



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.

Chapter 2

Genetic Instability of RNA Viruses

J.N. Barr¹, R. Fearn²

¹University of Leeds, Leeds, United Kingdom; ²Boston University School of Medicine, Boston, MA, United States

Chapter Outline

1. Introduction	21	10. Mechanisms Underlying Genetic Robustness in RNA Viruses	28
2. Overview of RNA Virus Multiplication	21	11. RNA Viruses on the Edge	29
3. Viruses as Quasispecies	23	12. Virus Genetic Variability and the Virus–Host “Arms Race”	30
4. Overview of RNA Virus Replication Mechanisms	23	13. Taking Advantage of the Mutability of RNA Viruses	31
5. The Viral Polymerase as a Source of Error	23	14. Conclusion	31
6. Other Viral Determinants of Mutation Rate	25	Glossary	31
7. Recombination	25	List of Abbreviations	32
8. The Effect of Replication Mode on Mutation Frequency	27	References	33
9. The Effect of Cellular Factors on Virus Mutation Rate	27		

1. INTRODUCTION

Viruses are enormously successful. They have been identified in organisms within all domains of life. Despite decades of scientific effort to combat viruses that cause disease in humans and economically important crops and animals, there are relatively few cases in which we have succeeded. Viruses have shown they are able to adapt and multiply to overcome almost any obstacle that is imposed on them. This remarkable adaptability can be attributed to their extremely high replication rate and their propensity for mutation. This is particularly true of the viruses that have RNA genomes: the riboviruses and retroviruses. This chapter will focus on these RNA viruses and on the exciting research that has provided valuable insight into how RNA viruses benefit from their genetic variability. In the first two sections of the chapter, two fundamental concepts are introduced: the intimate relationship between RNA viruses and their hosts, and the idea that viruses behave as quasispecies. Having introduced these concepts, the remainder of the chapter discusses the viral and host mechanisms that govern RNA virus genetic variability and the ability of viruses to withstand mutation. We then discuss evidence that at least some RNA viruses have a replication fidelity that is poised to maximize genome sequence space without incurring catastrophic lethal mutations and describe how this can be exploited to control viral infections. Throughout the chapter, we attempt to convey the diversity of RNA virus biology and mutation frequency and we conclude by speculating that each RNA virus has evolved an error rate that complements its genome replication strategy and mode of transmission.

2. OVERVIEW OF RNA VIRUS MULTIPLICATION

RNA viruses are very simple entities with small genomes that vary in length from about 2 to 32kb, depending on the virus. Thus, they have very limited coding capacity and so, similarly to DNA viruses, they are obligate intracellular parasites, depending on a host cell to provide energy generating systems, ribo- and deoxyribonucleotides, cellular translation machinery, tRNAs and amino acids to translate their mRNAs, cellular enzymes to posttranslationally modify their proteins, and

cellular structures such as membranes, vesicular compartments, and/or cytoskeleton networks to act as a scaffold for assembling and transporting components required to make virus particles. There are many RNA viruses and they vary enormously in their genome structures and mechanisms of replication. However, in its most distilled and generic form, the RNA virus infection cycle consists of the steps shown in Fig. 2.1.

First a protein on the surface of the virus particle attaches to a receptor molecule on a host cell enabling the viral genome to be delivered into the cell. The genome is expressed to produce viral proteins and replicated multiple times to produce progeny genomes. The progeny genomes are packaged with the proteins that make up the virus particle and are released to infect new cells. Thus, viruses multiply by a process of genome replication, expression, and assembly, rather than division, and a cell infected by a single infectious virus particle could release thousands of progeny virions in a matter of hours. This enables viruses to multiply very rapidly and to achieve large population sizes. Because viruses depend on a host cell to be able to replicate, their ability to multiply is heavily influenced by the biology of each cell that they encounter, such as the nature and density of surface molecules that can act as viral receptors, the cell's metabolic rate and availability of macromolecules, as well as the cell's innate antiviral defenses that have evolved to suppress viral replication. In addition to being able to replicate within a single cell type, most viruses require the capacity to replicate and spread within a multicellular host organism, which has tissues with varied cellular characteristics, physiological and anatomical constraints, and an adaptive immune response. While some viruses might only require the ability to infect one tissue to be successfully maintained in the environment, some viruses need to infect and multiply in different tissues to be spread to a new host and complete their transmission cycles. For example, measles virus initially infects alveolar macrophages and dendritic cells in the lung. It is then transferred to T- and B-lymphocytes and is amplified and spread systemically throughout the body. Infected lymphocytes can then transfer the virus to the basolateral surface of lung epithelial cells by attaching to an epithelial cell receptor. The virus multiplies further in the lung epithelium and is spread to new hosts by coughing and sneezing [1,2]. Thus, measles virus requires the ability to infect multiple cell types to complete its transmission cycle. Viruses must also be capable of replicating within populations of hosts whose immune responses are shaped by different histories of virus exposure and some viruses even require the ability to replicate in different host species. For example, West Nile virus transmission is dependent on the virus being able to replicate efficiently in both mosquitos and birds. In mosquitos the virus multiplies in the salivary glands and is transmitted to birds when the mosquito takes a blood meal. The virus is amplified in

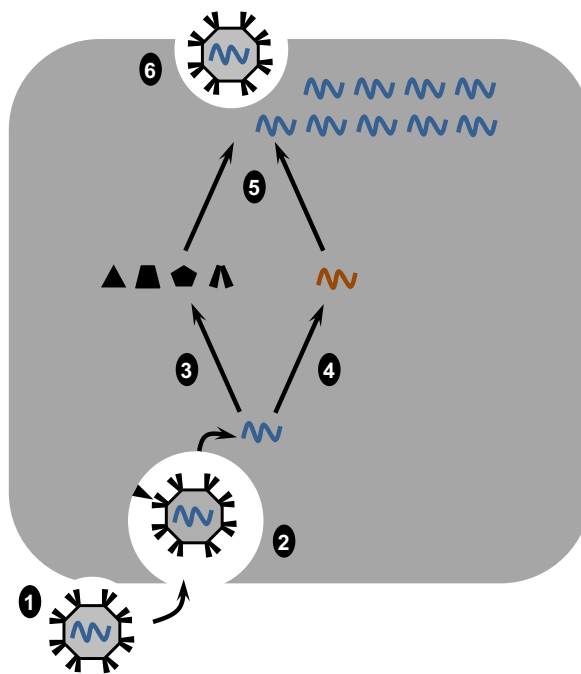


FIGURE 2.1 A schematic diagram showing the infection cycle of an RNA virus. An RNA virus particle, or virion, consists of an RNA genome (blue) surrounded by a protein coat or capsid (black). Some viruses also have a lipid envelope studded with viral proteins surrounding the capsid (not shown). The virus particle attaches to a receptor on the surface of a susceptible host cell (1), and becomes internalized (2). The viral genome codes for viral proteins (black shapes) (3) and is replicated via a replication intermediate (red) (4). Newly synthesized genomes and proteins assemble together (5) and newly made virus particles are released (6).

birds and can be transferred to further mosquitos [3]. Because RNA viruses need to replicate in these highly variable and dynamic environments, they need to be highly adaptable to maintain their existence. This adaptability is conferred by their genetic heterogeneity.

3. VIRUSES AS QUASISPECIES

In the late 1970s it was discovered that the nucleotide sequences of RNA bacteriophage, Q β , are highly heterogeneous [4], and this observation has since been extended to all RNA viruses. Accurate quantification of RNA virus mutation rates is challenging, but they have been estimated at 10^{-4} to 10^{-6} per nucleotide per round of copying [5,6]. This equates to approximately one mutation per genome replication event, which is a considerably higher rate than that of bacteria, estimated at one mutation per 300 genome replication events [5]. In addition to point mutations, recombination between viral genomes can occur in high frequency in some RNA viruses, resulting in replacement of different regions of genome sequence. Any particular RNA virus population is always in flux, with new mutations arising and deleterious mutations being lost through selection. The high mutation rate of RNA viruses, coupled with their very high levels of replication and the large population sizes that they can achieve means that RNA viruses exist as a swarm of variants rather than as a single genotype entity. Thus, RNA viruses are a genetic paradox: they are in one sense very simple entities, having very limited genetic information, but on the other hand, they are genetically complex, having the capability to access millions of sequence combinations. Adding to this complexity, there is evidence for some RNA viruses that they can exist as quasispecies in which the related genome sequences can complement each other and function cooperatively [7,8]. Thus, when a virus spreads from cell to cell and host to host, it is the properties of a swarm of genetically related but distinct viruses that enables this to occur, not the properties of a single, isogenic virus. As described in detail later in this chapter, RNA viruses require a high mutation rate to enable them to survive the varied environments that they encounter in the course of their transmission cycle. Interestingly, they also have evolved genome sequences that have a bias that allows them to rapidly adapt. However, there is also evidence that at least some have a mutation rate that is so high that they are poised at the edge of a threshold of viability, with small increases in mutation rate causing them to accumulate so many lethal mutations that they are extinguished. Together, these findings suggest that RNA viruses have evolved to have a specific mutation rate and mutation bias to enable them to survive in the particular environments in which they need to exist.

