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

Molecular basis of genetic variation of viruses: error-prone replication

Abbreviations

A adenine
ADAR adenosine deaminase acting on double-stranded RNA
AIDS acquired immune deficiency syndrome
AMV-RT avian myeloblastosis virus reverse transcriptase
APOBEC apolipoprotein B mRNA editing complex
BHK baby hamster kidney cells
bp base pairs
C cytosine
cDNA complementary DNA
CVB3 coxsackievirus B3
DI defective interfering
DI RNAs defective interfering RNAs
E. coli *Escherichia coli*
FMDV foot-and-mouth disease virus
G guanine
HBV hepatitis B virus
HCV hepatitis C virus
HIV-1 human immunodeficiency virus type 1
HIV-RT human immunodeficiency virus reverse transcriptase
HPV human papillomavirus
IFN interferon
I inosine
IRES internal ribosome entry site
Kb kilobase
miRNA micro-RNA
mRNA messenger RNA
NTP-dependent nucleoside triphosphate-dependent
OPV oral poliovirus vaccine
PCR polymerase chain reaction
PV poliovirus
RdRp RNA-dependent RNA polymerase
rRNA ribosomal RNA
RT reverse transcriptase
siRNA small interfering RNA
T thymine
U uracil
VSV vesicular stomatitis virus

2.1 Universal need for genetic variation

Genetic change was a prerequisite for the early life forms to be generated and maintained (Chapter 1), and it is also a requirement for the evolution of present-day life. We may willingly or inadvertently modify selective pressures, but genetic change is rooted in all replication machineries. The results of genetic modifications, regarding relative dominances of variant forms, are guided by selective pressures and random events. The replicative machinery itself has probably been influenced by natural selection; as an example, polymerases devoid of a capacity to generate variants should have endured a long-term selective disadvantage. However, once the replicative machinery was established, the mechanisms of variation acted independently of the selective pressures applied or to come.

https://doi.org/10.1016/B978-0-12-816331-3.00002-7

© 2020 Elsevier Inc. All rights reserved.

Viruses use the same molecular mechanisms of genetic variation than other forms of life: mutation (that encompasses point mutations and insertions-deletions of different lengths), hypermutation, several types of recombination, and genome segment reassortment. Mutation is observed in all viruses, with no known exceptions. Recombination is also widespread, but its role in the generation of diversity appears to vary among viruses. Its occurrence was soon accepted for DNA viruses, but it was considered uncertain for the RNA viruses. Pioneering studies of poliovirus (PV) by P. D. Cooper, V. I. Agol and colleagues, and of foot-and-mouth disease virus (FMDV) by A.M. King and colleagues provided the first evidence of recombination in RNA. The present perception is that recombination is more widespread than thought only a few decades ago and that its frequency and the types of genomic forms it generates are varied among viruses. For example, it appears that positive-strand RNA viruses recombine more easily than negative-strand RNA viruses to give rise to mosaic genomes of standard length. Several negative-strand RNA viruses, however, can yield defective genomes through recombination, frequently characterized by deletions in their RNA. A connection between the structure of replication complexes—as viewed by X-ray diffraction or high-resolution cryo-electron microscopy—and the propensity to produce defective genomes has not been established. Defective genomes are increasingly perceived not only as unavoidable side-products of blind replicative imperfections but as classes of genome subpopulations that perform relevant biological roles for the standard, infectious viruses. Genome segment reassortment, a type of variation close to chromosomal exchanges in sexual reproduction, is an adaptive asset of segmented viral genomes, as continuously evidenced by the ongoing evolution of the influenza viruses. The three modes of virus genome variation are not incompatible, and reassortant-recombinant-mutant genomes are continuously arising in present-day viruses. The potential for genetic variation of RNA and DNA viral genomes is remarkable, and it is the ultimate molecular mechanism that lies at the origin of the virus diversity delineated in Chapter 1.

2.2 Molecular basis of mutation

Mutation is a localized alteration of a nucleotide residue in a nucleic acid. It generally refers to an inheritable modification of the genetic material. In the case of viral genomes, mutations can result from different mechanisms: (i) template miscopying (direct incorporation of an incorrect nucleotide); (ii) primer-template misalignments that include miscoding followed by realignment, and misalignment of the template relative to the growing chain (polymerase "slippage" or "stuttering"); (iii) activity of cellular enzymes (i.e., deaminases), or (iv) chemical damage to the viral nucleic acids (deamination, depurination, depyrimidination, reactions with oxygen radicals, direct and indirect effects of ionizing radiation, photochemical reactions, etc.) (Naegeli, 1997; Bloomfield et al., 2000; Friedberg et al., 2006).

The basis of nucleotide misincorporation during template copying (defined as the incorporation of a nucleotide different from that expected from the template residue at that position) lies mainly in the electronic structure of the bases that make up DNA (adenine, A; guanine, G; cytosine, C; thymine, T) or RNA (with uracil, U instead of T). Each base includes potential hydrogen-bonding donor sites (amino or amino protons) and hydrogen-bonding acceptor sites (carbonyl oxygens or aromatic nitrogens) that contribute to standard Watson-Crick base pairs (Fig. 2.1), as well as wobble base pairs (nonstandard Watson-Crick, but fundamental for RNA secondary structure and mRNA translation) (Fig. 2.2). The conformation of the purine and pyrimidine bases is highly dynamic. Amino and methyl groups rotate about the bonds that link them to the ring structure. In dilute solution,

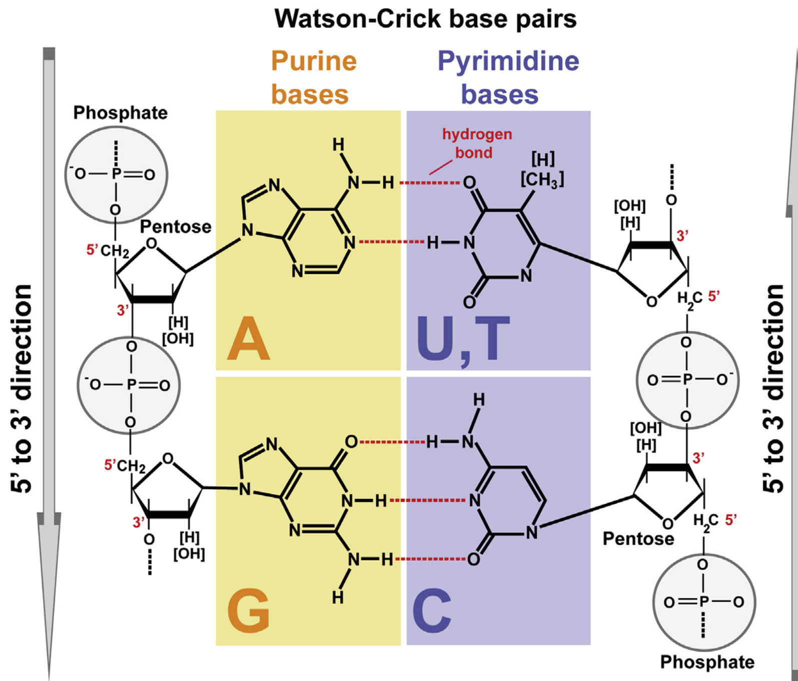

FIGURE 2.1 The standard Watson-Crick base pairs in DNA (A-T and G-C, with deoxyribose as pentose) and RNA (A-U and G-C, with ribose as pentose). Phosphodiester bonds of two potential polynucleotide chains of different polarity (outer arrows) are indicated.

hydrogen bonds are established with water, and they can be displaced by nucleotide or amino acid residues to give rise to nucleotide-nucleotide or nucleotide-amino acid interactions. The strength difference between hydrogen bonds established in a polynucleotide chain with water, and their strength between two bases in separate polynucleotide chains determines whether a double-stranded polynucleotide will be formed.

Purine and pyrimidine bases can acquire different charge distributions and ionization states. As a consequence, in addition to the standard Watson-Crick and wobble, other base pairs are found in naturally occurring nucleic acids (notably cellular rRNA and tRNA) and in synthetic oligonucleotides (A-U or A-T Hoogsteen, and A-G, C-U, G-G, and U-U pairs, as well as interactions involving ionized bases). One of the types of electronic redistribution leads to tautomeric changes, such as the keto-enol and amino-imino transitions, which modify the hydrogen-bonding properties of the base; tautomeric imino and enol forms of the standard bases can produce non-Watson-Crick pairs. The proportion of the alternative tautomeric forms can be influenced by modifications in the purine and pyrimidine rings, which, in turn, can favor either the *syn* or *anti* conformation of a nucleoside, which is defined by the torsion angle of the bond between the 1′ carbon of the ribose and either N1 in pyrimidines or N9 in purines (Fig. 2.1). The *anti*-conformation is usually the most stable in standard nucleotides and polynucleotides. The transition from the *anti* to the *syn* conformation may alter the hydrogen-bonding properties of the base, thereby inducing mutagenesis (Bloomfield et al., 2000; Suzuki et al., 2006). The understanding of conformational

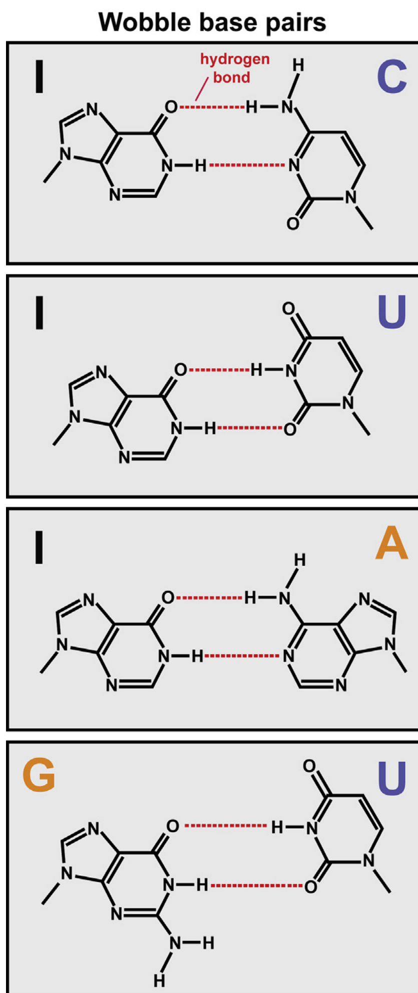

FIGURE 2.2 Examples of a class of non-Watson-Crick base pairs termed wobble base pairs. The drawing is similar to that of Fig. 2.1, except that the sugar residues and phosphodiester bonds have been omitted. Hydrogen bonds (*discontinuous lines in red*) are shown between I (inosine) and C, U, and A, and between G and U. Wobble base pairs are important for codon-anticodon interactions, as described in the text.

effects on the base-pairing tendencies of nucleoside analogs in the context of the active site of a polymerase is very relevant to the design of specific mutagenic analogs for viral polymerases in lethal mutagenesis-centered antiviral approaches (Chapter 9).

Base-base interactions are not only responsible for part of the mutations that occur during genome replication, but also for the formation of double-stranded nucleic acids, either within the same polynucleotide chain or between two different chains. Transitions from a coil-like into an organized double-stranded (or other) structure are functionally relevant for both RNA and DNA. In the case of RNA, double-stranded regions in the adequate alternation with single-stranded regions, determine key catalytic or macromolecule-attracting abilities, as for example, ribozyme activities (Chapter 1), the internal ribosome entry site (IRES) of several viral and cellular mRNAs, or multitudes of functional RNA-protein interactions (Denny and Greenleaf, 2018).

