	G99E
	T3071C	2	G	-	-
	T4965G	3	L	TTA -> TGA	L78*
	C5398T	1	L	CTC -> CTT	L222L
	C7134A	6	L	ACC -> AAC	T801N
72	T5671A	8	L	CAT -> CAA	H313Q
73	A249G	2	N	GGA -> GGG	G62G
	A1317G	1	N	TCA -> TCG	S418S
	A2329G	10	M	TAT -> TGT	Y27C
	A6070G	7	L	GAA -> GAG	E446E
	C8547A	4	L	ACG -> AAG	T1272K
	G9430T	1	L	GTG -> GTT	V1566V
	A9681T	10	L	TAT -> TTT	Y1650F
74	G5323T	6	L	ACG -> ACT	T197T
	C7135A	3	L	ACC -> ACA	T801T
75	T1044A	1	N	CTT -> CTA	L327L
	G4249A	7	G	AGG -> AAG	R391K
	C5832T	1	L	ACT -> ATT	T367I
	T6728C	2	L	TCA -> CCA	S666P
	C8308T	4	L	CCC -> CCT	P1192P
	A8671G	3	L	TCA -> TCG	S1313S
76	G535A	4	N	GAT -> AAT	D158N
	A1224G	1	N	AGA -> AGG	R387R
	G2601A	7	M	GCC -> ACC	A118T
	C2896G	7	M	GCA -> GGA	A216G
	A4969C	2	L	ACA -> ACC	T79T
	T7406C	4	L	TTG -> CTG	L892L
	G8593T	4	L	TGG -> TGT	W1287C
77	G3350A	10	G	CCG -> CCA	P91P
	A3870G	2	G	AGA -> GGA	R265G
	A5239G	6	L	TTA -> TTG	L169L
	A8714G	1	L	ACA -> GCA	T1328A
	T10830C	4	L	ATC -> ACC	I2033T
78	C1146T	4	N	ACC -> ACT	T361T
	G1318A	2	N	GAA -> AAA	E419K
	T10532A	9	L	TAT -> AAT	Y1934N
	A1407C	9	P	CTA -> CTC	L4L
	G3699A	2	G	GAC -> AAC	D208N

79	G3871A	1	G	AGA -> AAA	R265K
	T4964G	1	L	TTA -> GTA	L78V
	G7591A	1	L	CTG -> CTA	L953L
	G8551T	1	L	TTG -> TTT	L1273F
80	T1438C	5	P	TCT -> CCT	S15P
	A6787G	1	L	CAA -> CAG	Q685Q
	G8548A	1	L	ACG -> ACA	T1272T
	T10407C	1	L	ATT -> ACT	I1892T
81	C7135A	3	L	ACC -> ACA	T801T
	G8548T	1	L	ACG -> ACA	T1272T
82	C5315A	2	L	CAT -> AAT	H195N
	A5973C	2	L	GAC -> GCC	D414A
	C8552T	3	L	CTC -> TTC	L1274F
	A10146G	3	L	GAA -> GGA	E1805G
83	G3414T	2	G	GAA -> TAA	E113*
	A5549G	1	L	AAA -> GAA	K273E
	G6935A	1	L	GGT -> AGT	G735S
	G7501A	1	L	CTG -> CTA	L923L
	C8308T	4	L	CCC -> CCT	P1192P
84	A123G	4	N	GCA -> GCC	A20A
	G4249A	7	G	AGG -> AAG	R391K
	C5302A	2	L	GTC -> GTA	V190V
	T7586G	2	L	TCT -> GCT	S952A
85	A242T	4	G	TTC -> TAC	F221Y
	T1438C	5	P	TCT -> CCT	S15P
	G5150A	1	L	GAA -> AAA	E140K
	A5627G	1	L	AAA -> GAA	K299E
	A7573G	2	L	GTA -> GTG	V947V
	C8547T	3	L	ACG -> ATG	T1272M
86	A4462G	1	G	AAG -> AGG	K462R
	T8602C	7	L	AGT -> AGC	S1290S
87	A123G	4	N	GCA -> GCC	A20A
	C1368T	9	N	-	-
	G1672A	6	P	GAT -> AAT	D93N
	A4983C	2	L	AAT -> ACT	N84T
	A5152G	1	L	GAA -> GAG	E140E
	A5240G	5	L	ATC -> GTC	I170V
	G5794A	2	L	TCG -> TCA	S354S

	T7736G	1	L	TTC -> GTC	F1002V
	A8546G	1	L	ACG -> GCG	T1272A
	C8837T	6	L	CTA -> TTA	L1369L
88	C2416T	1	M	CCG -> CTG	P56L
	T930C	2	N	TAT -> TAC	Y289Y
	G1713A	1	P	TCG -> TCA	S106S
	G3083T	1	G	AAG -> AAT	K2N
89	A7867G	5	L	AAA -> AAG	K1045K
	A8374G	2	L	AAA -> AAG	K1214K
	C8689T	2	L	CCC -> CCT	P1319P
	T10210C	2	L	TTT -> TTC	F1826F