4. OVERVIEW OF RNA VIRUS REPLICATION MECHANISMS

There are several sources of genetic variability in RNA viruses, some are inherent to the biology of the virus and others are consequences of the cellular environment. The viral mutation rate is the rate at which a viral genome acquires mutations per genome replication event and is determined by the viral polymerase and any proofreading activities that the virus encodes. The mutation frequency of a virus is the frequency with which mutations accumulate over a virus infection cycle and can be impacted by the mode of virus replication, and cellular factors. Thus, to understand how viral genetic heterogeneity arises, it is helpful to have an appreciation of the mechanisms by which RNA viruses replicate their genomes. RNA viruses can be divided into different classes by virtue of their distinct genome structures and strategies of genome replication [9] (Fig. 2.2).

The riboviruses replicate their genomes via an RNA intermediate synthesized by a viral RNA-dependent RNA polymerase (RdRp). Riboviruses can have single or double-stranded RNA genomes; those with single-stranded genomes can be further characterized by being either positive or negative stranded (ie, having a genome that is of the same sense, or the opposite sense to mRNA, respectively). Riboviruses can also have genomes contained within a single piece of RNA, or a genome that is divided into multiple segments. Another class of RNA virus is the retroviruses. These viruses have an RNA genome, which is reverse transcribed by the viral reverse transcriptase enzyme into double-stranded DNA. The virus-specific double-stranded DNA then integrates into the host genome and becomes a template for cellular RNA polymerase II, which synthesizes multiple copies of RNA to generate the progeny viral genomes. It is important to appreciate that this classification system does not relate in any way to the tissues or hosts that a virus can infect, or the way in which it is transmitted to new hosts. For example, hepatitis C virus (HCV) and West Nile virus are both positive-strand RNA viruses, but they cause very different diseases and are spread in different ways.

5. THE VIRAL POLYMERASE AS A SOURCE OF ERROR

Both RdRps and reverse transcriptases have the potential to introduce deletions, insertions, and nucleotide mismatches into the nucleic acid product [10–12]. Unlike DNA-based life forms, most RNA viruses have no mechanisms to identify and repair mismatches [11,13] and so polymerase error is not corrected. The error-prone nature of polymerase activity, coupled

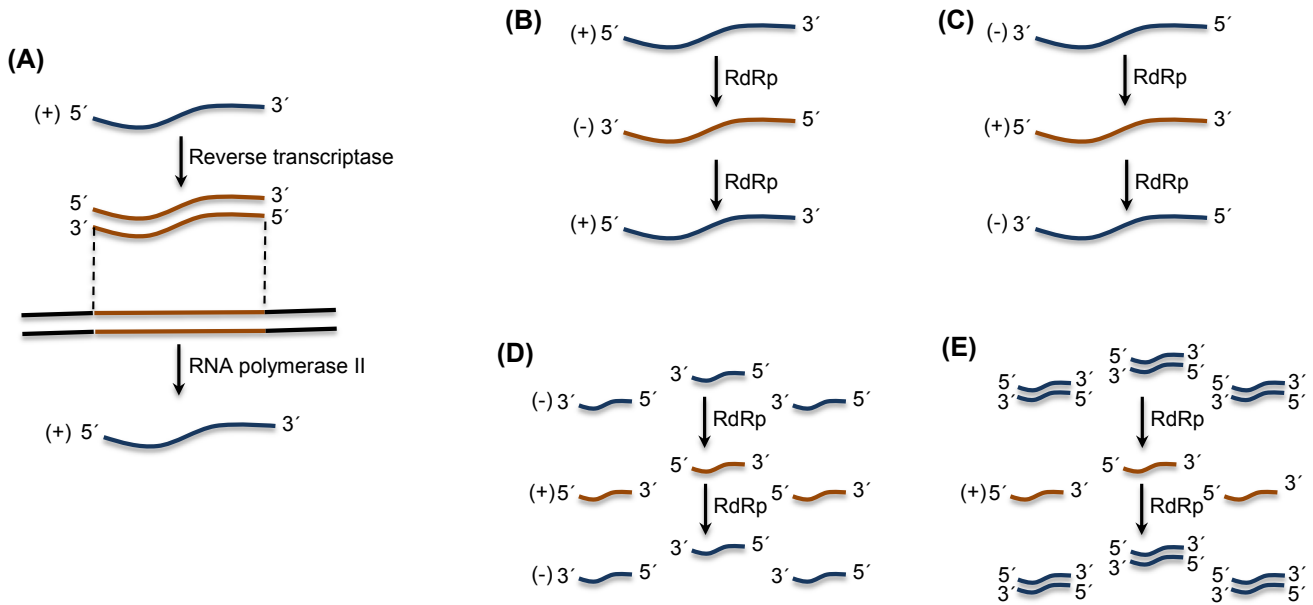


FIGURE 2.2 Examples of genome structures and genome replication strategies of RNA viruses. The examples shown are a retrovirus (A), a nonsegmented, positive-strand RNA virus (B), a nonsegmented, negative-strand RNA virus (C), a segmented, negative-strand RNA virus (D), and double-stranded RNA virus (E). The virus genomes are shown in blue and the replicative intermediates in orange. The (+) and (–) symbols indicate if the RNA is positive or negative sense (this is also indicated by the polarity of the RNA). The enzyme that performs each step of nucleic acid synthesis is shown. The replicative intermediate for the retrovirus (panel A) is DNA, whereas for the other viruses shown (panels B–E) it is RNA. Note that not all RNA virus genome structures are illustrated.

with the absence of a proofreading mechanism, is the key reason why RNA virus genomes acquire mutations and exist as a swarm of genetic variants. Although all RdRps and reverse transcriptases are capable of introducing mutations, they are not equally error prone. For example, the viral mutation rate inversely correlates with genome size, such that viruses with larger genomes have a lower per nucleotide mutation rate than those with small genomes [14]. This is intuitively logical as a high mutation rate in a virus with a large genome would increase the chance of genomes acquiring a lethal mutation and so viruses with low fidelity polymerases could not be sustained. This suggests that viruses with larger genomes have evolved to limit their mutation rate and some RNA viruses encode proteins that function to mitigate polymerase error, as described in the following. However, even when related viruses with similar genome lengths are compared, there are differences in polymerase fidelity [11,15]. For example, in a side-by-side comparison, using *in vitro* biochemical assays, the RdRp of coxsackievirus B is of higher fidelity than that of poliovirus, even though these are highly related viruses [16]. In sum, these facts suggest that polymerase error rate is determined by selection pressures related to viral genome size and other facets of virus biology.

The molecular mechanisms that govern polymerase fidelity have been elucidated by detailed enzyme kinetics studies of wild-type polymerases and by studying mutant versions of polymerase with altered fidelity [14,17–24]. These studies have shown that the error rate of the polymerase can be modulated by single amino acid substitutions in the enzyme, and that substitutions outside the active site can have an effect. Thus, the structure of the polymerase is tuned to enable it to manifest a particular fidelity. In addition to controlling the rate of replication error, polymerase determinants can also influence what substitution mutations are introduced. In a landmark study, a novel sequencing approach was employed to identify low-frequency mutations that accrued in the poliovirus genome under relatively constant conditions [25]. Viral populations present at different times were analyzed to determine what mutations accumulated in this stable environment, where selection pressure was minimized. This analysis showed that transitions occurred more frequently than transversions, and within these categories there was variation: C-to-U and G-to-A transitions accumulated more frequently than U-to-C or A-to-G. Thus, these studies indicate that there is directionality to the mutation pattern of the viral swarm. Similar findings had been made with HIV [10] and studies with West Nile virus have shown that different polymerase variants have different mutational biases [26]. Thus, RNA viruses do not incur substitutions randomly, but have a mutation bias that is likely governed by the molecular determinants of fidelity in the polymerase. This bias might play an important role in allowing the virus to generate a favorable spectrum of sequences following a genetic bottleneck.

6. OTHER VIRAL DETERMINANTS OF MUTATION RATE

Although the viral polymerase is typically the key viral factor that determines how faithfully the viral genome will be copied, it is not the only factor and there are examples of viruses in which other proteins can come into play to reduce polymerase error. As noted earlier, the genomes of all RNA viruses are relatively small compared to those of the largest DNA viruses, and it is thought that the high mutation rate of RNA virus polymerases imposes an upper limit on genome size. However, there is a wide range of genome sizes within the RNA viruses, with the largest being those of the coronaviruses, at up to 32 kb. This is more than twofold longer than most other RNA viruses. It has now become apparent that the reason why the coronaviruses can sustain this relatively large RNA genome is that they have an RNA proofreading activity [27,28] facilitated by an exonuclease that probably functions by removing incorrect insertions at the 3'-end of the RNA product during RNA synthesis [29]. Interestingly, the activity of the exonuclease is significantly enhanced by an additional coronavirus protein [30,31]. The fact that a multipartite complex performs the polymerization and RNA proofreading activities raises the intriguing possibility that coronavirus fidelity could be regulated.