Adjacent base stacking due to electronic interactions (rather than hydrophobic bonds as once thought), contribute also to the stability of double-helical regions in nucleic acid molecules. Structural transitions due to alternative stacking conformations, particularly within polypurine or polypyrimidine tracts, can affect nucleic acid-protein interactions. In turn, replication machineries (typically including viral and host proteins gathered in membrane structures) may also be affected by nucleic acid conformations; such effects are important in virology regarding consequences for mutant generation in a given template sequence context. These considerations on structural transitions are relevant to the non-neutral character of silent (also termed synonymous) mutations (those in open-reading frames that do not result in an amino acid substitution), a point to be addressed in the next section. Transitions from a single-stranded into a double-stranded nucleic acid structure and the relative stability of the two forms depend on multiple factors that include the nucleotide sequence of the nucleic acid, its being a ribo- or a deoxyribose-polynucleotide, temperature, ionic environment, and ionic strength. Positively charged counterions neutralize negatively charged phosphates, and favor duplex stability [as an overview of

physical and chemical properties of nucleic acids and their nucleotide components, see (Bloomfield et al., 2000)].

2.3 Types and effects of mutations

Mutations resulting from any of the mechanism just summarized can be divided into transitions, transversions (both referred to as point mutations), and insertions and deletions (referred to as *indels*) (Fig. 2.3). The latter occurs preferentially at homopolymeric tracts and also at short, repeated, sequences which are prone to misalignment mutagenesis (Fig. 2.4). An example is an editing mechanism for some viral mRNAs, such as the phosphoprotein mRNA of the *Paramyxovirinae* [(Kolakofsky et al., 2005) and references therein]. Other examples in vivo are hot spots for variation in reiterated sequences in complex DNA genomes (Yamaguchi et al., 1998; Barrett and McFadden, 2008; McGeoch et al., 2008), or the insertion of two amino acids (often Ser-Ser, Ser-Gly, or Ser-Ala between residues 69 and 70 of the HIV-1 reverse transcriptase), in concert with HIV-1 resistance to nucleoside inhibitors (Winters and Merigan, 2005) (Chapter 8). Hairpin structures in RNA and DNA may also induce deletions as a result of slippage mutagenesis (Pathak and Temin, 1992; Viguera et al., 2001). Transition mutations occur more frequently than either transversions or *indels* during virus replication. Nucleotide discrimination at the catalytic site of viral polymerases fits this observation because of the more likely replacement of a purine or pyrimidine nucleotide by its structurally more similar nucleotide. In some cases, however, an abundance of *indels* and similar numbers of transitions and transversions have been recorded (Cheynier et al., 2001; Malpica et al., 2002). The molecular bases of such unexpected behavior regarding mutational spectra are not well understood.

The generation of point mutations and *indels* is subject to thermodynamic and quantum-mechanical uncertainties inherent to atomic fluctuations, rendering mutagenesis a highly unpredictable event, thus introducing stochasticity (randomness) in a key motor of evolution: the generation of diversity at the molecular level (Domingo et al., 1995; Eigen, 2013).

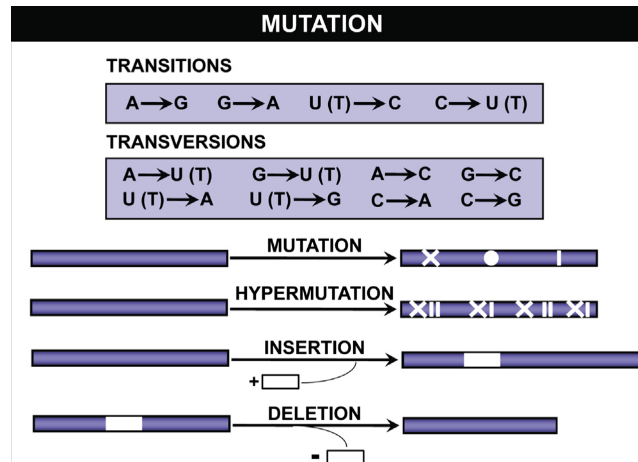

FIGURE 2.3 Major types of mutations in RNA (U) and DNA (T): four transitions and eight transversions. Below, a means to indicate point mutations, insertion or deletions (known as *indels*). A genome is depicted as an elongated rod. Symbols on the rod (*cross, circle, and line*) represent mutations. Hypermutation is generally associated with a high frequency of specific mutation types (*crosses and lines*). A region inserted or deleted from the genome is depicted as an empty rod.

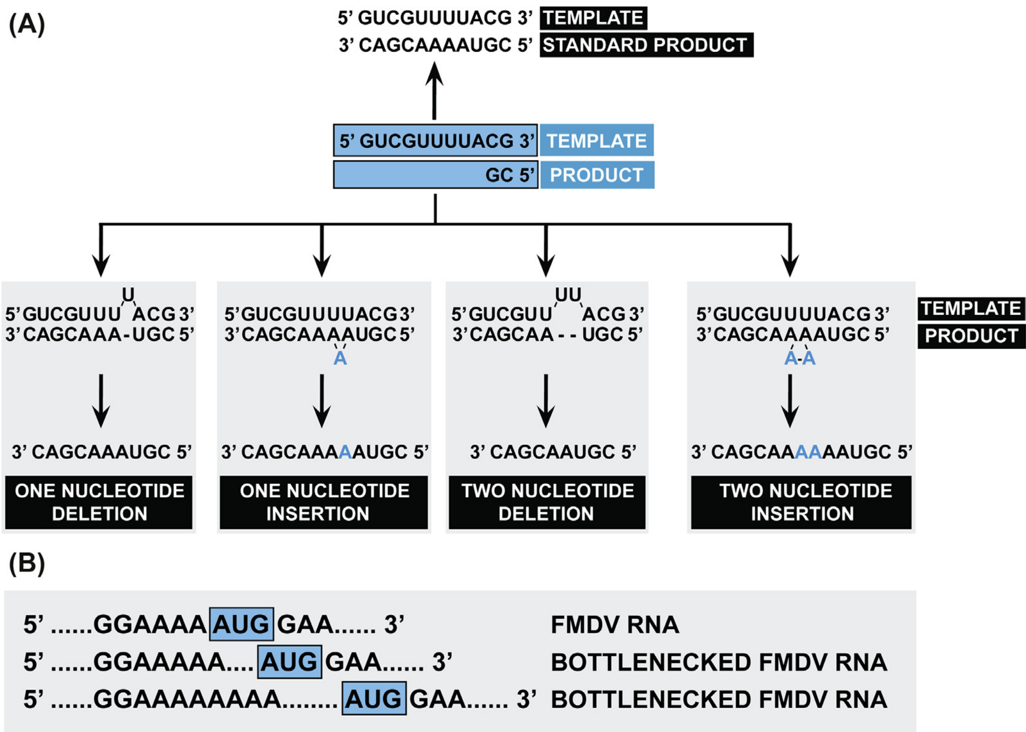

FIGURE 2.4 Misalignment mutagenesis. (A) Production of one or two nucleotide deletion or insertion at a homopolymeric tract in an RNA template. The basis of these events is a displacement of template or product residues during RNA synthesis. (B) A specific example of an internal oligoadenylated extension in the genome of FMDV subjected to plaque-to-plaque transfers. See text for biological implications and references.

The effect of mutations on the structure and function of proteins is extremely relevant to penetrate into the mechanisms that drive virus evolution since selection acts on phenotypes that are often embodied in protein molecules. Silent or synonymous mutations are those that do not give rise to an amino acid substitution despite being located in a protein-coding region of a genome. Their occurrence is due to the degeneracy of the genetic code: the same amino acid can be coded for by two or more triplets (codons), with the exception of AUG for methionine and UGG for tryptophan. Synonymous mutations are not necessarily selectively neutral, neutral meaning that they have no discernible consequence for any viral function. The assumption that synonymous mutations are selectively neutral, and the fact that the early comparison of nucleotide sequences of homologous genes showed a dominance of synonymous over nonsynonymous mutations, contributed to the foundations of the neutral theory of molecular evolution. This theory attributes the evolution of organisms at the molecular level mainly to the random drift of genomes carrying neutral or quasi- (or nearly-) neutral mutations (King and Jukes, 1969; Kimura, 1983, 1989). The terms quasi-neutral or nearly-neutral may seem ambiguous to molecular biologists. However, in the formulation of the neutral theory they had a precise meaning of the selection coefficient (a parameter that measures fitness differences) being lower than the inverse of the effective population size, with minor variations in the equations of some formulations (Kimura, 1983).

Despite random drift of genomes playing an important role in molecular evolution, evidence gathered over the last decades renders untenable the assumption that synonymous mutations are neutral. Evidence to the contrary has been obtained with viruses and cells, including mutations in the human genome that may affect enhancer functions (Hirsch and Birnbaum, 2015), mRNA folding (Faure et al., 2017; Mittal et al., 2018), and microRNA targeting (Brest et al., 2011), among other processes (Novella, 2004; Novella et al., 2004; Parmley et al., 2006; Hamano et al., 2007; Resch et al., 2007; Lafforgue et al., 2011; Nevot et al., 2011, 2018; Supek, 2016). There are several mechanisms by which synonymous mutations can affect virus behavior: alteration of *cis*-acting regulatory elements in viral genomes, decrease of the stability of duplex structures within the RNA genome or between viral sequences and miRNAs or siRNAs, or changes of viral gene expression [splicing precision or translation fidelity through the modification of RNA-RNA or RNA-protein interactions; reviewed in (Martínez et al., 2016)].

Synonymous codons use different tRNAs for protein synthesis, and different tRNAs do not have the same relative abundance in different host cell types. Thus, the rate of protein synthesis, an important phenotypic trait for cells and viruses, can be affected by the frequency of alternative synonymous codons present in mRNAs (Richmond, 1970; Akashi, 2001). Not only codon bias, but also specific codons or codon combinations may affect ribosome speed to regulate the folding of nascent proteins during translation (Makhoul and Trifonov, 2002; Rocha, 2004; Aragones et al., 2010; Brule and Grayhack, 2017). As a consequence, generation of rare codons by mutation of abundant codons (or vice versa) can modify viral fitness (Chapter 5). Rare codons may also limit the fidelity of amino acid incorporation when the frequency of the required aminoacyl-tRNAS is low (Ling et al., 2009; Zaher and Green, 2009; Czech et al., 2010). The frequency of codon pairs in RNA genomes is also a fitness determinant relevant to the preparation of attenuated viral vaccines.

To complicate matters further, a synonymous mutation may be neutral or quasi-neutral in one environment, but it may contribute to selection in a different environment, because of the phenotypic effects of RNA structure and codon usage. Neutrality is relative to the environment.

Regarding the effects of mutations (Box 2.1), the following general statements are applicable to viruses:

- Although difficult to prove due to the limited number of environments used for experimentation, truly neutral mutations (i.e., with no influence on the virus in any environment) are probably very rare. This applies to synonymous, as well as to non-synonymous mutations.
- Mutations resulting in chemically conservative amino acid substitutions are more likely to be tolerated than those leading to chemically different amino acids. Tolerability (quantified by substitution matrices among amino acids in protein evolution) should be distinguished from neutrality. A tolerated mutation may cause a reduction in fitness, which is nevertheless compatible with virus replication.
- A conservative amino acid substitution may have important biological consequences.
- The effect of any individual mutation is context-dependent in two ways: it may depend on other mutations in the same genome (epistasis, see also Section 2.8 and Chapter 5) or on the mutant cloud that surrounds the genome harboring the mutation (effects of complementation, cooperation, or interference, discussed in Section 3.8 of Chapter 3).
- The previous points do not deny the influence of random drift of genomes on intrahost and interhost evolution. The currently most accepted view is that positive and negative selection and random drift occur continuously during virus evolution (Chapter 3).

BOX 2.1

The effects of mutations on viruses

In noncoding regions

Mutations may affect stem-loop or other secondary and higher-order structures involved in regulatory processes through nucleic acid-nucleic acid or nucleic acid-protein interactions. The primary sequence in nonstructured, noncoding regions may also be functionally relevant.

In coding regions

- Synonymous or silent mutations do not affect the amino acid sequence of the encoded protein.
- Nonsynonymous mutations give rise to an amino acid substitution in the encoded protein.
- Some mutations may generate a stop codon, leading to a truncated protein.

