



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Genetic Diversity and Evolution of Viral Populations

Rafael Sanjuán and Pilar Domingo-Calap, Institute for Integrative Systems Biology (I2SysBio), University of Valencia – CSIC, Valencia, Spain

© 2021 Elsevier Ltd. All rights reserved.

This is an update of R. Sanjuán, Quasispecies, in Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00481-7.

Glossary

Basic reproductive number Average number of new infections generated by an infected host.

Clonal interference Failure of beneficial alleles to get fixed in a population due to the presence of other, more beneficial alleles.

Defective interfering virus Mutant lacking a portion of the viral genome that replicates at the expense of and interferes with functional viruses.

Error catastrophe Process whereby, at high mutation rates, lower-fitness mutants become more abundant than higher-fitness variants.

Fitness Production of viable progeny, usually averaged over individuals of a given genotype and expressed relative to a reference genotype.

Fitness landscape A representation of the genotype-fitness map.

Genetic complementation Compensation of a genetic defect by the presence of other, functional copies of the mutated gene in the same cell.

Genetic hitch-hiking Increase in the population frequency of a deleterious mutation driven by genetic linkage with a beneficial mutation.

Lethal mutagenesis Demographical extinction of a population due to mutation accumulation.

Muller's ratchet Irreversible accumulation of deleterious mutations in a population as a result of random genetic drift.

Mutation rate Probability that a new genetic variant is produced per generation, where a viral generation is often defined as the completion of a cellular infection cycle.

Mutation-selection balance Equilibrium between the influx of new deleterious mutations in a population and the removal of these mutations through selection.

Selection coefficient Difference between the fitness value of a genetic variant and the fitness value of a reference genotype.

Viral quasispecies A population of viruses replicating and competing at high mutation rates.

Introduction

Viruses exhibit highly heterogeneous genome structures and replication strategies. These basic features correlate with viral propensity to mutate and, hence, with viral genetic diversity and evolvability. The mutation rates of viruses vary according to genome composition (RNA or DNA), size, and structure (single- or double-stranded). In general, RNA viruses produce from 10^{-6} to 10^{-4} new base substitutions per nucleotide per cell, whereas these rates range from 10^{-8} to 10^{-6} in DNA viruses (Fig. 1). Furthermore, single-stranded viruses appear to mutate faster than double-stranded viruses, and an inverse correlation between genome size and mutation rate has also been demonstrated. Within the DNA virus group, thus, single-stranded viruses and, presumably, some double-stranded viruses with small genomes, mutate faster than double-stranded viruses with large genomes.

The term quasispecies has been used to refer to highly heterogeneous viral populations in which the frequency of mutations in the population is determined by mutation rate and fitness. The quasispecies concept has thus been applied mainly to RNA viruses, owing to their particularly high mutation rates. High population genetic diversity is believed to play a pivotal role in the ability of RNA viruses to adapt to changing selective pressures and invade new hosts. Indeed, RNA viruses comprise a large number of well-known human emerging or re-emerging pathogens such as HIV-1, hepatitis C virus, influenza virus, Ebola virus, and Zika virus among many others, as well as major plant and animal pathogens. In contrast, DNA viruses are typically less diverse and evolve more slowly, although they also include important pathogens such as herpesviruses and poxviruses. Nevertheless, the divide between RNA and DNA viruses is imperfect, since some RNA viruses evolve slowly and some DNA viruses are fast-evolving, particularly single-stranded and small double-stranded DNA viruses.

This article focuses on the sources of genetic diversity in viruses and on the resulting evolutionary processes at the population level (we do not review long-term viral evolution at the phylogenetic level). The main mechanisms responsible for introducing new mutations are discussed, including replication errors, lack of proofreading, repair avoidance, diversity-generating elements, host-encoded editing of viral genomes, and recombination (Table 1). We review how population diversity is modified by natural selection and genetic drift, and how this leads to different evolutionary processes in viruses. Finally, we discuss some practical implications of viral diversity for pathogenesis, immune escape, and drug resistance.

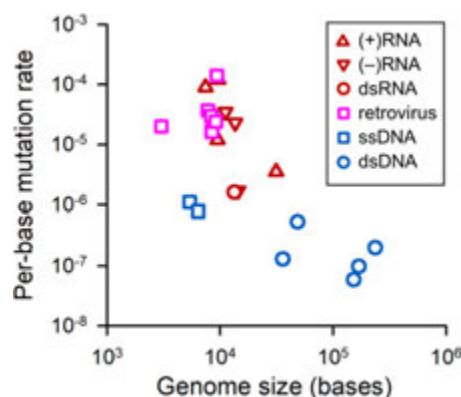


Fig. 1 Mutation rates in different types of viruses. Estimates obtained for viruses belonging to the major indicated groups are shown as a function of genome size. This reveals an effect of genetic material (RNA/DNA) and genome size on viral mutation rates. Data taken from Sanjuán, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. *Cellular and Molecular Life Sciences* 73, 4433–4448. doi:10.1007/s00018-016-2299-6.