Some of the nonsegmented, negative-strand RNA viruses also have an additional protein that might function to limit RdRp error. The pneumovirus subfamily has genomes of approximately 15 kb, and so they are in the midrange of RNA virus genome sizes. These viruses encode a small protein called M2-2. It has been found that deletion of M2-2 of human meta-pneumovirus results in increased accumulation of transitions, transversions, deletions, and insertions in the viral RNA, suggesting that M2-2 serves to increase the fidelity of the viral polymerase [32]. The mechanism by which M2-2 functions is not known, but it has no known enzyme activity and so it is unlikely to function as an exonuclease, but instead might serve to increase fidelity by altering RdRp structure.

A deoxyuridine-triphosphatase (dUTPase) enzyme is expressed by some, but not all, retroviruses [33,34]. This enzyme hydrolyzes dUTP and maintains low dUTP:TTP ratios, thus limiting misincorporation of deoxyuridine into viral DNA. The viral dUTPase has been shown to limit the mutation rate of feline immunodeficiency virus and caprine arthritis-encephalitis virus [35,36]. Interestingly, the primate lentiviruses, including HIV, do not encode a dUTPase, but might package a cellular DNA repair enzyme, uracil DNA glycosylase, into their virions to help limit the mutation rate [37,38].

7. RECOMBINATION

In addition to the mutations that can be introduced when the polymerase selects an incorrect nucleotide during RNA synthesis, genetic variation can also arise by recombination. Recombination can occur when two or more viral genomes enter the same cell and a part of one genome is incorporated into the other. This can result in significant changes in genome composition with dramatic impact on virus biology. For example, there is evidence that recombination might have been a factor that enabled the emergence of SARS coronavirus [39] and it is a key factor in emergence of pandemic influenza viruses [40]. However, while recombination can impact diversity, there is debate as to whether it has evolved as a means to generate variability, or is merely a consequence of viral genome replication [41]. In this respect, it is interesting that even viruses with similar genome structures can undergo different rates of recombination, perhaps suggesting that recombination is also finely tuned by evolutionary pressure. There are three mechanisms by which RNA viruses can recombine: template-switching recombination, nonreplicative recombination, and re-assortment (Fig. 2.3).

Template-switching, otherwise known as copy-choice recombination, can occur during the process of RNA synthesis if the viral polymerase transfers from one template to another, while remaining attached to the nascent nucleic acid chain [42]. This results in production of a mosaic genome. Template switching tends to occur between sequences of close similarity to give rise to a homologous recombination event. Nonhomologous recombination can also occur, but this typically results in defective genomes and is observed less frequently. Viruses differ significantly in the rate with which they can recombine by template-switching [41]. It can be highly frequent in retroviruses, particularly HIV, and also in coronaviruses. The high frequency of recombination in these viruses may be due to the replication strategies that they have. In the case of retroviruses, the reverse transcription process is highly complex and the reverse transcriptase must switch from one template to another during DNA synthesis [43]. Likewise, in coronaviruses, transcription of the genome, to allow gene expression, requires the RdRp to transfer from one site to another on the genomic template [44]. Thus, the fact that the polymerases of these viruses have evolved to transfer to a different template sequence probably means they are more likely to do so during other aspects of RNA synthesis. Recombination can occur in other positive-strand RNA viruses besides coronaviruses, and in double-stranded RNA viruses, although the recombination frequency apparently varies between viruses. For example, it occurs frequently in the positive-stranded enteroviruses, such as poliovirus, but less so in the flaviviruses, such as HCV [45]. Template switch recombination is much less frequent in the negative-strand RNA viruses [46], probably because their genomes are not naked RNA, but rather are encapsidated, or buried, in protein called nucleoprotein [47]. The polymerase

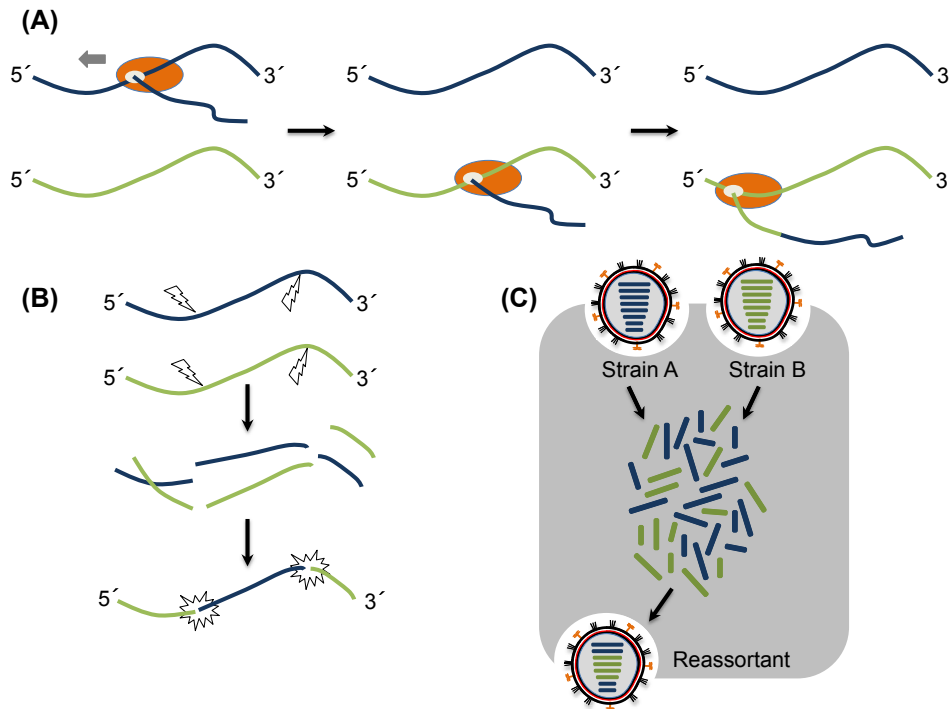


FIGURE 2.3 Mechanisms of RNA virus recombination. (A) The template-switching mechanism of recombination. The polymerase (orange) begins replicating the genome of one virus (blue), but then transfers to a genome of another related virus (green), resulting in a mosaic genome (blue and green). The *gray arrow* shows the direction of polymerase movement. (B) The nonreplicative mechanism of recombination. The genomes of two viruses (blue and green) are cleaved, either by nucleases, physical shearing, or cryptic ribozyme activity, resulting in partial genome fragments. The fragments are joined together to form a mosaic genome (blue and green). (C) A cell is infected with two viruses with different segmented genomes (blue and green). During the infection cycle the genome segments are amplified and a re-assortant virus containing a mixture of blue and green genome segments is released.

transiently displaces nucleoprotein as it moves along the template and only recognizes RNA sequences as the nucleoprotein is displaced. This probably prevents the RdRp from transferring from one genome to a similar sequence in another genome to yield functional recombinants. However, negative-strand RNA viruses containing gene duplications have arisen naturally [48] and defective interfering genomes, which contain promoters and partial genome sequences, but not complete genomes, are often detected. These findings suggest that the RdRps of the negative-strand RNA viruses are capable of jumping from one sequence to another, or that nonreplicative recombination (described later) can come into play, but perhaps most products of these events are nonviable and so are not detected.

In contrast to template switch recombination, nonreplicative recombination seems to be a relatively rare event that to date has only been described for a few positive-strand RNA viruses. This mechanism was documented by recovery of viable viruses (or replicative templates) following cotransfection of cells with two viral RNA fragments, each of which was unable to function in replication independently. The fragments recombined to form functional RNAs [49–52]. This mechanism of nonreplicative recombination might not involve a viral enzyme activity. Instead, it seems that the two RNA strands are joined together either by a transesterification reaction [49], or by ligation, presumably by cellular ligases [51–53]. Thus, RNA genome fragments created by physical shearing, nuclease cleavage, or cryptic ribozyme activity have the potential to be joined to form a novel viral genome, which can be further refined by homologous recombination to remove duplicated sequence [54].

Re-assortment is a process that can occur during coinfection of a cell with viruses with segmented genomes [55]. During re-assortment, a virus can exchange one of its own segments for that of another related virus. This process is well studied in influenza virus, in which it occurs frequently. Influenza A virus has eight genome segments that all need to be packaged into a virion for that virus particle to be infectious. This process is not completely random; there are packaging signals in the RNA segments and epistatic interactions enable the correct complement of segments to be incorporated into virions. There are many subtypes of influenza virus, but if the packaging signals of two viral subtypes coinfecting a cell are sufficiently similar, this enables a segment from one virus to be incorporated into another, resulting in release of virions with a new genome composition. Because different viral subtypes have different antigenic properties, this process has significant impact on influenza virus epidemiology [40].