Regarding functional effects

- Neutral mutations are those that have no functional effects.
- Nonneutral mutations can have a broad range of fitness effects: from nearly complete tolerability to lethality.
- Most proteins are multifunctional. A nonsynonymous mutation can affect one but not other functions performed by the same protein.

Context-dependence

The effect of a mutation may be context-dependent in two manners: it may be affected by other mutations in the same genome (epistasis) or by other genomes of the surrounding mutant spectrum.

2.4 Inferences on evolution drawn from mutation types

The proportion of transition versus transversion mutations may depend initially on the specific replication machinery of a virus that tends to produce some mutation types preferentially over others. For a given virus, short-term evolution is often reflected in the dominance of transitions, a dominance which is less apparent when distantly related sequences of the same virus are compared. The effect of evolutionary distance on the transition to transversion ratio was observed in the FMDV genome sequence comparisons carried out in our laboratory over several decades, that ranged from analyses of mutant spectra relative to their corresponding consensus sequence to independent viral isolates from disease outbreaks separated by several decades [review of the work on FMDV evolution in (Domingo et al., 1990, 2003)]. These two levels of sequence comparisons (within mutant spectra vs. independent isolates) can be highly significant, as discussed in Chapters 3 and 7.

The proportion of synonymous and nonsynonymous mutations that have mediated the diversification of viral genomic sequences that belong to the same phylogenetic lineage is often considered informative of the underlying evolutionary forces. Probably because of the rooted (albeit uncertain) notion that biological function is more likely to reside in protein than in DNA or RNA, the ratio of nonsynonymous substitutions (corrected per nonsynonymous site in the sequence under study) (d_n), to the number of synonymous substitutions per synonymous site (d_s), termed ω ($\omega = d_n/d_s$) is calculated to infer the dominant mode of evolution (Nei and Gojobori, 1986).

When $\omega = 1$ the evolution is considered neutral, when $\omega < 1$ purifying (or negative) selection is dominant, and when $\omega > 1$ positive (or directional) selection prevails (Yang and Bielawski, 2000). The types of selection undergone by viruses are discussed in Section 3.4 of Chapter 3.

There are several reasons to be cautious about the significance of ω: (i) synonymous mutations need not be neutral, for reasons discussed in Section 2.3. (ii) In the course of evolution, important but transient events of positive selection (termed episodic positive selection) due to one or a few amino acid substitutions may be accompanied by a larger number of synonymous, tolerated mutations. In this situation, ω computes as $\omega < 1$, thus indicative of purifying selection despite a critical role of positive selection triggered by one or few nonsynonymous mutations in the evolutionary outcome (Crandall et al., 1999). (iii) In a striking proof of the above arguments, statistically significant mutational biases led to a value of ω indicative of positive selection in an in vitro evolution experiment simulating pseudogene evolution in which positive selection was not possible (Vartanian et al., 2001); this study represents a warning which is rarely mentioned when discussing the limitations of conclusions based on the value of ω. (iv) A synonymous change may permit the mutant codon to acquire a relevant nonsynonymous change through a point mutation. The term *quasisynonymous* has been used to describe codons that encode the same amino acid, but that has a different evolutionary potential regarding the amino acids that they can access through a point mutation. Alternative codons for a given amino acid approximate a replicative system to points of sequence space from which a phenotypically relevant change has a different probability (Chapters 3 and 4). (v) Finally, ω was initially proposed to compare distantly related rather than closely related genomes, as is often the case in the short-term evolution of viruses (Kryazhimskiy and Plotkin, 2008).

For all these reasons, ω values as a diagnostic of forces mediating DNA and RNA virus evolution must be regarded only as indirect and suggestive, not as a definitive parameter. Despite these arguments, use of ω to propose a model of virus evolution continues being surprisingly unchallenged in the literature of virus evolution. We use ω only in a limited way in subsequent chapters because, in addition to the limitations just listed, it does not help in the interpretation of critical evolutionary events regarding viruses. Related shortcomings apply to other tests of neutrality developed to interpret the origin of DNA polymorphisms in the years following the summit of the neutralist-selectionist controversy (Fu, 1997; Achaz, 2009).

2.5 Mutation rates and frequencies for DNA and RNA genomes

Mutation rates quantify the number of misincorporations per nucleotide copied, irrespective of the fate (increase or decrease in frequency) of the mutated genome produced. A mutation rate for a genomic site measures a biochemical event dictated by the replication machinery and environmental parameters that affect the catalytic properties of the polymerase. In contrast, a mutant (or mutation) frequency describes the proportion of a mutant (or a set of mutants) in a genome population. The frequency of a mutant will depend on the rate at which it is generated (given by the mutation rate) and on its replication capacity relative to other genomes in the population (Drake and Holland, 1999) (Fig. 2.5). A specific mutation may be produced at a modest rate, but then be found at high frequency because the mutation is advantageous for genome replication in that environment. The converse situation may also occur. Some mutational hot spots (in the sense of genomic sites where mutations tend to occur with high probability) may never be reflected among the repertoire of mutations found in a

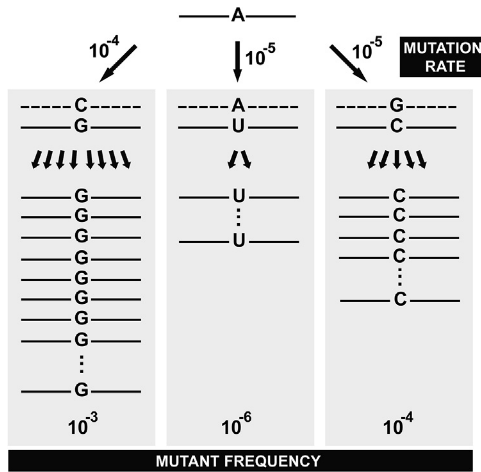

FIGURE 2.5 Scheme that illustrates the difference between mutation rate and mutant frequency. Residue A in a template residue (top) can be misread to incorporate a C, A, or G into the complementary strand (discontinuous lines), at a rate of 10^{-4}, 10^{-5}, and 10^{-5} substitutions per nucleotide, respectively. The replicative capacity of the newly generated templates (with G, U, and C, *continuous lines*) will determine widely different mutant frequencies with G > C > U.

genome population because of the selective disadvantage they inflict upon the genome harboring them.

A very significant example is the elongation of an internal oligoadenylate tract located between the two functional AUG initiation codons in the FMDV genome. The homopolymeric tract constitutes a hot spot for variation due to polymerase slippage (Fig. 2.4B). The elongation of the internal oligoadenylate was dramatic because it sextuplicated the number of adenylate residues present at that site; it was only observed when FMDV was subjected to repeated plaque-to-plaque (bottleneck) transfers, not large population passages. In fact, this drastic genetic modification has not been recorded among natural isolates of the virus. The molecular instruction to elongate the oligoadenylate was very strong because it was observed in many independent biological clones subjected to bottleneck transfers (Escarmís et al., 1996). Despite qualifying as a hot spot for variation, the first event in fitness recovery when the clones were subjected to large population passages was the reversion of the elongated tract to its original size (Escarmís et al., 1999). The interpretation of these findings, to be further analyzed in Chapter 6, is that during plaque-to-plaque transfers the negative selection to eliminate unfit genomes is less intense than during large, highly competitive population passages. Again, a clear molecular instruction to elongate a homopolymeric track may not be reflected in a high frequency of the affected genomes. Therefore, although mutation rates and frequencies for viruses bear some relationship, rates cannot be inferred from frequencies and vice versa (Fig. 2.5).

The first calculations of mutation rates for cellular organisms and for some DNA bacteriophages were carried out by J.W. Drake, who pursued comparative measurements that have generally supported a difference between mutation rates for DNA and RNA viruses. The rates estimated for bacteriophages λ and T4 were about 100 times higher than those of their host *E. coli*. An approximately constant rate of 0.003 mutations per genome per replication round was calculated for a number of DNA-based microbes (Drake, 1991), an observation sometimes referred to as "Drake's rule." This rather surprising constancy suggests that different DNA organisms have accommodated the template-copying fidelity of their replication machineries to achieve a narrow window in the mutational load measured as mutations fixed per genome, a remarkable fitting of biochemistry with evolutionary needs. The basal mutation rate in mammalian cells has been estimated at about 10^{-10} substitutions per nucleotide and cell generation [reviewed in (Naegeli, 1997; Domingo et al., 2001; Friedberg et al., 2006)] (Table 2.1).

The synonymous mutation rate measured with experimental populations of bacteria has been assumed to reflect the neutral mutation rate (despite limitations explained in Section 2.3). Values for *E. coli* have ranged from 2×10^{-11} up

TABLE 2.1 Mutation rates and frequencies for RNA and DNA genomes.

Viriods	**2×10^{-3}**
RNA viruses	10^{-5} to 10^{-3}
Retroviruses	10^{-6} to 10^{-4}
DNA viruses	10^{-8} to 10^{-3}
Cellular DNA	10^{-9} to 10^{-11}

Values are expressed as substitutions per nucleotide. The range of values is the most likely according to several independent studies. No distinction is made between mutation rates and frequencies. See text for comments and references.

to 5×10^{-9} substitutions per synonymous site per generation (Ochman et al., 1999) with 5×10^{-10} as the most likely estimate (Lenski et al., 2003). The latter value is in agreement with a rate of 3×10^{-10} to 4×10^{-10} substitutions per base pair and generation based on whole-genome deep sequencing of an experimentally evolved lineage of *Myxococcus xanthus* (Velicer et al., 2006). There are biological phyla for which no mutation rates have been calculated. From current knowledge, we can assume that mutation rates in cells and viruses depend on the replicative machinery (generally a multiprotein complex that includes the relevant viral polymerase with additional viral and host proteins and membrane structures) and on multiple environmental parameters (template nucleotide sequence context, ionic environment, temperature, metabolites in interaction with components of the replication apparatus, etc.). Whether bacteria are in the exponential or stationary phase of growth can affect intracellular metabolites and proton exchange rates which, in turn, may alter the proportion of tautomeric forms in nucleotides and misincorporation tendencies (Friedberg et al., 2006). The sequence context of the template nucleic acids (presence of repeated sequences that can induce misalignment mutagenesis or G-C vs. A-T rich regions in relation to relative nucleotide substrate abundances, etc.) may impel or attenuate mutability. Insertion elements may enhance mutation rates at neighboring sites in a bacterial genome (Miller and Day, 2004). Despite these influences, vesicular stomatitis virus (VSV) displayed comparable mutation rates in several host cells (Combe and Sanjuan, 2014) suggesting that there is a limited range of average error rates needed for a virus to maintain fitness (Chapters 5 and 9).

In addition to the general environmental and sequence context consequences for template-copying fidelity that may affect any genome type, mutation rates for DNA viruses will also be influenced by: (i) whether the DNA polymerase that catalyzes viral DNA synthesis includes or lacks a functional proofreading-repair activity. High copying fidelity is typical of DNA polymerases involved in cellular DNA replication (Bebenek and Ziuzia-Graczyk, 2018), and low copying fidelity is generally a feature of DNA polymerases involved in DNA repair (Friedberg et al., 2006; Ganai and Johansson, 2016). Thus, repair of lesions that by themselves might not be mutagenic may lead to the introduction of mutations during the error-prone repair process. (ii) Expression of proteins active in repair encoded in the viral genome, such as uracil-DNA glycosylase, DNA repair endonucleases, etc. (iii) The mechanism of viral DNA replication, particularly the occurrence of double-stranded versus single-stranded DNA in replicative intermediates. (iv) Intracellular site of replication and the availability of postreplicative DNA repair proteins (regarding both intracellular location and concentration) to the viral replication factories. Little is known of the spatial relationships and relative affinities of cellular and viral proteins and structures that may critically affect polymerase fidelity. Comparative measurements of mutation rates at specific genome sites of DNA viruses are needed, as a first step to define the cellular and biochemical influences on the fidelity of DNA virus genome replication.