Table 1 Sources of genetic diversity in seven major virus classes^{a,b}

	<i>dsDNA</i>	<i>ssDNA</i>	<i>Para retro</i>	<i>Retro</i>	<i>dsRNA</i>	<i>(+)RNA</i>	<i>(-)RNA</i>
Lack of 3'exonuclease proofreading	–	+/-	+	+	+	+	+
Avoidance of post-replicative repair	–	+/-	+	+	+	+	+
Use of error-prone repair polymerases	+/-	+/-	–	–	–	–	–
Diversity-generating retro-elements	+/-	–	–	–	–	–	–
APOBEC hyper-mutation	+/-	+/-	+	+	–	–	–
ADAR hyper-mutation	–	–	–	–	+/-	+/-	+/-
Reassortment	+/-	+/-	–	–	+/-	+/-	+/-
Template switching	–	–	+	+	+	+	–
Recombinases	+/-	–	–	–	–	–	–

^aAccording to the Baltimore classification of viruses.

^bSource of diversity generally present (+), present in some cases (+/-), or not shown or infrequent (–).

Error-Prone Replication

Intrinsic Selectivity of Viral Polymerases

The fidelity of a polymerase is first determined by its selectivity towards the correct base. Such selectivity is typically on the order of 10^4 – 10^5 , and there appear to be no major differences among RNA-dependent RNA polymerases, reverse transcriptases (RTs), and DNA polymerases. Polymerase variants with altered selectivity and fidelity have been reported in a wide variety of RNA viruses. Serial transfers in the presence of the base analog ribavirin were first shown to select for an increased-fidelity polymerase in polioviruses. Ever since, tens of fidelity mutants have been identified amongst widely different RNA virus families. The evolution of high- and low-fidelity variants in response to selective pressures such as chemically-induced mutagenesis demonstrates that selection can optimize RNA virus replication fidelity. However, most RNA virus fidelity variants show only modest differences in base selectivity compared to the wild type, revealing strong structural or biochemical constraints. Interestingly, the fidelity of viral DNA polymerases appears to be less constrained.

Other Factors Determining Polymerase Selectivity

In addition to the intrinsic fidelity of a polymerase, the rate of base misincorporations depends on factors such as sequence context, divalent cation concentrations, and the relative abundance of dNTPs. For instance, polymerase slippage occurs frequently in homopolymeric templates, producing insertions and deletions. In hepatitis C virus, mutation rates can vary extensively between consecutive bases in the viral genome, suggesting a major role for sequence context in determining replication fidelity. RNA secondary structure can also determine the rate at which replication errors occur.

Lack of Proofreading

Polymerases displaying 3' exonuclease activity can correct most nucleotide misincorporations. However, this activity is generally absent from RNA virus polymerases. Currently, only coronavirus RNA polymerases have been shown to perform

3' exonuclease proofreading. Lack of proofreading is a major factor responsible for the higher rates of spontaneous mutation of RNA viruses compared to DNA viruses. Because RTs also lack 3' exonuclease activity, retroviruses and pararetroviruses mutate at rates similar to those of most RNA viruses. Although DNA virus replicases exhibit 3' exonuclease activity, certain amino acid replacements in their polymerases inactivate proofreading and produce a strong mutator phenotype, as shown for instance in bacteriophage T4. Interestingly, there are also amino acid replacements in the T4 polymerase that increase fidelity up to a hundred-fold compared to the wild type. Finally, polyomaviruses, which use host polymerases for replication, can encode proteins that inactivate the 3' exonuclease proofreading domain of the host polymerase. Therefore, avoidance of proofreading may also be a mechanism whereby some DNA viruses increase their population diversity.

Role of Post-Replicative Repair

Repair Avoidance

Post-replicative repair is a highly efficient system for removing replication errors and repairing DNA damage. As such, lack of repair can increase mutation rates strongly. RNA viruses are not substrates of cellular repair systems, whereas some DNA viruses have evolved specific features to avoid repair. This is the case of bacteriophage Φ X174 and potentially other small DNA phages. In *Escherichia coli*, strand-specific mismatch repair uses the methylation status of GATC sequence motifs to discern between the parental and daughter DNA strands. Interestingly, the Φ X174 genome is devoid of GATC motifs. This prevents methylation and thus, post-replicative repair, allowing the phage to increase its mutation rate relative to that of the host. Eukaryotic viruses might also avoid repair or dysregulate repair pathways to increase their mutation rates, although this has not been investigated in detail.