8. THE EFFECT OF REPLICATION MODE ON MUTATION FREQUENCY

As described earlier, the mutation frequency of a virus differs from the mutation rate, in that it refers to the accumulation of mutations over a virus infection cycle, for example, from the point of entry of a virus into a cell until release of infectious progeny. In addition to having different genome structures and nucleic acid intermediates, different RNA viruses have different numbers of replication events per infection cycle and so this can impact on mutation frequency. Retrovirus reverse transcriptase only copies the genome twice during an infection cycle: once to generate cDNA from the RNA template and a second time to synthesize the complementary DNA strand to generate double-stranded DNA. Thus, these are the only two occasions in the retrovirus infection cycle where the viral polymerase can introduce mutations. The cellular RNA polymerase II enzyme is responsible for generating multiple copies of genome RNA that become packaged into viral particles, and while there is the potential for error to be introduced by RNA polymerase II, cellular proofreading mechanisms come into play at this step and so the major source of mutation during retrovirus genome replication is the reverse transcriptase [56]. In the riboviruses, the viral RdRp is responsible for all genome replication events and it copies the genome multiple times. Thus, in this case, there are many more opportunities for mutations to be introduced by polymerase error. Within the riboviruses, there are different modes of genome replication, referred to as a stamping machine or geometric modes, and the degree to which a virus employs one mode versus the other will affect mutation frequency [57]. In stamping machine mode, the infecting genome template is used to make multiple progeny genomes, but these genomes are not used as templates until they have been delivered into another cell. It is thought that double-stranded RNA viruses use this mode primarily. In contrast, in geometric replication, an incoming genome template acts as a template to make multiple complementary strands (or antigenomes), which in turn act as templates to make multiple genome sense strands, within the same infection cycle. In this case, there are many more opportunities for mismatch errors to be introduced than in the stamping machine mode. Positive and negative sense RNA viruses probably use a combination of both modes, but the exact contribution of each to the output virus is not well characterized, except in a few cases [57]. The mutation rate of the viral polymerase, coupled with the replication mode that the virus employs (and extrinsic factors, described in the following text) will determine the extent of genetic variability of viruses released from an infected cell.

9. THE EFFECT OF CELLULAR FACTORS ON VIRUS MUTATION RATE

The cellular environment can impact virus mutation rates and frequency. For example, dNTP pool imbalances can affect retrovirus mutation rates [58], and it has been suggested that differences in substitution rates between RNA viruses is a consequence of differences in virus RNA synthesis rates in different cell types [59]. In addition to these effects, there are also cellular factors that can result in increased mutation in RNA viruses. Adenosine-to-inosine modification by enzymes called adenosine deaminase acting on RNA (ADAR) is the most common form of RNA base modification that occurs in mammals. A-to-I conversion has important consequences in the coding potential of substrate RNAs, as inosine is decoded as a G by polymerases during template copying. The A-to-I conversion in a dsRNA duplex also has consequences to stability of RNA secondary structures, as the A:I pairing is less stable than a canonical A:U pair. This can have important consequences for RNAs that depend on their structure rather than sequence for their function [60]. ADAR modification of cellular double-stranded RNA was shown to prevent its recognition by the cytoplasmic sensor of nonself RNAs that would otherwise lead to chronic activation of innate immune pathways [61]. There is also evidence that ADAR can modify viral RNAs. Sequence analysis of RNA virus genomes has revealed that they preferentially accumulate A-to-G transitions, which are characteristic hallmarks of ADAR activity. Measles virus is a negative-stranded RNA virus, responsible for an acute disease predominantly in infants, but in rare instances associated with a fatal latent infection of the CNS known as subacute sclerosing panencephalitis (SSPE). Analysis of measles virus genomes from SSPE victims has revealed abundant A-to-G transitions, suggesting a role for ADAR in establishment of SSPE [62]. Consistent with an antiviral role for ADAR, measles virus infection of ADAR knock-out cell lines displayed increased cellular pathology, and similar findings were reported for other RNA viruses, implicating ADAR as a cellular restriction factor for a wide range of negative-stranded RNA viruses [63]. Direct evidence of ADAR modification of a viral RNA genome comes from studies of hepatitis delta virus (HDV). HDV is the smallest of the RNA viruses and encodes just two proteins, HDAg-L and HDAg-S, both of which are essential for virus viability. HDAg-L and HDAg-S share the same amino terminal open reading frame, but HDAg-L possesses a carboxyl terminal extension that is accessed when the stop codon at the end of the HDAg-S ORF is bypassed. Early during infection only the truncated HDAg-S is expressed, but then at later times expression of HDAg-L increases due to the site-specific modification of the stop codon by ADAR [64]. This editing event is highly specific and is promoted by the highly secondary structured HDV RNA genome. This action by ADAR is clearly proviral, in that without the activity of ADAR, no infectious HDV particles would form.

Another family of cellular factors that can modify the sequence of viral genomes is the APOBEC family of enzymes. These comprise an extensive arm of the innate immune system [65]. They are responsible for the modification by deamination of cytosine residues to uracil, which is an activity largely performed on single-stranded DNA substrates, leading to the phenomenon of hypermutation. APOBEC activity can affect the retroviruses. HIV infection is blocked by APOBEC, unless it expresses the viral infectivity factor (Vif). The mechanism for this blockade relies on the packaging of multiple APOBEC family members within HIV virions, which can act on the HIV genome once it has been copied by reverse transcriptase into a complementary DNA. The effect of APOBEC activity can be the modification of up to 10% of susceptible cytosine residues, resulting in a drop in infectivity of up to 100-fold.

10. MECHANISMS UNDERLYING GENETIC ROBUSTNESS IN RNA VIRUSES

Together, the studies described earlier show that there is a range of viral and host factors that combine to alter mutation frequency. The question that arises is: How do RNA viruses withstand mutation? The ability of a genome to withstand genetic or environmental perturbations without a change in phenotype is referred to as genetic robustness [66]. The high mutation rate that RNA viruses incur comes at a cost. It has been estimated that 30–40% of virus genomes generated during infection are defective [6] and so at an individual level, most viral genomes are not robust. This is not surprising: the small size of RNA virus genomes limits their coding potential and so they have limited genetic redundancy. Moreover, RNA viruses are highly compact, often containing overlapping reading frames, and nucleotide sequences that function at the RNA level, for example, as *cis*-acting elements that enable genome replication, as well as at the protein level. However, robustness is influenced by the genetic background in which it operates and so in the case of RNA viruses, genetic robustness is considered in the context of the viral swarm, rather than individual genotypes. RNA viruses are not all equally robust, and even closely related viruses can exhibit different degrees of robustness [15]. There are several factors that contribute toward this [67,68] which are described in the following paragraphs.

Robustness is conferred if a virus has an ability to more readily arrive at a new optimal or adapted genotype in the face of a changed environment, and the genetic composition of the viral swarm can facilitate this. Because the majority of nucleotide changes in RNA virus genomes are either strongly deleterious or lethal, the population is perpetually refined as deleterious genomes become purged through selection, leaving only mutations with phenotypically neutral or advantageous consequences to persist [69,70]. The neutral mutations can impact robustness. An explanation for this is that if the virus encounters a new environment, multiple nucleotide changes might need to occur for it to arrive at an optimal genotype. If some of these changes are already in place, then the jump to the new genotype is more likely to occur. This means that a population that includes a high proportion of neutral mutations will be more adaptable in the face of environmental change, as genomes with neutral mutations can act as stepping-stones toward reaching the new adapted genotype [71] (Fig. 2.4).

Thus one viral determinant of robustness is their high mutation frequency, which results in a more extensive neutral network [66,72–74]. Consequently, factors that affect mutation frequency, such as polymerase fidelity and replication mode, will also impact robustness. Interestingly, there is evidence that some RNA virus genomes have evolved to enable rapid adaptation. Experiments in which synonymous mutations were introduced into RNA virus genomes and fitness was assessed showed that the RNA nucleotide sequence has an effect on fitness, independently of its effect on protein sequence [75]. This could be due to effects on RNA structures and *cis*-acting elements. However, experiments with poliovirus showed that this might not be the only explanation. In this case, a region of the poliovirus genome that does not contain *cis*-acting RNA structures was recoded with synonymous mutations. The virus variant containing the synonymous mutations had reduced robustness and was attenuated in an animal model [76]. This finding suggests that wild-type poliovirus occupies a sequence space that enables it to rapidly adapt to environmental pressure.