General genetic variability affecting the entire virus genome should be distinguished from

localized variability at hot spots in a genome. Even the extremely complex human genome shows genetic instability at specific loci, some associated with genetic disease (Domingo et al., 2001; Alberts et al., 2002; Bushman, 2002). Genome size is a parameter pertinent to biological behavior, not only because it imposes a commensurate copying fidelity, but also because it affects the impact of genetic heterogeneity within infected organisms and upon the invasion of new hosts (Chapter 3).

Mutation frequencies measured by subjecting virus to a specific selective agent (e.g., mutants that escape the neutralizing activity of a monoclonal antibody or mutants that escape inhibition by a drug) span a broad range of values (10^{-3} to 10^{-8}) for DNA and RNA viruses (Smith and Inglis, 1987; Sarisky et al., 2000; Domingo et al., 2001) (Table 2.1). The technical details of any procedure used to calculate a mutation frequency should be carefully evaluated to translate its meaning to the genome level. Important variables are the efficacy of the antibody or drug (which will be concentration-dependent) or the possibility of phenotypic hiding-mixing in the escape mutants to be quantified (Holland et al., 1989; Valcarcel and Ortin, 1989). Unexpected low levels of escape mutants (that would imply $< 10^{-6}$ substitutions per nucleotide) for an RNA virus can mean either a general or site-specific high polymerase fidelity, a selective disadvantage of the genome that harbors the mutation or, when a phenotypic alteration is measured, the requirement of two or more mutations to produce the alteration. Conversely, a high mutation frequency for a DNA virus whose replication is catalyzed by a high-fidelity DNA polymerase may mean that either repair activities were not functional or that the mutant displayed a selective advantage and overgrew the wild type prior to the measurement of its frequency. Mathematical treatments that take into account reversion of a low fitness mutant and its competition with wild-type virus have been used to calculate mutation rates (Batschelet et al., 1976; Coffin, 1990).

Despite difficulties and limitations in the calculations, independent genetic and biochemical methods with different viruses support mutation rates for RNA viruses in the range of 10^{-3} to 10^{-5} substitutions per nucleotide copied [as representative articles and reviews see (Batschelet et al., 1976; Domingo et al., 1978, 2001; Steinhauer and Holland, 1986; Eigen and Biebricher, 1988; Varela-Echavarria et al., 1992; Ward and Flanegan, 1992; Mansky and Temin, 1995; Preston and Dougherty, 1996; Drake and Holland, 1999; Sanjuan et al., 2010; Bradwell et al., 2013)] (Table 2.1). A few early studies indicated unusual low mutation rates or frequencies for some RNA viruses. As discussed in some of the reviews listed above, there are technical reasons to suggest that such values were probably underestimates of the true average mutations rates or frequencies. Obviously, it cannot be excluded that some genomic sites or viruses under a given environment might be unusually refractory to introduce mutations, but most evidence supports the range of values listed in Table 2.1. The near million-fold higher mutation rates for RNA viruses than cellular DNA, whose biological implications were presciently anticipated by J. Holland and colleagues (Holland et al., 1982), have been confirmed. That is, for RNA viruses of genome length between 3 Kb and 32 Kb, an average of 0.1-1 mutation is introduced per template molecule copied in the replicating population. Unless most mutations impeded viral replication, a continuous input of mutant genomes is expected, as indeed found experimentally (Chapter 3).

High mutation rates for RNA genomes are also supported by measurements of template-copying fidelity by RNA polymerases, reverse transcriptases, and DNA polymerases devoid of 3–5′ proofreading exonuclease (or under conditions in which, such exonuclease is not functional) [(Steinhauer et al., 1992; Varela-Echavarria et al., 1992; Mansky and Temin, 1995; Domingo et al., 2001; Friedberg et al., 2002, 2006; Menéndez-Arias, 2002), and

references therein]. In vitro fidelity tests may be based on genetic or biochemical assays using homopolymeric or heteropolymeric template-primers. Measurements include the kinetics of incorporation of an incorrect versus the correct nucleotide directed by a specific position of a template or the capacity of a polymerase to elongate a mismatched template-primer 3′ end, [these and other assays have been reviewed (Menéndez-Arias, 2002)] (see also Section 2.6). Differences between related enzymes (i.e., AMV RT is more accurate than HIV-1 RT), and the fact that amino acid substitutions in the polymerases affect nucleotide discrimination, demonstrate that proofreading-repair activities together with the structure of the polymerase and replication complexes are determinants of template-copying fidelity.

2.5.1 Undesired consequences of the confusion between mutation rates and mutation frequencies

The term mutation rate if often used in a light manner in the literature of virus evolution, probably driven by nomenclature from classical population genetics. It is used to mean mutation frequency, rate of evolution, and sometimes to mean mutation rate in its real sense (as explained in Section 2.5). A particularly risky habit is to use mutation rate when what is measured is a mutation frequency. Some studies have claimed that they have a replication system devoid of selection, and therefore, the number of mutations counted corresponds to the true mutation rate of the system. This is incorrect. There is no replicative system devoid of selection because at least selection to maintain replication is in continuous operation. Furthermore, as taught by quasispecies dynamics (Stadler, 2016), supported by mutational waves in hepatitis C virus upon prolonged replication in a constant cellular environment (Moreno et al., 2017), the mutant spectrum per se is part of the environment. Since the mutant spectrum is constantly changing, so is the environment in which replication takes place. Unfortunately, some studies have proposed the existence of mutational cold spots (sites at which mutations as a biochemical event occur at a very low frequency) ignoring that negative selection might have eliminated newly arising mutations. These false conclusions imply the existence of genomic regions particularly suitable as drug or antibody targets because they cannot mutate. These are the types of studies and incorrect conclusions that keep perpetuating the problem of control of viral diseases by encouraging antiviral and vaccine designs doomed to failure (Chapters 8 and 9).

2.6 Evolutionary origins, evolvability, and consequences of high mutation rates: fidelity mutants

The amino acid substitutions in the core polymerase that affect fidelity can be located either close to or away from the active site of the enzyme. The change in fidelity can reach almost one order of magnitude, but virus viability is not compromised. Thus, error rates themselves can be subjected to selection, as supported by theoretical studies on evolvability (Earl and Deem, 2004). Early studies documented heterogeneity in the mutation rates among individual plaque isolates of influenza A virus (Suarez et al., 1992). It is not clear whether mutation rates of viruses have evolved to procure a balance between adaptability and genetic stability, or whether other selective constraints have imposed the observed values. It has been suggested that because of the generally deleterious nature of most mutations, the adaptive value of the high mutation rates for RNA viruses is debatable and that there might have been a trade-off between replication rate and copying fidelity (Elena and Sanjuan, 2005). Mutation rates would be a consequence of rapid RNA replication, and an increase in copying fidelity would come at a

cost, resulting in a lower replication rate. A connection between elongation and error rate has been suggested by results with some viral and cellular polymerases [review (Kunkel and Erie, 2005)]. In an early study with the poliovirus RdRp in vitro, an increase in the error frequency was observed when the pH and Mg^{2+} ion conditions were modified, and the decreased fidelity correlated with increased RNA elongation rate (Ward et al., 1988). A possible connection between elongation rate and copying fidelity cannot be ruled out, but current evidence points to template-copying fidelity as being the result of multiple factors, not necessarily linked to the rate of genome replication (Vignuzzi and Andino, 2010; Campagnola et al., 2015; Domingo and Perales, 2019).

There is ample support for an adaptive value of high mutation rates for RNA viruses, independently of their biochemical origins. A poliovirus mutant, whose RdRp displayed a three-to five-fold higher fidelity than the wild-type enzyme, replicated at a slightly lower rate than wild-type virus in cell culture but displayed a strong selective disadvantage regarding the invasion of the brain of susceptible mice (Pfeiffer and Kirkegaard, 2005; Vignuzzi et al., 2006). The impediment to cause neuropathology was due at least partly to the limited complexity of the mutant spectrum since its broadening through mutagenesis restored the capacity to produce neuropathology. These and other studies have provided evidence that mutant spectrum complexity, by virtue of its impact on fitness, can be a virulence determinant. The work by M. Vignuzzi, J. Pfeiffer, R. Andino, C. Cameron, K. Kirkegaard, and their colleagues on poliovirus fidelity mutants opened a much-needed branch of research in virus evolution and quasispecies implications. As proof of this statement, the field of fidelity mutants is rapidly expanding (Bordería et al., 2016), and references to the information they provide will be made in several chapters.

Theoretical models and experimental observations suggest that mechanisms for error correction had to evolve to maintain functionality of increasingly complex genomes (Swetina and Schuster, 1982; Eigen and Biebricher, 1988; Domingo et al., 2001; Eigen, 2002, 2013) (here complexity means genome size, provided no redundant information is encoded). The coronaviruses have the largest genomes among the known RNA viruses, with 30–33 kb. This is about 10-fold more genetic information than encoded in the simple RNA bacteriophages, such as MS2 or Qβ. Coronaviruses are replicated by complex RNA-dependent RNA polymerases, which include a domain that corresponds to a 3′–5′ exonuclease, proofreading-repair activity. The protein displays exonuclease activity in vitro, and its inactivation affects viral RNA synthesis (Minskaia et al., 2006), and results in increases of about 15-fold in the average mutation frequency (Eckerle et al., 2007, 2010). A coronavirus mutant devoid of this repair function is more susceptible to lethal mutagenesis than the corresponding, nonmutated virus (Smith et al., 2013; Smith and Denison, 2013), as expected from a connection between replication accuracy and proximity to an error threshold for the maintenance of genetic information (Chapter 9). Thus, it is likely that a proofreading activity evolved (or was captured from a cellular counterpart) in RNA genomes, whose genomic complexity was in the limit compatible with the fidelity achievable by standard RNA replicases. It would be interesting to discover new RNA viruses with a single RNA molecule longer than 30 Kb as a genome to analyze whether they have evolved more accurate core polymerases or exhibit a proofreading-repair function during replication. Toward the other end of the RNA size scale, viroid RNAs display a mutation rate higher than (or close to the highest) recorded for RNA viruses, consistent with the correlation between genome size and template-copying accuracy (Gago et al., 2009).

Studies with bacteria have identified some of the factors that successively increase copying fidelity. It has been estimated that during *E. coli* DNA

replication the error rate would be 10^{-1} to 10^{-2} mutations per nucleotide copied if accuracy relied only upon the strength of interactions provided by base pairing (Section 2.2). The error rate would decrease to 10^{-5} to 10^{-6} with base selection and proofreading-repair, to about 10^{-7} with the contribution of additional proteins present in the replication complex, and to about 10^{-10} misincorporations per nucleotide with the participation of postreplicative mismatch correction mechanisms (Naegeli, 1997; Kunkel and Erie, 2005; Friedberg et al., 2006). Reduction of bacterial genome size results in the increase of mutation frequency (Nishimura et al., 2017). The error rate of the bacteriophage ϕ29 DNA polymerase is about 10^{-6} without the proofreading exonuclease activity, and it decreases to 10^{-8} with the correcting activity [(de Vega et al., 2010) and references therein]. Postreplicative repair pathways act on double-stranded DNA, but not (or very inefficiently) on RNA or DNA-RNA hybrids. Therefore, the known postreplicative repair systems that operate in cellular DNA do not make a significant contribution to error correction in RNA viruses (Steinhauer et al., 1992).