Error-Prone Repair Polymerases

Eukaryotic viruses are known to interact with DNA damage response pathways. For instance, some small DNA viruses activate these pathways to prolong the S cell-cycle phase and promote their replication. DNA damage response pathways comprise error-prone DNA polymerases for re-synthesis of excised strands, which might contribute to elevating viral mutation rates. On the other hand, some large DNA viruses encode their own DNA repair system. For instance, African swine fever virus encodes an endonuclease, a repair polymerase (pol X) and an ATP-dependent DNA ligase. Pol X is highly error-prone due to a poor base selectivity and lack of 3' exonuclease activity. Use of pol X might contribute to the relatively high genetic diversity shown by African swine fever virus isolates.

Diversity-Generating Retroelements

First described in a *Bordetella* phage, diversity-generating retroelements are genetic cassettes found in different bacteria, plasmids, and DNA bacteriophages. They contain two short repeated sequences, called the template and the variable repeats. The template repeat is transcribed and then reverse-transcribed by an RT encoded by the cassette. During this process, adenines are systematically substituted for random bases. The resulting cDNA is integrated, replacing the former variable repeat with a new, highly mutagenized sequence. This process allows hyper-mutation of specific phage genes involved in host attachment, a trait that is often subject to rapidly changing selective pressures.

Viral Hyper-Mutation Mediated by Host Enzymes

Cellular Cytidine Deaminases

Apolipoprotein B mRNA editing catalytic polypeptide-like enzymes (APOBECs) function as an innate immune response against reverse-transcribing viruses. APOBECs are cytidine deaminases that massively introduce C-to-U base substitutions in the retroviral cDNA, leading to a characteristic G-to-A hyper-mutation pattern in retroviral genomes. It has been estimated that 98% of new HIV-1 mutations are produced by APOBECs, versus only 2% caused by HIV-1 RT errors. Hyper-mutation produces non-infectious genomes, which accumulate as inactive proviruses. In cases of moderate mutation, though, APOBECs might contribute to viral evolution and could even promote immune escape and drug resistance. In addition to retroviruses, APOBECs can edit hepatitis B virus as well as other non-reverse transcribing DNA viruses such as papillomaviruses, herpesviruses, and human polyomaviruses.

Cellular Adenosine Deaminases

Double-stranded RNA-dependent adenosine deaminases (ADARs) constitute another type of host-encoded enzymes capable of editing and hyper-mutating RNA virus genomes. ADARs deaminate adenosines in long double-stranded RNA regions, converting them into inosines, which leads to A-to-G base substitutions. ADAR-induced hyper-mutation was first discovered in measles virus and was then extended to other viruses including human parainfluenza virus, respiratory syncytial virus, lymphocytic choriomeningitis virus, Rift Valley fever virus, influenza virus, noroviruses, and Zika virus, although in many cases evidence for ADAR-mediated mutagenesis is still indirect.

Recombination

Recombination plays an important role in viral evolution. Similar to mutation rates, recombination rates vary extensively among viruses. High rates of recombination may have been favored by natural selection to promote the fixation of new genotypes, to purge deleterious mutations, or as a by-product of other evolutionary processes. For recombination to introduce new genetic variants, it is required that the involved genomes differ by at least two mutations. Furthermore, the recombining molecules have to be present in the same cell. Hence, the emergence of new variants through recombination typically occurs in cells infected with different pre-existing variants. However, this may not happen frequently because super-infection exclusion renders infected cells refractory to re-infection with the same or a related virus. On the other hand, even if rare, recombination between distantly related viruses has the ability to introduce major genetic changes in a single event, such as gene transfer, leading to the emergence of new viral subtypes or even new species. Hence, recombination differs from mutation in the type and abundance of the genetic changes produced.

Reassortment

In segmented and multicomponent viruses, recombination is greatly facilitated by the possibility of re-assorting genome segments. However, barriers to reassortment also exist. For instance, co-packaging of a given genome segment with the segments of another virus requires that their packaging signals are compatible. Reassortment has been extensively studied in influenza virus, for which inter-subtype reassortments between swine, avian and/or human influenza A viruses are known to be at the origin of several pandemics. Reassortment between viruses of the same subtype can also be an important source of diversity, can promote adaptation, and can accelerate immune escape. Reassortment also plays a major role in the evolution of other viruses such as rotaviruses, favoring the emergence of new variants and complicating vaccination efforts.