Another viral determinant of robustness relates to the ability of RNA viruses to generate large numbers of genomes within individual infected cells. A consequence of the resulting polyploidy is that a genome containing a detrimental change can be complemented by the properties of another genome that is unaltered. This mechanism also has a downside in that it reduces the ability to purge poorly adapted genotypes, and thus their persistence in a population may lead to a reduction in its overall fitness. Interestingly, the huge range in the extent of polyploidy that occurs throughout the infection cycle may allow different levels of robustness at different times of the virus life cycle, with more opportunity for complementation at later stages of infection when the copy number of viral genomes is at its highest. Such a scenario may have important consequences for viruses that stimulate innate immunity early on in the infection cycle. The innate immune response poses a high adaptive requirement at a time when viral genome numbers are at their lowest. Conversely, persistent viruses that maintain high copy numbers for extended periods of time without inducing cell death, such as HCV, may be particularly robust due to the wide range of genotypes contained within the massive population of persisting RNAs. The presence of multiple genomes within the same cell can also enable recombination. Recombination is another factor that influences robustness, as it can result in purging of multiple mutations from a genome in a single recombination or re-assortment event [73].

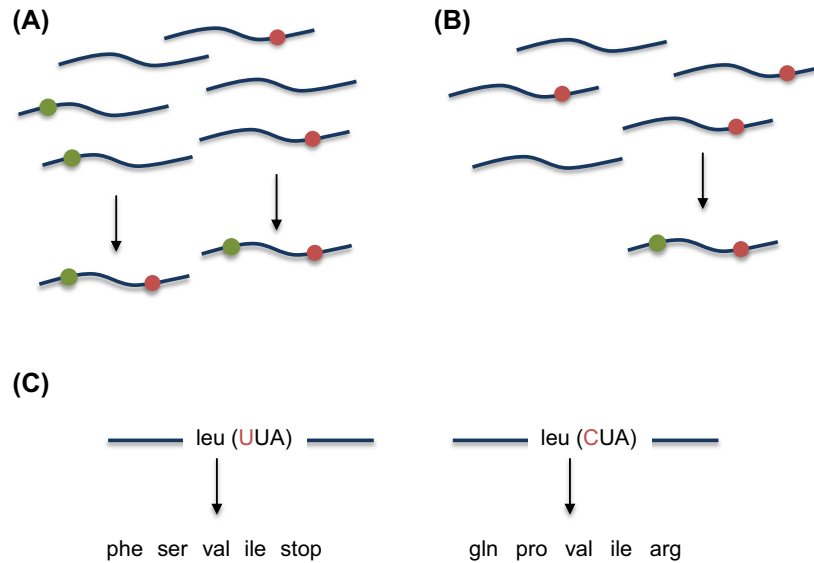


FIGURE 2.4 Mechanisms by which neutral networks can enable new genotypes to arise. (A) A neutral network containing a mixture of genomes containing green or red mutations provides a stepping-stone to a double mutant containing both mutations. (B) A neutral network can provide a stepping-stone to a new genotype by epistatic interactions. In this example, the green mutation alone is deleterious, but is neutral or beneficial in combination with the red mutation. If the neutral network contains genomes with red mutations, it provides a stepping-stone to enable introduction of the green mutation. (C) A neutral network containing genomes that have different codons for the same amino acid can provide a stepping-stone to genomes containing different spectra of amino acids following a single nucleotide substitution. In this example, a neutral network contains genomes coding for leucine at a given position, but the genomes differ by coding for leucine with either UUA or CUA. This expands the range of amino acids that could arise following a single nucleotide change.

RNA virus robustness can also be impacted by host cell factors. The ability of chaperones to buffer mutations was first proposed for the GroEL molecular chaperone [77]. Subsequently it has been experimentally observed that chaperones, such as members of the heat shock protein 70 and 90 families, play important roles in the infection cycles of many RNA viruses. It has been proposed that viruses have evolved the ability to interact with chaperones in order to buffer the effects of deleterious coding mutations that would otherwise prevent their correct folding [67,68]. This provision is particularly important as viruses depend on assembly of high-order multimers to build their capsids, a major component of the virions that are released. In these cases, a single misfolded protein has the potential to disrupt the function of the entire complex and so mechanisms that facilitate appropriate protein folding can have a significant impact.

Although there are a number of properties of RNA viruses that contribute to genetic robustness, the role of robustness in the natural history of RNA viruses is a controversial topic. A virus population with increased neutral genotypic diversity and thus high robustness can readily adapt to new environments due to its inherent diversity, and increased availability of adaptive pathways. This has important implications for viral pathogenesis and robustness has been shown to increase virulence in host organisms [76]. However, it appears that the converse can also be true and under certain conditions the neutral network can be composed of genotypes that are unable to reach a high level of fitness in the new environment [78]. This suggests that it may be difficult to make generalizations over how robustness shapes virus adaptability.

11. RNA VIRUSES ON THE EDGE

As mentioned at the beginning of this chapter, genetic heterogeneity of RNA viruses is such a key facet of their biology that it brings up the question of whether their high mutation rates have been selected for and are of evolutionary benefit. Fidelity comes at the price of elongation efficiency [16]. Thus, it is possible that the high mutation rates of RNA viruses are simply a consequence of polymerases that are under selective pressure to replicate genomes very rapidly to ensure efficient viral infection [79–81]. According to this view, RNA viruses have evolved a balance between rapid genome synthesis and error, such that the mutations that they incur are tolerable and on occasion advantageous, but are not necessary for virus survival. However, while genome synthesis rate is certainly an important factor in virus fitness [82]; for some viruses there is also evidence that the high mutation rate is beneficial and that RNA virus polymerase fidelity is tuned, enabling the virus to maximize sequence space while avoiding the accumulation of so many deleterious mutations that the genomes become nonviable. This is the concept that RNA viruses are “on the edge.”

As described earlier, most mutations that arise are deleterious and so there is a significant cost to having an error-prone polymerase. Furthermore, while complementation between defective genomes can occur, enabling genetic robustness, it is also possible for defective genomes to have an antagonistic effect, for example, by expressing mutant proteins that function as dominant negatives. Nonetheless, despite these disadvantages, it is possible to generate mutant viruses that have changes in the polymerase that result in its increased accuracy; these are known as high-fidelity mutants. Elegant studies performed with a poliovirus high-fidelity mutant showed that efficient spread within a host requires a quasispecies, and an error-prone RdRp to generate it [83,84]. Naturally, poliovirus replicates in the gut, but it can replicate in other tissues and spread to the spinal cord and brain. The ability to infect this variety of tissues requires poliovirus to overcome significant barriers to replication [85]. Experiments comparing the growth characteristics of wild type and a variant of poliovirus with a high-fidelity RdRp showed that the high-fidelity variant could replicate relatively efficiently compared to the wild-type virus in a single multiplication cycle in cell culture [83,84], and if introduced into mice intravenously, it could replicate efficiently in the spleen, kidney, and small intestine [84]. Thus, in this case, genome replication was not significantly delayed by the increased accuracy in RNA synthesis. However, in contrast to wild-type poliovirus, this high-fidelity mutant virus could not efficiently spread to the central nervous system (CNS), hence the 50% lethal dose (LD₅₀) was increased 300-fold [83,84]. To examine if the defect in virus spread was due specifically to the mutation (perhaps this variant of the RdRp could not function in a neuronal environment), or to the lack of genome diversity within its population, Vignuzzi and coworkers increased the diversity of the high-fidelity virus by treating it with mutagens. This had the dramatic effect of increasing the ability of the high-fidelity virus to replicate in the spinal cord and brain, and the LD₅₀ was restored to the same level as wild-type poliovirus. This result showed that poliovirus spread to the CNS is dependent on the virus being able to establish a highly diverse population. In addition, it was shown that coinfection of mice with wild-type and high-fidelity mutant virus enabled the high-fidelity virus to reach the brain [84]. This indicates that different viral genotypes in the quasispecies can complement each other to facilitate infection spread. It is not known exactly how complementation functions in this case, but it is easy to imagine that one variant of a virus might be more efficient at subverting innate immune defenses (which could impact virus genomes within the same cell and neighboring cells), whereas another variant might express a capsid protein better adapted to bind to a new cell receptor. In its natural context, poliovirus is spread through ingestion of contaminated water and so there is no necessity for poliovirus to be able to spread to the CNS to be able to complete its transmission cycle. However, these studies are important because they show that viruses can benefit from polymerase infidelity and a high mutation rate, particularly under conditions where they encounter a change in environment [86,87]. Studies with a number of viruses indicate that these findings are widely applicable in RNA virology and so it seems likely that RNA viruses have evolved a high mutation rate that enables them to rapidly adapt to the dynamic and varied environments in which they exist.