The importance of copying fidelity for complex genomes is reflected in the fact that more than 100 proteins are directly or indirectly involved in the repair of the human genome (Liu et al., 2017). Elevated mutation rates in the range of those operating for RNA viruses would be lethal for large DNA genomes. Localized genetic modification occurs physiologically in processes, such as somatic hypermutation and class-switch recombination in B cells of the germinal centers, as mechanisms of diversification of immunoglobulin genes (Upton et al., 2011; Methot et al., 2017). Chromosomal instability has long been associated with cancer (Gatenby and Frieden, 2004; Stratton et al., 2009). Surveys have been (and are currently being) used to identify genes associated with chromosomal instability and their role in aging and disease (Aguilera and Garcia-Muse, 2013; Vijg et al., 2017) (see also Chapter 10). While uncontrolled high mutability is deleterious for differentiated cellular organisms, it constitutes a *modus vivendi* for a great majority of viruses.

Despite its attractiveness, definitive proof of the hypothesis of a direct relationship between error rate and limited genome complexity will require additional functional and biochemical studies. Exceptions to the absence of repair activities in simple genetic elements have been described. A satellite RNA of the plant virus turnip crinkle carmovirus evolved a 3′-end RNA repair mechanism. It implicates the synthesis of short oligoribonucleotides by the viral replicase using the 3′-end of the viral genome as a template. The mechanism consists probably of template-independent priming at the 3′-end of the damaged RNA to generate wild type, negative strand, and satellite RNA (Nagy and Simon, 1997). A reversible, NTP-dependent excision of the 3′ residue of the nascent nucleic acid product has been described in some retroviruses and hepatitis C virus (Meyer et al., 1998; Jin et al., 2013). This activity is important for drug resistance, and it may also modulate the overall fidelity of some polymerases. It cannot be excluded that some type of point mutation correction may operate in RNA genetic elements of less than 30 Kb. Such putative mechanisms may even diminish mutation rates that would otherwise be prohibitively deleterious, and they do not overshadow high mutation rates as a feature of RNA and some DNA genomes (Table 2.1).

Limited copying fidelity in the absence of proofreading-correction mechanisms can be regarded as an unavoidable consequence of the molecular mechanisms involved in template copying by viral polymerases. Most nucleic acid polymerases share a structure that resembles a right hand, with fingers, palm, and thumb domains (Fig. 2.6A). Three-dimensional structures of viral RdRps and RTs indicate that interactions between the incoming nucleotide or residues of the template-primer with amino acids of the polymerase must permit displacement of the growing polymerase chain along

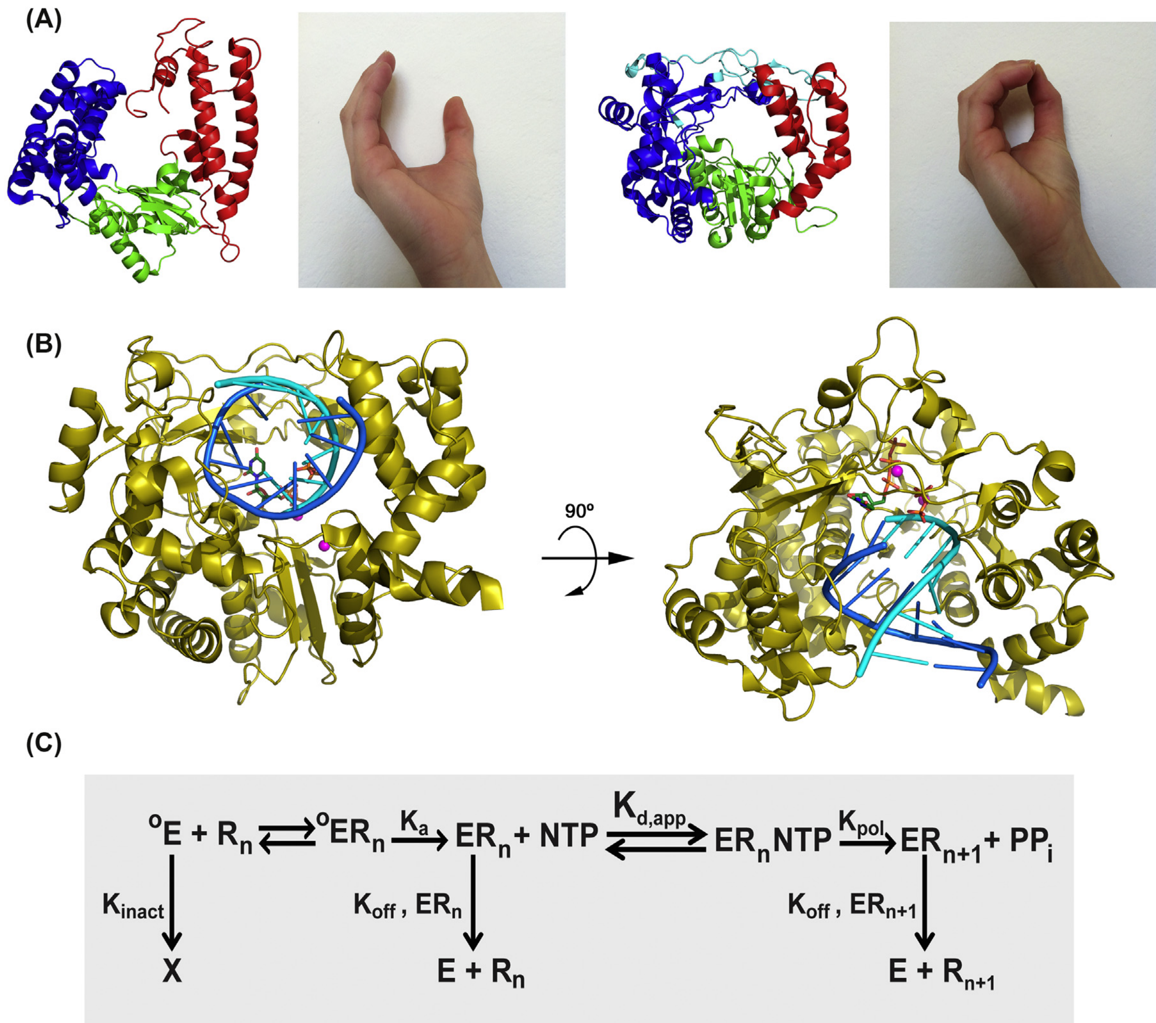

FIGURE 2.6 Polymerization by viral polymerases. (A) The structure of nucleic acid polymerases resembles that of a right hand. This was first evidenced with the structure of the Klenow fragment of the *E. coli* DNA polymerase I by T.A. Steitz and colleagues in 1985. On the left, the Klenow fragment is represented with colored fingers, palm and thumb domains, next to an open right hand. The structure on the right is that of PV RdRp, next to a closed right hand. Courtesy of N. Verdaguer and L. Vives-Adrián (the hand is that of L. Vives-Adrián). (B) The structure of the ternary complex between FMDV 3D, an RNA molecule, and UTP as the substrate (PDB id. 2E9Z). The left panel is a front view of the complex, depicting the polymerase chain as a yellow ribbon, the RNA in dark blue (template) and cyan (primer). The incoming UTP and the pyrophosphate product are shown in atom type, and 2 Mg^{2+} ions as magenta balls. The right panel is the same complex in a top-down orientation. (Figure courtesy of C. Ferrer-Orta and N. Verdaguer). (C) Scheme of the minimum number of steps involved in nucleotide incorporation. The first step consists of the binding of polymerase ^{0}E to the template-primer R_n (elongated up to nucleotide n) to form a complex $^{0}ER_n$. Formation of the activated complex ER_n is governed by the rate of constant $k_{assembly}$ (k_a). The activated ER_n complex binds a nucleotide NTP with an apparent binding affinity given by $K_{d,app}$ to form the ER_nNTP complex. Catalysis to covalently incorporate the NTP to the growing primer chain to yield ER_{n+1} and pyrophosphate (PP_i) is governed by the rate constant k_{pol}. Other constants depicted in the scheme are the inactivation rate constant (k_{inact}) of ^{0}E, and dissociation of E from RNA (k_{off}, ER_n and k_{off}, ER_{n+1}). *Based on Arias, A., Arnold, J.J., Sierra, M., Smidansky, E.D., Domingo, E., et al., 2008. 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 82, 12346—12355, and previous studies with PV polymerase 3D by C. E. Cameron and his colleagues.*

the channel located at the palm domain of the polymerase (Steitz, 1999; Ferrer-Orta et al., 2006; Wu and Gong, 2018) (Fig. 2.6B). The polymerase of classical swine fever virus and other pestiviruses, such as bovine viral diarrhea virus includes an N-terminal extra-domain of about 100 amino acids; its interaction with the palm domain is important for template copying fidelity (Liu et al., 2018). If interactions around the catalytic site to ensure the correct nucleotide incorporation were so strong as to preclude misincorporations, the movement of the growing polynucleotide chain would be hampered. Again, this compromise suggests a match between biochemical and evolutionary needs.

The orientation of the triphosphate moiety of the incoming nucleotide substrate is important for nucleotide incorporation (Menéndez-Arias, 2002; Graci and Cameron, 2004; Ferrer-Orta et al., 2009). One of the several steps involved in nucleotide incorporation is the formation of a ternary complex (polymerase with template-primer and the incoming nucleotide) that undergoes a conformational change (reorientation of the divalent ion-complexed triphosphate moiety of the incoming nucleotide). This conformational change activates the complex for phosphoryl transfer, to link the nucleoside-monophosphate to the 3′-terminus of the primer (or growing chain). Steps involved in the nucleotide incorporation are represented in Fig. 2.6C. Both the conformational change and the relative rate of phosphoryl transfer for an incorrect nucleotide versus the correct nucleotide influence the error rate at each site of the growing chain. Critical kinetic constants in Fig. 2.6C that are determined experimentally to quantify relative nucleotide incorporations and misincorporations are $K_{d,app}$ (expressed as μM), k_{pol} (expressed as s^{-1}), and the ratio $k_{pol}/K_{d,app}$ ($\mu M^{-1} s^{-1}$) termed the catalytic efficiency. The ratio of $k_{pol}/K_{d,app}$ for the incorporation of an incorrect nucleotide to $k_{pol}/K_{d,app}$ for a correct nucleotide gives the frequency of that particular misincorporation, and an assessment of polymerase fidelity [(Castro et al., 2005) and references therein]. Modifications of polymerase residues by site-directed mutagenesis, combined with comparisons of the relevant structures, have identified critical amino acid residues involved in template-copying fidelity.

High-fidelity mutants are frequently obtained by selecting viruses resistant to mutagenic nucleotide analogs such as the antiviral agent ribavirin (Beaucourt and Vignuzzi, 2014). Limited incorporation of a deleterious nucleotide can be attained either through specific discrimination against the analog (ribavirin or other) or through a general decrease of all types of misincorporations, that is, a high-fidelity phenotype. Structural modifications of viral polymerases that lead to high fidelity have inspired the design of mutant viral polymerases displaying either an increase or decrease of copying fidelity achieved through a single amino acid substitution (Wainberg et al., 1996; Menéndez-Arias, 2002; Mansky et al., 2003; Pfeiffer and Kirkegaard, 2003; Arnold et al., 2005; Domingo, 2005; Vignuzzi et al., 2006; Coffey et al., 2011; Gnadig et al., 2012; Meng and Kwang, 2014; Rozen-Gagnon et al., 2014; Borderia et al., 2016). The capacity of the virus to evolve at higher or lower rates than their ancestors is achievable through modest numbers of mutations (limited movements in sequence space, Chapter 3), again emphasizing the evolvability of mutation rates.

The rates of mutation and recombination need not be independent. M. Vignuzzi and colleagues have shown that a mutator Sindbis virus displays a higher recombination rate and enhanced production of DI particles than the wild type virus (Poirier et al., 2015). A connection between mutation and recombination rates strengthens the evolutionary consequences of the modifications of template copying fidelity that can be achieved through a single amino acid substitution.