Template Switching

Even if reassortment is not possible, recombination can still take place through a template-switching mechanism whereby the viral polymerase and the nascent chain dissociate from one template and associate with another. The rate at which this template-switching process occurs has been measured experimentally using genetic markers, and typically ranges from 10^{-5} to 10^{-3} events per nucleotide site per co-infected cell in different retroviruses such as HIV-1, murine leukemia virus, and spleen necrosis virus, as well as in positive-stranded RNA viruses such as mouse hepatitis virus, poliovirus, and tobacco etch virus. Whereas the production of new variants through template switching also requires that the cell is coinfecting with two different variants, in retroviruses such as HIV coinfection with two independent virions is not required because virions are diploid. In contrast to retroviruses and positive-stranded RNA viruses, most negative-stranded RNA viruses show very low rates of template switching. Similar to replication fidelity variants, viral polymerases with different propensities to template switching have been isolated. Such variants have allowed investigating the evolutionary implications of recombination from an experimental evolution approach using poliovirus. This has suggested that recombination increases the ability to purge deleterious mutations from the population and promotes the spread of beneficial mutations. This may accelerate adaptation and increase viral pathogenicity.

Recombination in DNA Viruses

Recombination is a more complex process in DNA viruses than in RNA viruses and involves specific host or viral recombinases. DNA virus homology-dependent recombination can be mediated by cellular double-stranded break repair systems. These systems act on double-stranded DNA viruses but also on single-stranded DNA viruses because the latter produce double-stranded replicative intermediates. Some DNA viruses encode their own recombination systems. For instance, phage λ recombination is active in cells deficient for the *E. coli* recombinase RecA. Also, the herpesviruses recombination machinery plays a major role during viral replication and repair. There is ample evidence that this type of recombination is sequence-dependent. For instance, the inverted terminal repeats of the vaccinia virus genome undergo rapid changes in size because this region contains multiple 10–100 bp repeats that are highly prone to unequal crossovers. Some viruses can use recombinogenic sequence motifs for targeting diversity towards certain genes. In poxviruses, this allows for a recombina-

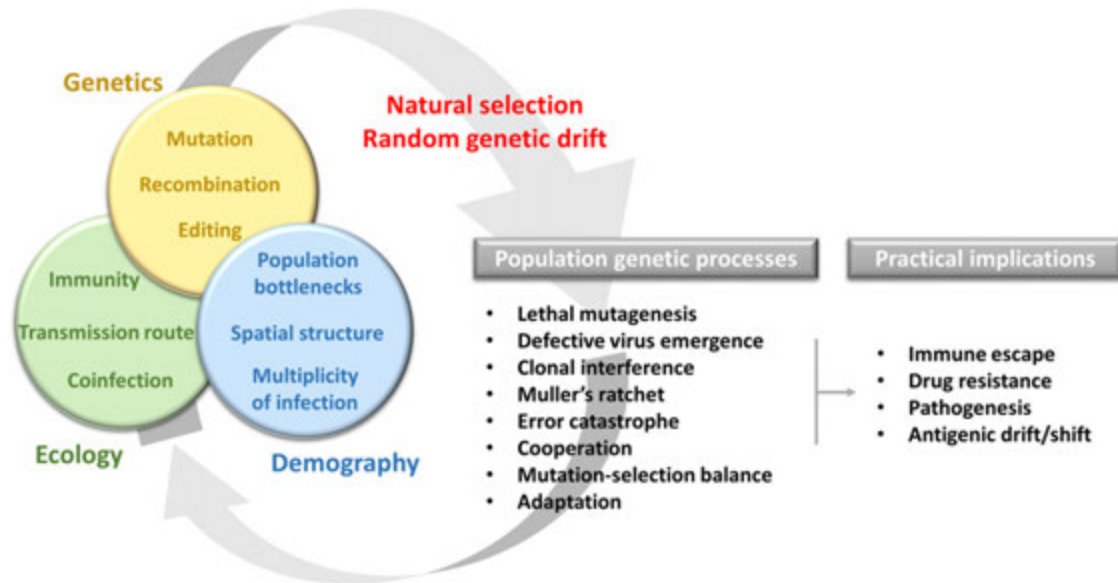


Fig. 2 Genetic, demographical, and ecological processes determining viral population diversity. Mutation is the ultimate source of genetic variation, but recombination also produces new genotypes. Ecological and demographical processes determine how natural selection and random genetic drift acts on these new genetic variants. This leads to a variety of possible population genetic processes, which in turn feedback on viral demography, ecology and genetics. Immune escape, drug resistance, and pathogenesis are a visible consequence of such processes. The list of factors shown is not exhaustive.

tion-mediated gene amplification process that can temporarily increase the dosage of viral genes involved in evading host innate immunity.