The studies described earlier show that RNA viruses benefit by having an error-prone polymerase to enable them to adapt to new conditions. However, there is also a cost if the polymerase has mutations that decrease its fidelity, so that the error rate is increased. Experiments performed with coxsackievirus B3 and poliovirus showed that low-fidelity mutants were able to replicate efficiently in cell culture when propagated at high multiplicity of infection (ie, when the population size was large), but were extinguished when the viruses were propagated under low multiplicity conditions, which mimics conditions when a virus first establishes infection in a host or when it has overcome a barrier, such as adaptation to a new host cell type. Consistent with these findings, both the coxsackievirus B3 and poliovirus low-fidelity mutants were attenuated in vivo. In the case of the coxsackievirus B3, low-fidelity mutants were unable to establish productive infection in the heart, the usual site for coxsackievirus B3 virus replication, and in the case of poliovirus they were unable to reach the CNS [82,88]. Comparison of the high- and low-fidelity poliovirus variants indicates how much latitude there is in the mutation rate for this virus. The high-fidelity RdRp had an approximately twofold decrease in nucleotide misincorporation rate, and the low-fidelity RdRp had a twofold increase [82]. Thus, the range in misincorporation rate that can lead to virus extinction in an animal host is not that substantial, even in a virus that is relatively genetically robust. This indicates that the fidelity of the polymerase, coupled with the impact that accuracy has on the rate of RNA synthesis, is optimized to enable viruses to adapt to the many environments in which they need to exist while avoiding extinction [82].

12. VIRUS GENETIC VARIABILITY AND THE VIRUS–HOST “ARMS RACE”

The propensity that RNA viruses have for mutation seems to have opened this up as an avenue for host cell defense. Pathogenic viruses and their hosts are engaged an epochal “arms race” in which the host evolves immune defenses to suppress virus infection and the virus in turn evolves countermeasures to disable host defenses. The existence of APO-BEC and ADAR, cellular proteins that can increase virus mutation frequency, suggests that mammalian hosts have taken advantage of the high mutation rate of viruses and evolved mechanisms to induce further mutations in the viral genomes

and push viruses toward extinction [89]. Conversely, primate lentiviruses have evolved vif, a protein that can target APO-BEC for proteosomal degradation, indicating that these retroviruses have evolved a mechanism to counter this cellular defense [90,91]. Likewise, the nonsegmented, negative-strand RNA viruses, which are susceptible to ADAR, maintain their genomes encased in a ribonucleoprotein complex throughout the infection cycle, reducing the opportunity for them to adopt double-stranded RNA structures, the substrate for ADAR. This perhaps prevents ADAR causing as much damage as it otherwise might.

13. TAKING ADVANTAGE OF THE MUTABILITY OF RNA VIRUSES

The high mutation rate of RNA viruses has often been an impediment to drug and vaccine development as viruses can rapidly gain resistance to antiviral drugs and to the immune response elicited by vaccines. However, our increasing understanding of function and consequences of genetic variability has opened new avenues for controlling viral infection. As described earlier, small decreases in polymerase fidelity can have dramatic effects on viral infectivity. Similarly, studies have shown that small increases in viral mutation rate caused by treatment with mutagenic compounds can result in significant decreases in viral fecundity [92,93]. Thus, treatment with mutagens that increase the accumulation of mutations in the viral genome can lead directly to virus extinction, or can reduce virus infection to enable effective clearance with other inhibitors, given in combination, or by host immune responses [94,95]. The identification of high-fidelity mutant viruses that can infect animals has also suggested a means to exploit these mutants as vaccine candidates. Live-attenuated virus vaccines can be highly effective, but have the disadvantage that they can potentially revert to a wild-type pathogenic phenotype. By engineering recombinant viruses with increased fidelity, it is possible to generate viruses that are attenuated, as described earlier, and that elicit protective immune responses, with reduced risk of reversion [96].

14. CONCLUSION

The RNA viruses are hugely diverse, not only in their genome structures and replication strategies, but also in their “lifestyles,” which can differ significantly, even between closely related viruses. What has emerged from studies of virus genetics is that RNA viruses are also highly divergent in terms of their polymerase fidelity, recombination rates, replication modes, and genetic robustness. We speculate that RNA viruses have evolved such that there is an intricate balance between these factors that is tuned to match the “lifestyle” of each virus, enabling it to occupy the niche in which it exists. There is some evidence to support this idea. For example, a side-by-side comparison of influenza virus RdRp and HIV reverse transcriptase showed that influenza virus RdRp is much less error prone than HIV reverse transcriptase. This may be a reflection of the fact that the influenza virus RdRp performs many more genome replication events during an infection cycle than the HIV reverse transcriptase and needs to be less error prone to avoid having a mutation frequency that is too high [97]. Another example comes from studies of West Nile virus. While the fidelity of the West Nile virus RdRp has not been directly compared to that of other viruses, there is a greater difference in fidelity between the wild-type West Nile virus RdRp and a high-fidelity mutant than has been found for most RdRps [26]. This could suggest that West Nile virus RdRp is naturally more error prone than most. This could be a necessary feature of West Nile virus to enable it to cycle back and forth between mosquito and avian hosts. Understandably, much of the work that has been performed so far has focused on viruses that are “model” viruses—those that are relatively easy to culture in vitro and replicate rapidly. However, a fuller understanding of how the factors that influence genetic diversity intertwine with virus biology will come from extending the work that has been performed so far and applying it to other viruses that have similar genome structures and replication strategies, but diverse lifestyles, such as West Nile virus and HCV, or vesicular stomatitis virus and measles virus. Research in this area will potentially be transformed by new sequencing techniques, such as CirSeq, which can detect low-level genetic variants above the background of errors introduced during RNA sequencing [25], and BAsE-Seq, a method for obtaining long stretches of sequence that can be used to identify haplotypes [98]. Ultimately, application of cutting-edge sequencing technologies, mathematical analyses, and virology studies to a range of viruses will enhance our understanding of the genesis and functional consequences of RNA virus genetic instability.

GLOSSARY

Antigenome The RNA replicative intermediate formed during viral genome replication.

Capsid A protein coat that surrounds the viral genome in the packaged virus particle. The capsid is important for protecting the viral nucleic acid and for delivering it to cells and/or to the appropriate location within cells.

cis-Acting RNA signal An element in the RNA genome that has a functional role in the replication (or transcription or translation) of the genome.

- Complementation** The ability of the products of one viral genome to provide a function that cannot be performed by the products of another viral genome.
- Copy-choice recombination** A recombination event that occurs when the viral polymerase switches to another template while remaining attached to the nascent RNA. Also known as template switch recombination.
- Epistatic mutation** A phenomenon in which mutations have different effects in combination than individually.
- Fidelity** The accuracy with which the polymerase copies the template. A high-fidelity polymerase will make fewer errors than a low-fidelity polymerase.
- Genetic robustness** The degree to which a genome can withstand environmental or genetic perturbation.
- Geometric mode** A mode of genome replication in which the newly synthesized genomes become templates for further rounds of genome replication during the infection cycle.
- Infection cycle** The cycle of events by which an infectious virus particle infecting a cell results in release of virus progeny. In the virology field, this is often referred to as the virus replication cycle, but infection cycle was used here to avoid confusion with genome replication.
- Lethal dose 50 (LD₅₀)** The quantity of infectious virus that is required to cause death in 50% of inoculated hosts.
- Live-attenuated virus vaccine** A vaccine that consists of a live (infection-competent) virus that contains mutations that reduce the disease symptoms, usually by impairing its ability to efficiently complete its infection cycle.
- Mutation rate** The rate at which a viral genome acquires mutations per genome replication event.
- Mutation frequency** The frequency at which a viral genome acquires mutations per viral infectious cycle. This frequency could be affected by cellular factors and the mode of viral replication, as well as by polymerase fidelity.
- Nonreplicative recombination** A recombination event in which two RNA fragments are joined together by either a *trans*-esterification reaction, or ligation by cellular ligases.
- Persistent virus** A virus that can infect a host and maintain the infection for extended periods of time. HIV and HCV are examples of persistent viruses.
- Polyploidy** The presence of multiple viral genomes within the same cell.
- Quasispecies** A collection of closely related viral genomes, genetically linked through mutation, that compete within a highly mutagenic environment, interact cooperatively, and collectively contribute to the population phenotype.
- Re-assortment** A recombination event that can occur with viruses with segmented genomes, in which a genome segment from one virus is packaged into a virus particle in place of a genome segment from another virus, thus producing a virus with a novel complement of genome segments.
- Retrovirus** A class of RNA viruses that replicate their genomes via a double-stranded DNA intermediate.
- Reverse transcriptase** A viral enzyme encoded by retroviruses that is responsible for generating a double-stranded DNA copy of the viral RNA genome.
- Ribovirus** RNA viruses that replicate their genomes via an RNA intermediate.
- RNA-dependent RNA polymerase** A viral enzyme encoded by riboviruses that is responsible for generating the viral genome RNA and the RNA replication intermediates.
- RNA virus** A virus that carries a genome composed of RNA in the virus particle.
- Stamping machine mode** A mode of genome replication in which the incoming genome is reiteratively used as a template to produce multiple copies of replication product.
- Swarm** A population of closely related viruses, connected through mutation, similarly to a quasispecies. We have used the term swarm in many instances here because a population of virus variants might not always fully fulfill the definition of quasispecies.
- Synonymous mutation** A nucleotide substitution that does not result in an amino acid change.
- Template switch recombination** A recombination event that occurs when the viral polymerase switches to another template while remaining attached to the nascent RNA, also known as copy-choice recombination.
- Transmission cycle** The cycle of events by which a virus is transmitted from one host to another host in the same species.