2.7 Hypermutagenesis and its application to generating a variation: APOBEC and ADAR activities

Some viral genomes either isolated from biological samples or evolved in cell culture show biased mutation types (e.g., monotonous G → A or C → U substitutions in the same genome), generally at frequencies of around 10^{-2} substitutions per nucleotide (10- to 1000-fold higher than standard mutation rates and frequencies) (Table 2.1). Biased hypermutation was first observed in some defective interfering (DI) RNAs of vesicular stomatitis virus (VSV) (Holland et al., 1982), and in variant forms of measles virus, associated with postmeasles neurological disease, such as subacute sclerosing panencephalitis (Cattaneo and Billeter, 1992). Hypermutation is mainly due to the activity of cellular deaminases, such as the apolipoprotein B mRNA and the editing complex (APOBEC), or the adenosine deaminase acting on double-stranded RNA (ADAR) families, that are involved in cellular editing and regulatory functions (Sheehy et al., 2002; Santiago and Greene, 2008; Nishikura, 2010; Stavrou et al., 2014; Pfaller et al., 2018; Venkatesan et al., 2018). In the event of a viral infection, such cellular functions can become part of an innate defense mechanism against the invading virus. Viral proteins (i.e., Vif in HIV-1) bind some APOBEC proteins, thus inhibiting mutagenesis and permitting virus survival (Sheehy et al., 2002). In oncoretroviruses, retroviruses, and hepatitis B virus (HBV), the APOBEC-3 cytidine deaminase acts on single-stranded DNA and results mainly in G → A and C → U hypermutation, that may affect 40%−100% of the G residues. The preferred sequence context for G hypermutation in HIV-1 observed in vivo is GpA > GpG > GpT ≈ GpC. The specific dinucleotide context of the hypermutated sites provides a means to distinguish genomes that have undergone hypermutation by cellular activities from those that are heavily mutated by other mechanisms, such as the action of mutagenic agents (Chapter 9).

APOBEC 3 proteins play a role in cancer through cytidine deaminase mutagenesis and generation of double-strand breaks in chromosomal DNA (Wang et al., 2016). APOBEC 3 levels in the cell may be regulated by cellular and viral proteins, for example, human papillomavirus (HPV) oncoprotein E that stabilizes APOBEC 3A in human keratinocytes that may promote cervical cancer (Westrich et al., 2018).

The ADAR-associated hypermutation was identified in negative-strand RNA riboviruses and results mainly in A → G and U → C hypermutation. It is originated by A → I (inosine) modifications in double-stranded viral RNA catalyzed by ADAR-1 L, one of more than 100 proteins inducible by type I IFN (Maas et al., 2003). Inosine can be recognized as G by the replication machinery (Valente and Nishikura, 2005), although it can form wobble base pairs also with A and U (Fig. 2.2). Hypermutation can contribute to genetic variation of viruses (Hirose et al., 2018).

There are additional mechanisms of hypermutagenesis. Higher than average mutation frequencies can occur as a result of replication in the presence of biased concentrations of the standard nucleotide substrates; this has been applied to the in vitro generation of genes mutated at frequencies of 10^{-1} to 10^{-2} (mutagenic PCR), as a powerful tool to study sequence-function relationships and functional robustness of nucleic acids and proteins (Meyerhans and Vartanian, 1999). Error-prone PCR has been used in experiments of in vitro evolution of nucleic acid enzymes to generate heterogeneous collections of nucleic acid sequences to select for molecules capable of catalyzing specific reactions (Joyce, 2004) (Chapter 1).

2.8 Error-prone replication and maintenance of genetic information: instability of laboratory viral constructs

High mutation rates have practical implications in laboratory studies on the behavior of

virus mutants obtained by molecular cloning of a biological sample, or constructed by site-directed mutagenesis. A transition mutation that causes a strong fitness decrease but that still allows residual RNA genome replication will most likely revert following infection or transfection of cells with the mutant construct and subsequent viral replication. Double or triple mutants (preferentially including transversions) should be engineered (when possible according to the genetic code) to study the behavior of a viral mutant with an amino acid replacement of interest that may produce a fitness decrease. As an example, a C → U transition found in an open reading frame of an RNA virus may convert a Pro into a Ser (CCG → UCG). Since Ser will revert to Pro through a U → C transition in the triplet (a common type of misincorporation by most polymerases), Ser should be engineered to be encoded by AGU; in the course of replication, reversion to Pro would require at least two transversions since the codons for Pro are CCU, CCC, CCA, or CCG. Thus, if effects derived from the difference in the primary sequence of the RNA or codon bias do not intervene in the behavior of the viral genome, codons with a high genetic barrier to reversion should be engineered for studies involving viral replication.

In general, deletions revert at a much lower frequency than point mutations, and when appropriate for the question under study, a deletion should be introduced within the gene of interest to probe gene function in reversegenetics studies. High mutation rates also imply that infection or transfection with debilitated mutant viruses may result in progeny with sequences that differ from the input. V. I. Agol and colleagues have coined the term *quasi-infectious* to refer to mutant viruses that are capable of yielding progeny, but the progeny differs from the initial genome (pseudorevertants) (Gmyl et al., 1993; Agol and Gmyl, 2018). The difference between the input mutant and the rescued progeny virus will depend on the type of genetic lesion in the input virus and its consequences for virus multiplication. A single point mutation that decreases replication is likely to evolve to yield a true revertant (return to the original sequence) upon replication. If the same reversion depends on two or more mutations, a true revertant will require extended replication for exploration of sequence space (Chapter 3), and selection of compensatory mutations elsewhere in the genome (sometimes referred to as second site revertants) becomes an alternative for fitness gain. The term compensatory applies to mutations that compensate for the deleteriousness of other mutations. A typical example is a mutation that decreases the stability of a stem in an RNA stem-loop that functions as a *cis*-acting element. A compensatory mutation restores a stable stem needed for the activity. Transfection of cells by an engineered virus with some preselected genetic modification (produced either from cDNA copies of a viral genome or by chemical synthesis) may yield progeny genomes, which differ from the parent. If a substantial loss of replicative capacity is produced by a drastic genetic change (an *indel*, loss of a stem-loop structure, etc.) selection of a true revertant becomes extremely unlikely. The compensatory generation of alternative structures (or constellations of point mutations) that restores replication (partially or completely) becomes an interesting and informative possibility.

Procedures to copy an entire viral RNA genome into a cDNA for reverse genetics studies are now available (Fan and Di Bisceglie, 2010). If for technical reasons an infectious cDNA clone is constructed from several molecules, which were copied from different genomes present in the mutant spectrum, the ligation product may be transcribed into an RNA, which is not infectious. This is because, while some constellations of mutations may be compatible with infectivity, others may not, or may allow limited, suboptimal replication, thus favoring the selection of additional mutations or reversions. The same applies to a synthetic genome based on one of the multiple genomic sequences from a viral

isolate. Individual mutations may be detrimental either per se, or by the combined presence of other mutations. The joint effect of different mutations in the same genome is often referred to as epistasis. Mutations that reinforce each other with regard to a viral function are said to produce positive epistasis, and those that interfere with each other produce negative epistasis (also mentioned in Section 2.3). Epistasis in RNA viruses may be blurred by the weight of mutant spectra in determining viral behavior through intergenomic interactions (Chapter 3).

An interesting contrast that recapitulates concepts given in Sections 2.5 and 2.6 is the effect of an active proofreading-repair activity in maintaining the infectivity of a viral genome upon its extended replication in vitro (in a test tube, in the absence of cellular extracts). The 19,285 bp bacteriophage ϕ29 DNA can be amplified at least 4000-fold without detectable loss of infectivity due to the fidelity of ϕ29 DNA polymerase conferred by a 3′−5′ proofreading-repair exonuclease activity (Bernad et al., 1989). Engineered ϕ29 DNA polymerases provide a powerful amplification tool in genomics (de Vega et al., 2010). In contrast, the 4220 nucleotides long Qβ RNA rapidly loses its infectivity when replicated by Qβ replicase in vitro due to the accumulation of mutations and deletions in the viral RNA (Mills et al., 1967; Sabo et al., 1977). The error-prone Qβ replicase is not adequate to amplify infectious viral RNA, but it was at the origin of the quasispecies concept to be discussed in Chapter 3. Mutagenic DNA polymerases (generally those involved in DNA repair) are an alternative to mutagenic PCR (Section 2.7) to produce randomly mutated collections of nucleic acid molecules (Forloni et al., 2018).

2.9 Recombination in DNA and RNA viruses

Recombination is the formation of a new genome by covalent linkage of genetic material from two or more different parental genomes (Fig. 2.7). Recombination can also involve different sites of the same genome to yield insertions or deletions, such as in the formation of defective interfering (DI) genomes. It is a widespread mechanism of genetic variation in all biological systems, and in cells, it underlies critical physiological and developmental processes (splicing, generation of diversity in

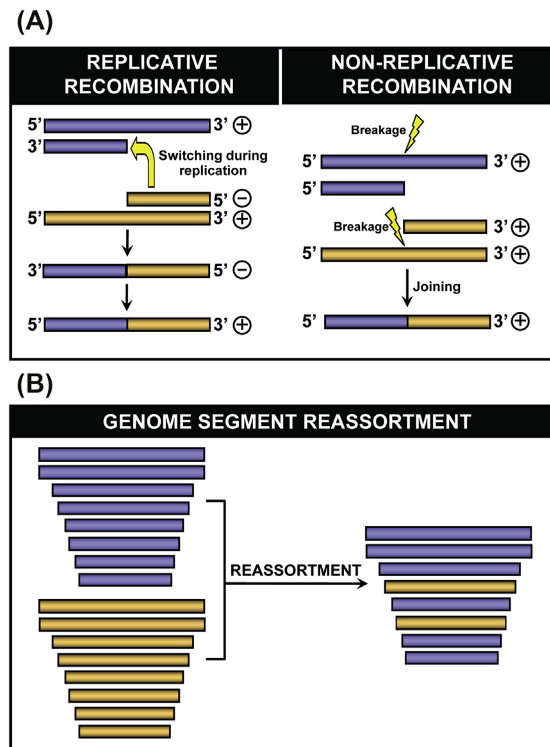

FIGURE 2.7 RNA recombination and segment reassortment. (A) Scheme of replicative and nonreplicative RNA recombination. RNA polarity is indicated by +, − symbols. Replicative recombination is displayed as the result of template switching during minus-strand RNA synthesis. Nonreplicative recombination is depicted as the outcome of breakage and ligation (joining) of fragments of plus-strand RNA. (B) An example of genome segment reassortment with the formation of a new segment constellation in which six genomic segments originate from one parent (*blue*) and two from the other (*gold*). Influenza virus is the best-known example (see text).

immunoglobulin genes and T cell receptors, transposition events, phase variation in bacteria, repair pathways that promote postreplicative error correction, etc.). Cellular DNA recombination relates to replication, repair, and completion of DNA replication, operations that involve multiple proteins displaying a variety of activities (Smith and Jones, 1999; Alberts et al., 2002; Nimonkar and Boehmer, 2003; Friedberg et al., 2006).

Recombination occurs both with DNA and RNA viruses, often with the participation of the virus replication machinery. Several types of recombination have been distinguished in viruses: homologous versus nonhomologous recombination, according to the extent of nucleotide sequence identity around the recombination (crossover) site, and replicative versus nonreplicative recombination, according to the requirement of viral genome replication for recombination to occur (Kirkegaard and Baltimore, 1986; King, 1988; Lai, 1992; Nagy and Simon, 1997; Plyusnin et al., 2002; Boehmer and Nimonkar, 2003; Gmyl et al., 2003; Chetverin et al., 2005; Agol, 2010; Simmonds, 2010; Bujarski, 2013; Perez-Losada et al., 2015; Agol and Gmyl, 2018; Bentley and Evans, 2018).