Basic Population Genetic Processes Shaping Viral Diversity

The fate of new genetic variants is largely determined by selection. The population frequency of a new mutant will tend to increase, fluctuate randomly, or decrease depending on whether it has a beneficial, neutral, or deleterious effect on viral fitness, respectively. Yet, the fitness of a given allele is context-dependent in several ways, since it varies according to genotype-genotype and genotype-environment interactions. To a large extent, the environment of a virus is dictated by the host, but it can also be modulated by other members of the viral population via competition for cellular resources or cooperative interactions. Allele frequencies can also change as a result of random genetic drift. Drift is determined by the effective size of the viral population, which in turn depends on spatial structure, population size bottlenecks during host-to-host transmission, viral growth dynamics, and so on. Therefore, both selection and drift are conditioned by multiple demographical and ecological factors, which results in a variety of population genetic processes (Fig. 2). Below, we briefly summarize some of these processes.

Mutation-Selection Balance

Since natural populations tend to be adapted to their environments, most new mutations with a fitness effect are deleterious and, hence, selection tends to remove them from the population. When the influx of genetic diversity produced by mutation equals the outflow due to selection, the population reaches an equilibrium called the mutation-selection balance. In haploid populations, the population frequency of a given deleterious allele at the mutation-selection balance simply equals μ/s , where μ is the mutation rate and s the selection coefficient. The selection coefficient ranges from zero for neutral alleles to one for lethal alleles. The time for this equilibrium to be reached depends inversely on s , the equilibrium being reached instantaneously (in one generation) for $s = 1$. In RNA viruses, the statistical distribution of selection coefficients for new mutations has been characterized using site-directed mutagenesis. This has revealed that RNA viruses show very high s -values (on average, $s > 0.1$) compared to other microorganisms. Hence, selection against deleterious mutations acts strongly on RNA viruses. This implies that the mutation-selection balance tends to be reached in few generations and that, at equilibrium, most genetic variants have low population frequencies. Supporting this view, high-fidelity Illumina sequencing of experimental poliovirus populations has revealed a large number of deleterious variants at frequencies that are not detectable by conventional next-generation sequencing. A likely explanation for why s -values are high in RNA viruses is that their genomes are extremely compact, with few and often multifunctional and overlapping genes, many of which

are essential. These genome features also apply to small DNA viruses and, consequently, mean s -values are also very high for DNA viruses such as microviruses (e.g., phage Φ X174) and inoviruses (e.g., phage M13).

Viral Quasispecies

Quasispecies theory is primarily a theory of mutation-selection balance at extremely high mutation rates. Quasispecies were originally studied within the framework of physical chemistry as models for the dynamics of self-replicating sequences in a hypothetical primitive RNA world. However, due to the high mutation rates exhibited by RNA viruses, the theory subsequently became popular amongst virologists. Despite using different terminology, quasispecies and classical mutation-selection models essentially deal with the same process and are largely equivalent. Yet, quasispecies models make a more explicit approach to mutant dynamics by explicitly considering the replicative rate of each sequence as well as the probability of each sequence mutating to another specific sequence. A transition matrix is thus defined, whose eigenvector with the highest eigenvalue defines a stable mutation-selection balance. This approach defines fitness as a property of the ensemble of sequences present at equilibrium (the quasispecies). As a consequence, fitness is not merely determined by the initial sequence, but also by its mutational neighborhood. This means that a quasispecies containing a maximally fit individual sequence might be outcompeted by another quasispecies that does not contain the fittest sequence but shows higher average fitness. This phenomenon is called the survival of the flattest. A related prediction of quasispecies models is that, beyond a certain mutation rate, the fittest sequence within a given quasispecies can be outnumbered by its own cloud of deleterious mutants, a process known as error catastrophe. How this process occurs depends on the topology of the fitness landscape. Original quasispecies models assumed a landscape with a single high-fitness sequence (master) in which all mutants had the same fitness value. In this particular system, beyond the error threshold all mutants reach a frequency similar to that of the master and, since there is an extremely large number of such possible mutants, the actual frequency of the master sequence becomes vanishingly small.