LIST OF ABBREVIATIONS

- ADAR** RNA-specific adenosine deaminase
- APOBEC** Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
- CNS** Central nervous system
- dUTPase** Deoxyuridine-triphosphatase
- HCV** Hepatitis C virus
- HDV** Hepatitis delta virus
- HIV** Human immunodeficiency virus
- kb** Kilobase
- LD₅₀** 50% lethal dose
- RdRp** RNA-dependent RNA polymerase
- SARS coronavirus** Severe acute respiratory syndrome coronavirus
- SSPE** Subacute sclerosing panencephalitis
- Vif** Viral infectivity factor

REFERENCES

- [1] de Vries RD, Mesman AW, Geijtenbeek TB, Duprex WP, de Swart RL. The pathogenesis of measles. *Curr Opin Virol* 2012;2(3):248–55.
- [2] Noyce RS, Richardson CD. Nectin 4 is the epithelial cell receptor for measles virus. *Trends Microbiol* 2012;20(9):429–39.
- [3] Chancey C, Grinev A, Volkova E, Rios M. The global ecology and epidemiology of West Nile virus. *Biomed Res Int* 2015;2015:376230.
- [4] Domingo E, Sabo D, Taniguchi T, Weissmann C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 1978;13(4):735–44.
- [5] Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics* 1998;148(4):1667–86.
- [6] Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *J Virol* 2010;84(19):9733–48.
- [7] Andino R, Domingo E. Viral quasispecies. *Virology* 2015;479–480:46–51.
- [8] Holland JJ, De La Torre JC, Steinhauer DA. RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* 1992;176:1–20.
- [9] Baltimore D. Expression of animal virus genomes. *Bacteriol Rev* 1971;35(3):235–41.
- [10] Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. *Science* 1988;242(4882):1168–71.
- [11] Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science* 1988;242(4882):1171–3.
- [12] Castro C, Arnold JJ, Cameron CE. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. *Virus Res* 2005;107(2):141–9.
- [13] Steinhauer DA, Domingo E, Holland JJ. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 1992;122(2):281–8.
- [14] Bradwell K, Combe M, Domingo-Calap P, Sanjuan R. Correlation between mutation rate and genome size in riboviruses: mutation rate of bacteriophage Q β . *Genetics* 2013;195(1):243–51.
- [15] Graci JD, Gnadig NF, Galarraga JE, Castro C, Vignuzzi M, Cameron CE. Mutational robustness of an RNA virus influences sensitivity to lethal mutagenesis. *J Virol* 2012;86(5):2869–73.
- [16] Campagnola G, McDonald S, Beaucourt S, Vignuzzi M, Peersen OB. Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *J Virol* 2015;89(1):275–86.
- [17] Arias A, Arnold JJ, Sierra M, Smidansky ED, Domingo E, Cameron CE. Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. *J Virol* 2008;82(24):12346–55.
- [18] Arnold JJ, Cameron CE. Poliovirus RNA-dependent RNA polymerase (3D^{pol}): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg²⁺. *Biochemistry* 2004;43(18):5126–37.
- [19] Arnold JJ, Vignuzzi M, Stone JK, Andino R, Cameron CE. Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. *J Biol Chem* 2005;280(27):25706–16.
- [20] Garforth SJ, Domaal RA, Lwataula C, Landau MJ, Meyer AJ, Anderson KS, et al. K65R and K65A substitutions in HIV-1 reverse transcriptase enhance polymerase fidelity by decreasing both dNTP misinsertion and mispaired primer extension efficiencies. *J Mol Biol* 2010;401(1):33–44.
- [21] Gohara DW, Arnold JJ, Cameron CE. Poliovirus RNA-dependent RNA polymerase (3D^{pol}): kinetic, thermodynamic, and structural analysis of ribonucleotide selection. *Biochemistry* 2004;43(18):5149–58.
- [22] Moustafa IM, Korboukh VK, Arnold JJ, Smidansky ED, Marcotte LL, Gohara DW, et al. Structural dynamics as a contributor to error-prone replication by an RNA-dependent RNA polymerase. *J Biol Chem* 2014;289(52):36229–48.
- [23] Skasko M, Weiss KK, Reynolds HM, Jamburuthugoda V, Lee K, Kim B. Mechanistic differences in RNA-dependent DNA polymerization and fidelity between murine leukemia virus and HIV-1 reverse transcriptases. *J Biol Chem* 2005;280(13):12190–200.
- [24] Weiss KK, Chen R, Skasko M, Reynolds HM, Lee K, Bambara RA, et al. A role for dNTP binding of human immunodeficiency virus type 1 reverse transcriptase in viral mutagenesis. *Biochemistry* 2004;43(15):4490–500.
- [25] Acevedo A, Brodsky L, Andino R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 2014;505(7485):686–90.
- [26] Van Slyke GA, Arnold JJ, Lugo AJ, Griesemer SB, Moustafa IM, Kramer LD, et al. Sequence-specific fidelity alterations associated with West Nile virus attenuation in mosquitoes. *PLoS Pathog* 2015;11(6):e1005009.
- [27] Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol* 2011;8(2):270–9.
- [28] Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol* 2007;81(22):12135–44.
- [29] Sevajol M, Subissi L, Decroly E, Canard B, Imbert I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res* 2014;194:90–9.
- [30] Bouvet M, Imbert I, Subissi L, Gluais L, Canard B, Decroly E. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl Acad Sci USA* 2012;109(24):9372–7.
- [31] Smith EC, Case JB, Blanc H, Isakov O, Shomron N, Vignuzzi M, et al. Mutations in coronavirus nonstructural protein 10 decrease virus replication fidelity. *J Virol* 2015;89(12):6418–26.
- [32] Schickli JH, Kaur J, Macphail M, Guzzetta JM, Spaete RR, Tang RS. Deletion of human metapneumovirus M2-2 increases mutation frequency and attenuates growth in hamsters. *Virol J* 2008;5:69.
- [33] Elder JH, Lerner DL, Hasselkus-Light CS, Fontenot DJ, Hunter E, Luciw PA, et al. Distinct subsets of retroviruses encode dUTPase. *J Virol* 1992;66(3):1791–4.
- [34] Threadgill DS, Steagall WK, Flaherty MT, Fuller FJ, Perry ST, Rushlow KE, et al. Characterization of equine infectious anemia virus dUTPase: growth properties of a dUTPase-deficient mutant. *J Virol* 1993;67(5):2592–600.