As in the case of cells, homologous recombination in double-stranded DNA viruses is intimately connected with DNA replication and repair. It implicates multiple viral gene products (DNA polymerase, single-stranded DNA-binding proteins, processivity factors, helicase-primase, eukaryotic topoisomerase I, etc.), and a succession of protein-catalyzed steps (Czarnecki and Traktman, 2017). In the copy choice (or template switching) mechanism, the nascent DNA switches from one template molecule to another, resulting in the synthesis of recombinant, daughter DNAs. In its basic form, recombination by breakage and rejoining starts with the introduction of a nick at one of the strands of each parental DNA, strand invasion of one parental DNA by the other, branch migration, ligation at the nicks (linking DNA strands from the two parents), and further isomerization and cleavage reactions. DNA recombination is responsible for the endonuclease-mediated isomerization of herpesvirus genomes [four isomers defined by the orientation of the long (L) and short (S) regions of the viral genome]. During the late phase of herpes simplex virus-1 replication, the frequency of recombination has been estimated at 0.6% per Kb of the genome (Boehmer and Nimonkar, 2003). Integration or excision of proviral DNA or temperate bacteriophage DNA, are examples of site-specific recombination that involves specific enzyme activities (i.e., retroviral integrases), and requires a short stretch of nucleotide sequence identity.

The copy choice mechanism of homologous RNA recombination is also associated with genome replication. An RNA polymerase molecule with its nascent RNA product jumps into the corresponding position of another template molecule, to complete synthesis of the RNA product (Fig. 2.7). Given the large numbers of viral genomes often present in replication complexes (also termed replication factories) in each infected cell, it is not surprising that this mechanism may give rise to frequent recombinant progeny genomes. Formation of mosaic genomes has long been recognized as an essential feature of the genetics of some retroviruses and plant RNA viruses. For HIV-1 and some plant RNA viruses recombination frequencies have been estimated at 2%–10% of progeny per 100 nucleotides; for picornaviruses and coronaviruses the number of recombinants amounts to 10%–20% of the progeny (King, 1988; Lai, 1992; Nagy and Simon, 1997; Levy et al., 2004; Urbanowicz et al., 2005; Sztuba-Solinska et al., 2011). Using a phylogenetic approach the average recombination rate of HIV-1 in vivo was estimated in 1.4×10^{-4} recombination events/site/generation, which is about five-fold greater than the average point mutation rate (Shriner et al., 2004). A ten-fold lower value of 1.4×10^{-5} recombination events/site/

generation was estimated from the changes in the genetic composition of HIV-1 within single patients (Neher and Leitner, 2010). Recombination is required for HIV-1 replication and genome integrity (Rawson et al., 2018). In negative-strand RNA viruses recombination may be inefficient or absent, but some of them can display homologous recombination (Plyusnin et al., 2002), and a high rate of generation of DI RNAs and other types of defective genomes (Roux et al., 1991; Rezelj et al., 2018).

Recombination frequency may be altered by environmental factors that affect viral replication. A decrease of intracellular nucleotide levels as a result of treatment of cells with hydroxyurea may favor template switching reflected in an increase of intra and intermolecular recombination (Pfeiffer et al., 1999; Svarovskaia et al., 2000). Homologous RNA recombination can also be influenced by amino acid substitutions in the polymerase, the primary sequence in the RNA (i.e., high frequency of template switching in AU-rich regions), the sequence identity between the nascent strand and acceptor template, and secondary structures at or around the crossover sites, among other influences (Nagy and Simon, 1997; Alejska et al., 2005; Agol, 2010; Agol and Gmyl, 2018). Since recombination necessitates coinfections of the same cell by at least two parental genomes, the persistence of a viral genome in a cell increases the likelihood of sequential coinfections, unless some reinfection or superinfection exclusion mechanism operates [(Webster et al., 2013) and references therein]. Without such restrictions, persistently infected cells may be an environment with a higher probability of recombination than transiently infected cells, assuming comparable genome loads at the sites of replication.

Nonreplicative recombination does not require replication of the viral genome, and has been described upon cotransfection of cells with viral RNA fragments that could not replicate by themselves (Gmyl et al., 2003; Gallei et al., 2004; Agol, 2010; Agol and Gmyl, 2018; Bentley and Evans, 2018). It appears to be a promiscuous event with a required 3′-phosphate in the 5′ partner RNA and a 5′-hydroxyl residue in the 3′ partner RNA mediated by cellular components whose mechanisms of activity are not understood.

2.9.1 Molecular occurrence versus observed recombination

The emerging picture is that the frequency of recombination varies among viruses, and that as new tools for genome analyses have become available, recombination has been detected in an increasing number of viruses. Recognition of recombination in a viral system is facilitated when a cell culture system is available. Controlled infection of cells with genetically marked parental viruses has been essential to estimate recombination frequencies, and to distinguish true recombination from mutation-reversion events that may mimic the formation of recombinants. As with the concept of high genetic variation in RNA viruses, recombination has often gone from being considered marginal to prominent and relevant; HCV is a typical example (Galli and Bukh, 2014), with presently at least one chimera established as a circulating recombinant form in the field.

Viral replicative machineries may be endowed with features that influence the occurrence of recombination. One such feature is processivity of the viral polymerase (capacity of continued copying of the same template molecule). Genome detachment of the polymerase complex from one genome to bind either to a different genome or to a distant site of the same genome is part of the standard replicative cycle of viruses, such as retroviruses and coronaviruses. Reverse transcriptase participates in strand transfer during DNA synthesis, and coronavirus polymerase switches from one template site to another during discontinuous RNA synthesis. It may be significant that they belong

to viral families displaying high recombination frequencies (Makino et al., 1986). Thus, here again, we encounter a "molecular instruction" that evolved as an essential feature of viral genome replication, and that can be exploited to generate variation, and permit new genomic forms to undergo the scrutiny of selection (Neher and Leitner, 2010) (Chapter 3).

Recombination is diagnosed by discordant positions of different genes or genomic regions in phylogenetic trees, as a result of the transfer of part of a viral genome from representatives of one lineage to representatives of another lineage (Chapter 7). A commonly used procedure measures similarity values between sequences using a sliding-window scanning method. The recombination crossover point (where the two parental sequences meet) is identified by the point (or region) where the similarity plot crosses from one sequence into another (Salemi and Vandamme, 2004; Martin et al., 2005; Kosakovsky Pond et al., 2006, Perez-Losada et al., 2015). Crossover points along a viral genome are not distributed at random, either because polymerase detachment from the template is sequence-dependent or because many of the recombination events do not lead to viable progeny. Absence of recombinant viability introduces a parallel with the distinction between mutation rate and mutation frequency (Section 2.5); that is, a difference between what does occur at the biochemical level during replication and what is subsequently observed upon analysis of the replication products. Newly arising recombinants may be subjected to negative selection, and only a viable subset might be detectable in the progeny virus (King, 1988; Lai, 1992). An elegant study by D. J. Evans and colleagues documented "imprecise" enterovirus recombinant intermediates that were lost upon serial virus passage (Lowry et al., 2014). Recombination is viewed as a biphasic process consisting of initial imprecise events followed by a stage of resolution in favor of fit recombinants.

The distinction between generation and resolution events that applies both to mutants and recombinants has yet another implication for RNA virus genetics. Some mutants or recombinants that in isolation do not exhibit sufficient replicative fitness to acquire dominance in a population may nevertheless persist as minority genomes. They may display low-level replication or be maintained by complementation by partner genomes (as in the case of two FMDV genome segments that are described in Section 2.11). As minority genomes, they may engage in modulatory activities (Chapter 3).

2.10 Genome segment reassortment

In viruses whose genomes are composed of two or more RNA or DNA segments, genome segment reassortment consists in the formation of new constellations of viral genomic segments from two or more parental genomes (McDonald et al., 2016) (Fig. 2.7). Reassortment can produce new phenotypic traits. It is the main mechanism of antigenic shift of influenza A viruses—often associated with new influenza pandemics (Webster et al., 1992; Morse, 1994; Gibbs et al., 1995; Domingo et al., 2001)—as opposed to antigenic drift, which is mediated by amino acid substitutions in the surface proteins hemagglutinin and neuraminidase (Barbezange et al., 2018). Reassortments occur among the 9–12 double-stranded RNA segments of the widespread *Reoviridae* family (Tanaka et al., 2012). Fitness differences among all possible segment combinations (2^n, for two types of coinfecting particles with n genome segments) determine the types of genome segment groupings that dominate subsequent rounds of infection. In the laboratory, analysis of reassortant viruses has been applied to map a viral function into one segment or a combination of segments.

Genomic segments can be encapsidated either into a single virus particle (as in Orthomyxoviruses or Arenaviruses) or into separate particles

(as in multipartite plant viruses). A multipartite virus can have either RNA or DNA as genetic material. The plant Nanoviruses have 6-8 molecules of single-stranded circular DNA of about 900−1000 nucleotides, and each segment encodes a single protein. In the case of the nanovirus Faba bean necrotic stunt virus, its eight segments vary in frequency in a host-dependent manner (Sicard et al., 2013). This observation led the authors to propose a "set-point genome formula," which may reflect the control of segment (gene) copy number that may provide some still unrecognized benefit to the multipartite phenotype (see next Section 2.11). In principle, replication of multipartite viruses requires that each cell be coinfected by at least one of each type of particle harboring a different genome type, which in fact represents a remarkable cost for replicative efficiency. The fact that unsegmented and segmented RNA viruses are well represented in our biosphere suggests that neither of the two organizations confers a definitive and general advantage for long-term survival.

2.11 Transition toward viral genome segmentation: implications for general evolution

The origin of viral genome segmentation is a debated issue, although there is general agreement that it may confer adaptive flexibility to viruses. Most proposals have been based on theoretical studies. Segmentation has been viewed as a form of sex that facilitates genomic exchanges to counteract the effect of deleterious mutations (Chao, 1988; Szathmary, 1992). An alternative, not mutually exclusive model is that segmentation confers an advantage because replication of shorter RNA molecules is completed earlier than the unsegmented counterparts (Nee, 1987). Yet another possibility is that the lifestyle of a virus (in particular, the particle yield in connection with the number of surrounding susceptible cells), shaped over long evolutionary periods, may favor segmentation over intactness of a genetic message or vice versa.

An experimental system of genome segmentation is available with the picornavirus foot-and-mouth disease virus (FMDV). Its single-stranded RNA genome underwent a modification akin to genome segmentation when the standard virus was subjected to 260 passages in BHK-21 cells at high multiplicity of infection. The experiment was originally intended to investigate the limits of fitness gain following prolonged multiplication in a defined environment, in this case, BHK-21 cells in culture. The starting FMDV had not been well adapted to the BHK-21 cell culture environment since it derived from a diseased swine during a disease outbreak, and it was minimally propagated in BHK-21 cells to obtain a biological clone by plaque isolation. Upon extensive replication of the clone in BHK-21 cells, the virus evolved toward a bipartite genome (García-Arriaza et al., 2004). Each of the two pieces of RNA that composed the bipartite (or segmented) genome version contained in-frame deletions affecting *trans*-acting proteins (Fig. 2.8). Each segment in isolation could not infect cells productively, but, when present together, they were infectious by complementation, and killed cells in the absence of standard FMDV. A low multiplicity of infection rapidly selected the full-length genome as a result of recombination of the two parental, defective segments (García-Arriaza et al., 2004). The particles containing the shortened RNA were thermally more stable than the standard particles (Ojosnegros et al., 2011), but this difference did not explain the initial trigger of the segmentation event. The solution to this question came with the demonstration that the transition toward genome segmentation was possible because of an extensive exploration of the mutational sequence space by the standard virus. Indeed, the mutations that accumulated during serial passages enhanced the fitness of the segmented