Error Catastrophe and Lethal Mutagenesis of Viruses

The theory of error catastrophe suggested that viral populations could be effectively extinguished by increasing the viral mutation rate sufficiently. Treatment of RNA virus experimental populations with chemical mutagens has provided ample support for the feasibility of mutation-driven extinction in cell cultures and even in animal models. However, error catastrophe models only deal with the fitness of certain sequences relative to others, and thus ignore the fact that extinction is an absolute, demographical process that takes place when the basic reproductive number of a population decays below one. It has been thus suggested that extinction of experimental RNA populations following mutagen treatments is better defined as lethal mutagenesis process. Regardless of the models used and the mechanisms underlying extinction, empirical evidence strongly support the view that RNA viruses naturally exist at the verge of mutation-driven extinction, since extinction can be induced by modest increases in mutation frequencies. In contrast, the lower rates of spontaneous mutation shown by most DNA viruses makes them less likely to undergo lethal mutagenesis or error catastrophe.

Adaptation via Directional Selection of Beneficial Mutations

The high mutation rates exhibited by RNA viruses reduce the waiting times required for the emergence of selectively beneficial mutations compared to other microorganisms. The emergence of beneficial mutations is also favored by the extremely high population sizes of RNA viruses, which is on the order of billions of infectious particles per infected host. For a mutation rate of 10^{-5} substitutions per nucleotide per round of genome copying, replication of a population of 10^9 viral particles should produce as many as 10^4 mutations, which approaches the total number of possible single-point mutations in a typical 10 kb RNA virus genome. This makes RNA viruses extremely responsive to host adaptive immunity, antiviral drugs, and other rapidly-changing selective pressures. Yet, despite favoring the appearance of beneficial alleles, high mutation rates can hamper their fixation process in the population. First, at high mutation rates, many genomes carry several deleterious mutations, and beneficial mutations may fail to be selected if they occur in low-fitness genetic backgrounds. Alternatively, if the beneficial effect of a mutation exceeds the disadvantageous effect of other mutations present in the same genome, deleterious mutations can undergo fixation, a process known as genetic hitch-hiking. Another “traffic issue” in fast-mutating systems is clonal interference, a process whereby multiple beneficial mutations coexist in the same population and only the fittest variant is selected, leading to extinction of other beneficial variants. Since lower mutation rates increase waiting times for the emergence of beneficial mutations but higher mutation rates reduce the efficiency with which beneficial mutations are selected, adaptation should be maximal for some intermediate, optimal mutation rate. A similar optimality argument can be used for neutral evolution. Analysis of empirical data suggests that RNA viruses replicate at nearly-optimum mutation rates. Lastly, clonal interference, counter-selection of beneficial alleles in low-fitness genetic backgrounds and hitch-hiking can all be dissipated by recombination, since recombination can break disadvantageous genetic combinations and create new ones.

Frequency-Dependent Selection

The fitness of a given virus can be modified by other viruses of the same population. Such interactions can be cooperative or competitive, and are more likely to occur among viruses infecting the same cell because this allows them to share proteins. Coinfection of a cell with independent virions requires a high local density of viral particles. However, many viruses spread as collective infectious units composed of multiple particles, which promotes coinfection even at low overall viral densities. When a situation in which each cell receives multiple viral particles is maintained for several infection cycles, defective viruses tend to invade the population. Defective viruses, produced by illegitimate recombination, lack large portions of the genome and can only replicate as hyper-parasites at the expense of complete viral genomes present in the same cell. In addition to replicating faster than the wild type because their genomes are shorter, defective viruses often interfere with some steps of the infectious cycle. Hence, the fitness values of both the wild type and the defective viruses are frequency-dependent. Viral interference can produce sharp drops in fitness, which has led to suggest defective interfering viruses as possible antivirals. The emergence of interfering viruses can also be promoted by chemically induced mutagenesis in RNA viruses. On the other hand, coinfection of cells with different virus variants may allow for cooperative virus-virus interactions. In principle, two viruses each carrying a different genetic defect can mutually increase their fitness values if they share gene products (genetic complementation). It is also possible that each variant has optimized a different aspect of the infection cycle, and that such variants achieve higher fitness in coinfection than alone. This type of cooperative interaction has been suggested for measles virus and influenza virus, and might contribute to the maintenance of viral genetic diversity because fitness would be maximal at intermediate frequencies of each variant in the population. Therefore, diversity can lead to negative or positive virus-virus interactions. In turn, these interactions feedback on diversity by modifying the population frequency of certain variants.