- [35] Lerner DL, Wagaman PC, Phillips TR, Prospero-Garcia O, Henriksen SJ, Fox HS, et al. Increased mutation frequency of feline immunodeficiency virus lacking functional deoxyuridine-triphosphatase. *Proc Natl Acad Sci USA* 1995;92(16):7480–4.
- [36] Turelli P, Guiguen F, Mornex JF, Vigne R, Querat G. dUTPase-minus caprine arthritis-encephalitis virus is attenuated for pathogenesis and accumulates G-to-A substitutions. *J Virol* 1997;71(6):4522–30.
- [37] Chen R, Wang H, Mansky LM. Roles of uracil-DNA glycosylase and dUTPase in virus replication. *J Gen Virol* 2002;83(Pt 10):2339–45.
- [38] Selig L, Benichou S, Rogel ME, Wu LI, Vodicka MA, Sire J, et al. Uracil DNA glycosylase specifically interacts with Vpr of both human immunodeficiency virus type 1 and simian immunodeficiency virus of sooty mangabeys, but binding does not correlate with cell cycle arrest. *J Virol* 1997;71(6):4842–6.
- [39] Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol* 2010;84(7):3134–46.
- [40] Yoon SW, Webby RJ, Webster RG. Evolution and ecology of influenza A viruses. *Curr Top Microbiol Immunol* 2014;385:359–75.
- [41] Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? *Nat Rev Microbiol* 2011;9(8):617–26.
- [42] Kirkegaard K, Baltimore D. The mechanism of RNA recombination in poliovirus. *Cell* 1986;47(3):433–43.
- [43] Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* 2012;2(10).
- [44] Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. *J Virol* 2007;81(1):20–9.
- [45] Gonzalez-Candelas F, Lopez-Labrador FX, Bracho MA. Recombination in hepatitis C virus. *Viruses* 2011;3(10):2006–24.
- [46] Spann KM, Collins PL, Teng MN. Genetic recombination during coinfection of two mutants of human respiratory syncytial virus. *J Virol* 2003;77(20):11201–11.
- [47] Ruigrok RW, Crepin T, Kolakofsky D. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* 2011;14(4):504–10.
- [48] Trento A, Galiano M, Videla C, Carballal G, Garcia-Barreno B, Melero JA, et al. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. *J Gen Virol* 2003;84(Pt 11):3115–20.
- [49] Chetverin AB, Chetverina HV, Demidenko AA, Ugarov VI. Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell* 1997;88(4):503–13.
- [50] Gallei A, Orlich M, Thiel HJ, Becher P. Noncytopathogenic pestivirus strains generated by nonhomologous RNA recombination: alterations in the NS4A/NS4B coding region. *J Virol* 2005;79(22):14261–70.
- [51] Gmyl AP, Belousov EV, Maslova SV, Khitrina EV, Chetverin AB, Agol VI. Nonreplicative RNA recombination in poliovirus. *J Virol* 1999;73(11):8958–65.
- [52] Gmyl AP, Korshenko SA, Belousov EV, Khitrina EV, Agol VI. Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? *RNA* 2003;9(10):1221–31.
- [53] Austermann-Busch S, Becher P. RNA structural elements determine frequency and sites of nonhomologous recombination in an animal plus-strand RNA virus. *J Virol* 2012;86(13):7393–402.
- [54] Holmblat B, Jegouic S, Muslin C, Blondel B, Joffret ML, Delpyroux F. Nonhomologous recombination between defective poliovirus and coxsackievirus genomes suggests a new model of genetic plasticity for picornaviruses. *MBio* 2014;5(4):e01114–9.
- [55] Vijaykrishna D, Mukerji R, Smith GJ. RNA virus reassortment: an evolutionary mechanism for host jumps and immune evasion. *PLoS Pathog* 2015;11(7):e1004902.
- [56] O’Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J Biol Chem* 2002;277(41):38053–61.
- [57] Safari M, Roossinck MJ. How does the genome structure and lifestyle of a virus affect its population variation? *Curr Opin Virol* 2014;9:39–44.
- [58] Julias JG, Pathak VK. Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. *J Virol* 1998;72(10):7941–9.
- [59] Hicks AL, Duffy S. Cell tropism predicts long-term nucleotide substitution rates of mammalian RNA viruses. *PLoS Pathog* 2014;10(1):e1003838.
- [60] Samuel CE. ADARs: viruses and innate immunity. *Curr Top Microbiol Immunol* 2012;353:163–95.
- [61] Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* 2015;349(6252):1115–20.
- [62] Bass BL, Weintraub H, Cattaneo R, Billeter MA. Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA. *Cell* 1989;56(3):331.
- [63] Ward SV, George CX, Welch MJ, Liou LY, Hahn B, Lewicki H, et al. RNA editing enzyme adenosine deaminase is a restriction factor for controlling measles virus replication that also is required for embryogenesis. *Proc Natl Acad Sci USA* 2011;108(1):331–6.
- [64] Casey JL. Control of ADAR1 editing of hepatitis delta virus RNAs. *Curr Top Microbiol Immunol* 2012;353:123–43.
- [65] Harris RS, Dudley JP. APOBECs and virus restriction. *Virology* 2015;479–480:131–45.
- [66] de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, et al. Perspective: evolution and detection of genetic robustness. *Evolution* 2003;57(9):1959–72.
- [67] Elena SF. RNA virus genetic robustness: possible causes and some consequences. *Curr Opin Virol* 2012;2(5):525–30.
- [68] Lauring AS, Frydman J, Andino R. The role of mutational robustness in RNA virus evolution. *Nat Rev Microbiol* 2013;11(5):327–36.
- [69] Martinez F, Lafforgue G, Morelli MJ, Gonzalez-Candelas F, Chua NH, Daros JA, et al. Ultradeep sequencing analysis of population dynamics of virus escape mutants in RNAi-mediated resistant plants. *Mol Biol Evol* 2012;29(11):3297–307.
- [70] Wright CF, Morelli MJ, Thebaud G, Knowles NJ, Herzyk P, Paton DJ, et al. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *J Virol* 2011;85(5):2266–75.
- [71] Draghi JA, Parsons TL, Wagner GP, Plotkin JB. Mutational robustness can facilitate adaptation. *Nature* 2010;463(7279):353–5.

- [72] Codoner FM, Daros JA, Sole RV, Elena SF. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog* 2006;2(12):e136.
- [73] Montville R, Froissart R, Remold SK, Tenaillon O, Turner PE. Evolution of mutational robustness in an RNA virus. *PLoS Biol* 2005;3(11):e381.
- [74] Sanjuan R, Cuevas JM, Furio V, Holmes EC, Moya A. Selection for robustness in mutagenized RNA viruses. *PLoS Genet* 2007;3(6):e93.
- [75] Cuevas JM, Domingo-Calap P, Sanjuan R. The fitness effects of synonymous mutations in DNA and RNA viruses. *Mol Biol Evol* 2012;29(1):17–20.
- [76] Lauring AS, Acevedo A, Cooper SB, Andino R. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe* 2012;12(5):623–32.
- [77] Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature* 2002;417(6887):398.
- [78] Stern A, Bianco S, Yeh MT, Wright C, Butcher K, Tang C, et al. Costs and benefits of mutational robustness in RNA viruses. *Cell Rep* 2014;8(4):1026–36.
- [79] Furio V, Moya A, Sanjuan R. The cost of replication fidelity in an RNA virus. *Proc Natl Acad Sci USA* 2005;102(29):10233–7.
- [80] Furio V, Moya A, Sanjuan R. The cost of replication fidelity in human immunodeficiency virus type 1. *Proc Biol Sci* 2007;274(1607):225–30.
- [81] Regoes RR, Hamblin S, Tanaka MM. Viral mutation rates: modelling the roles of within-host viral dynamics and the trade-off between replication fidelity and speed. *Proc Biol Sci* 2013;280(1750):20122047.
- [82] Korboukh VK, Lee CA, Acevedo A, Vignuzzi M, Xiao Y, Arnold JJ, et al. RNA virus population diversity, an optimum for maximal fitness and virulence. *J Biol Chem* 2014;289(43):29531–44.
- [83] Pfeiffer JK, Kirkegaard K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* 2005;1(2):e11.
- [84] Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 2006;439(7074):344–8.
- [85] Kuss SK, Etheredge CA, Pfeiffer JK. Multiple host barriers restrict poliovirus trafficking in mice. *PLoS Pathog* 2008;4(6):e1000082.
- [86] Borderia AV, Stapleford KA, Vignuzzi M. RNA virus population diversity: implications for inter-species transmission. *Curr Opin Virol* 2011;1(6):643–8.
- [87] Lancaster KZ, Pfeiffer JK. Viral population dynamics and virulence thresholds. *Curr Opin Microbiol* 2012;15(4):525–30.
- [88] Gnadig NF, Beaucourt S, Campagnola G, Borderia AV, Sanz-Ramos M, Gong P, et al. Coxsackievirus B3 mutator strains are attenuated in vivo. *Proc Natl Acad Sci USA* 2012;109(34):E2294–303.
- [89] Dapp MJ, Patterson SE, Mansky LM. Back to the future: revisiting HIV-1 lethal mutagenesis. *Trends Microbiol* 2013;21(2):56–62.
- [90] Bonnac LF, Mansky LM, Patterson SE. Structure-activity relationships and design of viral mutagens and application to lethal mutagenesis. *J Med Chem* 2013;56(23):9403–14.
- [91] Strebel K. HIV accessory proteins versus host restriction factors. *Curr Opin Virol* 2013;3(6):692–9.
- [92] Crotty S, Cameron CE, Andino R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci USA* 2001;98(12):6895–900.
- [93] Loeb LA, Essigmann JM, Kazazi F, Zhang J, Rose KD, Mullins JI. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc Natl Acad Sci USA* 1999;96(4):1492–7.
- [94] Anderson JP, Daifuku R, Loeb LA. Viral error catastrophe by mutagenic nucleosides. *Annu Rev Microbiol* 2004;58:183–205.
- [95] Graci JD, Cameron CE. Therapeutically targeting RNA viruses via lethal mutagenesis. *Future Virol* 2008;3(6):553–66.
- [96] Vignuzzi M, Wendt E, Andino R. Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* 2008;14(2):154–61.
- [97] Aggarwal S, Bradel-Tretheway B, Takimoto T, Dewhurst S, Kim B. Biochemical characterization of enzyme fidelity of influenza A virus RNA polymerase complex. *PLoS One* 2010;5(4):e10372.
- [98] Hong LZ, Hong S, Wong HT, Aw PP, Cheng Y, Wilm A, et al. BAsE-Seq: a method for obtaining long viral haplotypes from short sequence reads. *Genome Biol* 2014;15(11):517.