Random Genetic Drift

Despite their enormous population sizes, viruses experience significant genetic drift. This is because the strength of drift depends on the effective population size, not on the census size. Demographical bottlenecks have a strong and long-lasting effect on effective population sizes, and such bottlenecks occur extensively during inter-host transmission in nearly all viruses. The size of the founder population following inter-host transmission has been estimated using genetic markers, and is typically lower than 10 genome copies. This number, though, can vary widely depending on transmission route (aerosol, fecal-oral, sexual, insect-borne, etc.). Additionally, viral infections show extensive intra-host spatial structure caused by restrictions to viral trafficking among tissues or organs (organ compartmentalization) as well as by the limited diffusion of virions in the extracellular milieu (localized infection foci). Spatial structure also reduces effective population sizes, hence promoting drift. The immediate consequence of drift is the loss of genetic diversity, because only a subset of the population contributes to the next generation. A longer-term consequence is the random fixation of alleles in the population. In highly variable populations such as those of RNA viruses and at sufficiently low population sizes, drift can result in the loss of the wild-type sequence. In the absence of back mutation or recombination, this loss is irreversible and hence leads to the accumulation of random mutations in the population, most of which are neutral or deleterious. This process, known as Muller's ratchet, reduces mean population fitness and can endanger the survival of small RNA virus populations.

Implications for Viral Disease

Viral population genetic diversity plays a major role in the ability of viruses to cause disease. In general, pathogens evolve faster than their hosts owing to their shorter generation times and higher population sizes. In addition, in their evolutionary arms race against hosts, most viruses benefit from a faster mutation rate, particularly RNA viruses. On the other hand, hosts have the ability to deploy complex antiviral responses such as adaptive immunity. As a result, viruses are subject to continual selective pressure for avoiding antiviral defenses.

Short-Term Pathogenesis

Different early processes including efficient entry, local replication, and rapid spread determine pathogenesis. The association between RNA virus population diversity and pathogenesis has been examined using polymerase fidelity variants. Increased-fidelity variants are typically attenuated in animal models, suggesting a role for viral population diversity in causing disease. This was first shown using a variant of poliovirus carrying a single amino acid replacement in the polymerase that confers an approximately three-fold increase in replication fidelity. Since these experiments involved acute infections, adaptive immunity was probably not responsible for the observed association between population diversity and pathogenesis, and the mechanism linking intra-host viral diversity with pathogenesis has not been elucidated. Importantly, low-fidelity variants have also been tested and found to be attenuated. A candidate alternative explanation for these results is that in RNA viruses, high mutation rates have been selected as a byproduct of selection for high replicative speed. Increasing fidelity would come at the cost of slower growth, hence producing an attenuated phenotype. On the other hand, decreasing fidelity further would elevate the mutational load of the population, reducing mean population fitness at mutation-selection balance or promoting error catastrophe.

Immune Escape

In chronic infections such as those caused by HIV or hepatitis B and C, among others, adaptive immune responses involving neutralizing antibodies and cytotoxic T cells select for escape mutants during the course of infection. In HIV-1, the ability of the virus to escape immunity partially correlates with its ability to generate a genetically diverse intra-host population. DNA viruses such as, for instance, herpes simplex virus also cause chronic infections and, hence, have to cope with the adaptive immune response as well. Yet, their strategy for evading immune responses appears to be based on latency rather than on the continual production of immune escape variants. In viruses causing acute infections, adaptive immunity escape mutants are also selected, but this process occurs mainly at the inter-host population level. For this selection to occur, hosts need to recover from infection. This creates a host subpopulation immunized against the virus variant that originally infected them, thus favoring escape variants. This process has been amply documented for influenza virus, and is called antigenic drift or antigenic shift, depending on the diversity-generating process at play. Antigenic drift is mainly based on point mutations that change epitope sequences, whereas antigenic shift occurs as a result of segment reassortments that produce new subtype combinations. Finally, immune escape is not restricted to vertebrate viruses. For instance, bacteria use clustered regularly interspaced short palindromic repeats (CRISPR) and associated proteins as an adaptive immune system. Bacteriophages can escape CRISPR via point mutations and have also evolved anti-CRISPR proteins (Acrs) of variable strength.

Drug Resistance

Antiviral drugs impose a strong selective pressure on viruses causing chronic diseases such as HIV-1. The first attempts to treat AIDS using base analogs failed due to the emergence of drug-resistant mutants. Monotherapies were subsequently replaced with combination therapies aimed at preventing drug resistance, which have been successful despite the remarkable genetic variability of HIV-1 intra-host populations. A similar scenario applies to hepatitis C virus, in which resistance against protease inhibitor-based monotherapies emerges rapidly through one or few mutations, thus requiring the use of combination therapies. Targeting different steps of the viral infection cycle, such as for instance nucleoside analog-based polymerase inhibitors and protease inhibitors reduces the risk of resistance compared to combination therapies aiming at a single target. An alternative antiviral treatment consists in taking advantage of the highly error-prone replication of RNA viruses to promote viral extinction via lethal mutagenesis using nucleoside analogs or other mutagenic compounds. This type of treatment can select for high-fidelity polymerases that partially compensate the mutagenic effect of the drugs, although this might not be sufficient for preventing viral extinction. In other cases, drug-resistance against mutagenesis may fail to evolve. Finally, yet another alternative treatment strategy that has been designed to minimize the evolution of drug resistance consists in targeting host factors such as molecular chaperones, the viral receptor, or proteins involved in the immune response.

Conclusions

Viruses are a highly heterogeneous group of molecular parasites showing varied replication strategies, genome organizations, mutability, and evolutionary properties. RNA viruses are characterized by their elevated mutation rates and extremely high genetic diversity, but reverse-transcribing DNA viruses, single-stranded DNA viruses, and some small double-stranded DNA viruses can also achieve high levels of population genetic diversity. Here, we have shown how viral diversity is generated by widely different molecular mechanisms, including lack of proofreading, avoidance of DNA repair, host-mediated viral genome editing, and specific diversity-generating modules, as well as different mechanisms of recombination. However, viral diversity does not only depend on these purely genetic factors. Ecological and demographic processes determine the action of natural selection and random genetic drift, producing a myriad of different population genetic processes, of which pathogenesis, immune escape, and drug resistance are observable consequences. Therefore, viral population diversity and evolution should help us develop more efficient antiviral control and treatment strategies.

Further Reading

- Andino, R., Domingo, E., 2015. Viral quasispecies. *Virology* 479–480, 46–51. doi:10.1016/j.virol.2015.03.022.
- Cvijović, I., Nguyen Ba, A.N., Desai, M.M., 2018. Experimental studies of evolutionary dynamics in microbes. *Trends in Genetics* 34, 693–703. doi:10.1016/j.tig.2018.06.004.
- Dolan, P.T., Whitfield, Z.J., Andino, R., 2018. Mechanisms and concepts in RNA virus population dynamics and evolution. *Annual Review of Virology* 5, 69–92. doi:10.1146/annurev-virology-101416-041718.
- George, C.X., Gan, Z., Liu, Y., Samuel, C.E., 2011. Adenosine deaminases acting on RNA, RNA editing, and interferon action. *Journal of Interferon Cytokine Research* 31, 99–117. doi:10.1089/jir.2010.0097.
- Gutiérrez, S., Michalakis, Y., Blanc, S., 2012. Virus population bottlenecks during within-host progression and host-to-host transmission. *Current Opinion in Virology* 2, 546–555. doi:10.1016/j.coviro.2012.08.001.
- Harris, R.S., Dudley, J.P., 2015. APOBECs and virus restriction. *Virology* 479–480, 131–145. doi:10.1016/j.virol.2015.03.012.
- Irwin, K.K., Renzette, N., Kowalik, T.F., Jensen, J.D., 2016. Antiviral drug resistance as an adaptive process. *Virus Evolution* 2, vew01. doi:10.1093/ve/vew014.
- Kautz, T.F., Forrester, N.L., 2018. RNA virus fidelity mutants: A useful tool for evolutionary biology or a complex challenge? *Viruses* 10, E-600. doi:10.3390/v10110600.
- McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassortment in segmented RNA viruses: Mechanisms and outcomes. *Nature Reviews Microbiology* 14, 448–460. doi:10.1038/nrmicro.2016.46.

- Perales, C., Domingo, E., 2016. Antiviral strategies based on lethal mutagenesis and error threshold. *Current Topics in Microbiology and Immunology* 392, 323–339. doi:10.1007/82_2015_459.
- Retel, C., Märkle, H., Becks, L., Feulner, P.G.D., 2019. Ecological and evolutionary processes shaping viral genetic diversity. *Viruses* 11, 220. doi:10.3390/v11030220.
- Sanjuán, R., 2017. Collective infectious units in viruses. *Trends in Microbiology* 25, 402–412. doi:10.1016/j.tim.2017.02.003.
- Sanjuán, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. *Cellular and Molecular Life Sciences* 73, 4433–4448. doi:10.1007/s00018-016-2299-6.
- Simon-Lorière, E., Holmes, E.C., 2011. Why do RNA viruses recombine? *Nature Reviews Microbiology* 9, 617–626. doi:10.1038/nrmicro2614.
- Weller, S.K., Sawitzke, J.A., 2014. Recombination promoted by DNA viruses: Phage λ to herpes simplex virus. *Annual Review of Microbiology* 68, 237–258. doi:10.1146/annurev-micro-091313-103424.