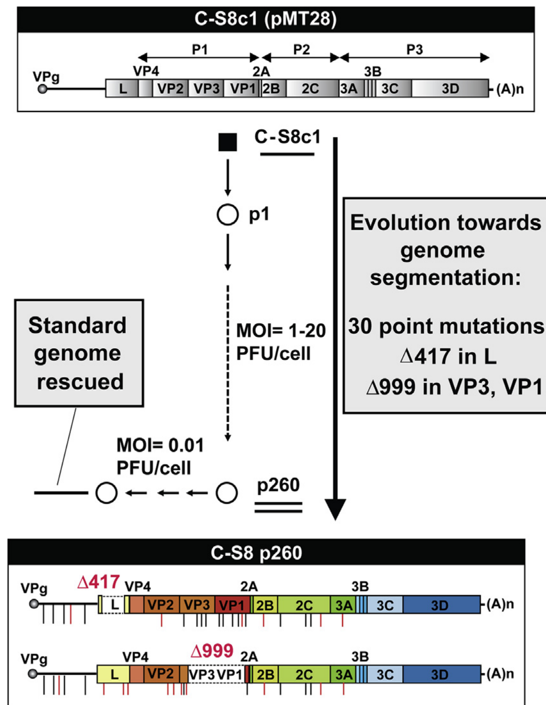

FIGURE 2.8 Evolution toward RNA genome segmentation in the laboratory. The monopartite, standard FMDV genome (clone C-S8c1 or pMT28, top) was subjected to 260 passages in BHK-21 cells. The resulting population p260 lacked detectable standard genome that could be rescued by low MOI passages. The evolved C-S8p260 accumulated 30 point mutations (depicted as *vertical lines* on the genome at the bottom) and consisted in two segments that were infections by complementation: Δ417, that lacked most of the L protease-coding region, and Δ999 that lacked most of the capsid proteins VP3, VP1-coding region. See text for further details and references.

genome version to a much higher extent than the fitness of the standard genome (Moreno et al., 2014) (Fig. 2.8). Thus, gradual evolution (drift in sequence space) was a requirement for the major transition toward segmentation, thus adding reassortment to mutation and recombination as potential mechanisms of genetic variation of this laboratory-adapted picornavirus. Cooperation and complementation are discussed in Section 3.8 of Chapter 3.

It should be noted that segmented forms of RNA viruses have been engineered, but little is known of the incurred fitness cost; it is relevant to have shown that segmentation was possible upon unperturbed replication of a virus with an unsegmented RNA genome. The experimental result suggests that in evolution there is no unsurmountable barrier that allows the conversion between intact and split forms of the same genome, reflecting remarkable genome flexibility that will be emphasized in Chapter 7 in the context of the relevance of virus variation in the emergence of viral pathogens.

2.12 Mutation, recombination, and reassortment as individual and combined evolutionary forces

Mutation, recombination, and segment reassortment contribute to the evolution of most DNA and RNA viruses. Sometimes one form of genetic change appears to be more prominent than another, and sometimes the concerted action of recombination or reassortment with the mutation is apparent [i.e., antigenic drift in influenza virus, following the origin of a new antigenic type through reassortment (Ghedin et al., 2005)]. A mutation is a universal form of genetic change. It underlies numerous adaptive responses and critical biological transitions in viruses, and it is a prerequisite for recombination and reassortment to have a biological impact. If mutations were not present in different template molecules during replication, recombinants with the crossover point at equivalent positions of the parental genomes would be "silent," and display the same behavior as the parental genomes. Apparently, "silent" recombination events may take place within replicative units; even if some mutations distinguished individual genomes of the same quasispecies swarm, a recombinant would not be distinguished from a mutant genome. The frequency of recombination in HIV-1 was noticed only when the acquired

immune deficiency syndrome (AIDS) pandemic had advanced, and the virus had diversified through the accumulation of mutations. Similar arguments apply to segment reassortment. Genomes necessitate mutation-driven diversification for reassortment to provide a biological difference; detection of a reassortant will be easier the larger the replicative advantage it confers to the virus (see also Chapter 10).

The evolutionary significance of recombination has been viewed in two opposite ways: as a means to rescue fit genomes from less fit parents (a conservative force that eliminates deleterious mutations), or as a means to explore new genomic forms for adaptive potential (a vast substrate for the exploration of sequence space; Chapter 3) [reviewed in (Zimmern, 1988; Lai, 1992; Worobey and Holmes, 1999; Simmonds, 2010; Perez-Losada et al., 2015)]. Recombination has been probably at the origin of new viruses that presently occupy a well-established niche, and it is also at play today to expand diversity during the spread of viruses. As a historical event, the coronavirus mouse hepatitis virus appears to have acquired its hemagglutinin-esterase gene by recombination with an influenza C virus. The alphavirus Western equine encephalitis virus originated probably by recombination between Sindbis-like and Eastern equine encephalitis-like viruses [reviewed in different chapters of (Domingo et al., 2008)].

Several recent poliomyelitis outbreaks have been associated with recombinants between oral poliovirus vaccine (OPV) viruses and other circulating enteroviruses (Gavrilin et al., 2000; Kew et al., 2002; Oberste et al., 2004; Muslin et al., 2015). Intersubtype HIV-1 recombinants play a key role in current HIV-1 diversification, with around 100 circulating recombinant forms (and the number is growing) displaying complex mosaic structures (multiple crossover sites) (Thomson et al., 2002; Gerhardt et al., 2005). In addition, other HIV-1 recombinants have been characterized that are not established epidemiologically. Fewer HCV recombinants have been identified, but the number is likely to increase as the virus diversifies in nature. Positive selection of HIV-1 recombinants that unite different drug-resistant mutations in the same genome offers an example of the conservative force of recombination to rescue fit viruses in the face of a strong selective constraint (Menéndez-Arias, 2002). Recombination is expected to play an increasing role in the spread of drug resistance among viruses for which new antiviral agents are in use, such as HBV and HCV.

Some defective DNA and RNA genomes that include *indels*, notably DI RNAs, which originate from recombination events can play an important role in the establishment and maintenance of persistent infections in cell culture, and can modulate viral infections in vivo (Holland and Villarreal, 1974; Roux et al., 1991; Rezelj et al., 2018). Detailed genetic and biochemical analyses by A. Huang, J.J. Holland and their colleagues on the generation of VSV DI's and their interplay with the standard, infectious VSV contributed to unveil a continuous dynamics of genetic variation, competition, and selection, observable within short time intervals, a hallmark of RNA genetics (Palma and Huang, 1974; Holland et al., 1982), fully confirmed by application of new sequencing techniques. DI particles and defective genomes are present in populations of positive- and negative-strand RNA viruses as they multiply in their natural hosts (Nüesch et al., 1989; Drolet et al., 1995; Li et al., 2011; Saira et al., 2013; Ke et al., 2013; Rezelj et al., 2018). Their widespread presence in vivo may mean that they are an unavoidable side-product of the replication machineries (i.e., instruction to recombine) or that selection might have favored their generation. Both possibilities are compatible. An instruction whose result is a means to modulate replication of the corresponding standard viruses or the antiviral immune response will be selected as a consequence of its biological effects.

DI RNAs can be regarded as the tip of the iceberg of many classes of defective genomes with a range of interfering or potentiating

capacities that may coexist with standard animal, plant, insect, and bacterial viruses, and that may facilitate persistence and modulate disease symptoms (Holland et al., 1982; Vogt and Jackson, 1999; López-Ferber et al., 2003; Rosario et al., 2005; Sachs and Bull, 2005; Villarreal, 2005; Aaskov et al., 2006; Rezelj et al., 2018). Noncytopathic coxsackievirus B3 (CVB3) variants with deletions at the untranslated 5′-genomic region were isolated from hearts of mice inoculated with CVB3. The variants replicated in vivo and were associated with long-term viral persistence (Kim et al., 2005).

Despite the continuous dynamics of the escape of infectious virus to the interfering activities of DIs, some authors consider DIs as potential antiviral agents (Dimmock and Easton, 2014). If defective genomes are competent in RNA (or DNA) synthesis or are complemented to replicate, they can act as dominant-negative swarms, provided they reach a sufficient load. In this manner, they may underlie the suppressive effects of mutant spectra of viral quasispecies. Intramutant spectrum decrease of replicative capacity due to the presence of defective genomes is one of the mechanisms of virus extinction evoked by enhanced mutagenesis (Chapters 3 and 9).

Recombination events must have been the last step in ancestral processes of horizontal gene transfer that mediated the incorporation of host genes (or gene segments) into viral genomes, and vice versa. Host genes related to immune responses were probably captured by complex DNA viruses at early stages of their evolution (Alcami, 2003; McFadden, 2005). Mosaicism associated with nonhomologous recombination events is the norm among tailed bacteriophages (Canchaya et al., 2003). Nonhomologous recombination can give rise to genomic sequences with a viral and a nonviral moiety. They include DI RNAs of Sindbis virus-containing cellular RNA sequences at their 5′ ends, some cytopathic forms of bovine viral diarrhea virus, RNA of potato leafroll virus-containing tobacco chloroplast RNA, or an influenza virus with an insertion of ribosomal RNA into the hemagglutinin gene, mentioned in Chapter 5 regarding transient, high fitness levels [reviewed in (Domingo et al., 2001)].

Phylogenetic analyses have suggested that recombination between RNA and DNA viruses might have occurred to give rise to some present-day single-stranded DNA viruses (Stedman, 2013). However, the evidence for this attractive possibility is indirect and, to my knowledge, no experimental evidence of viral RNA-viral DNA recombination in cell culture or in vivo has been reported. Viability of mutant and recombinant viral genomes is severely constrained by the evolutionary history of the virus that has shaped viral genomes as coordinated sets of modules (Botstein, 1980, 1981; Zimmern, 1988; Koonin and Dolja, 2014). Experimental studies with engineered recombinant viruses have shown that modularity can restrict recombination (Martin et al., 2005).

The three molecular mechanisms of viral genome variation (mutation, recombination, and reassortment) are not incompatible, although it may sometimes be difficult to discern their occurrence (Varsani et al., 2018). It would be truly remarkable if a viral system could be proven to be totally devoid of one of the mechanisms of genetic variation. It would imply that there are powerful molecular reasons to dispense with an effective adaptive mechanism. Absence of a mechanism is extremely difficult to demonstrate but, if we could, its basis would open a new chapter of molecular virology.

2.12.1 Mechanistically unavoidable versus evolutionarily relevant genetic variation

There is an ongoing controversy regarding clonality versus nonclonality in biological evolution, particularly concerning the evolution of cellular parasites (Heitman, 2010; Tibayrenc

and Ayala, 2012, 2014; Ramirez and Llewellyn, 2014; Hauser and Cushion, 2018). Clonality means asexual progeny from a single ancestor. In the case of viruses, clonal evolution emphasizes reproduction without the exchange of genetic material among two or more parental genomes. Sexual reproduction necessarily involves the exchange of genetic material. The question for viruses is interesting because we would be inclined to propose clonal evolution despite considerable promiscuity of recombination and reassortment. A tentative solution was offered based on one assumption and some experimental observations. The assumption is that recombination is not a requirement for viruses to complete their infection cycles. Despite the possibility that recombination might be imprinted in the replication apparatuses (rendering it inherent to replication), its occurrence is not a necessity. The experimental observations are of two sorts. One is that historically recombination and reassortment have been at the origin of the emergence and reemergence of viral pathogens [western equine encephalitis virus, pandemic influenza viruses, emergent circulating poliovirus, HIV-1 and HCV recombinants, etc. (Section 2.12)]. The second observation is that recombination-based genome segmentation can occur given the adequate population dynamics and competitive environment, as documented with FMDV (Section 2.11). In consequence, the distinction was made between mechanistically unavoidable but biologically irrelevant, and meaningful recombination (Perales et al., 2015). The latter form of recombination requires prior diversification of parental genomes by mutation and a number of cellular and epidemiological conditions. Despite its relevance to evolutionary transitions and viral emergence, it is not a requirement for virus survival, propagation, and evolution. This marks a contrast with the genomic exchanges associated with sexual reproduction. The proposal is that viruses evolve clonally at widely different time scales (intrahost or within-host evolution vs. long-term evolution at the epidemiological level). Similar arguments apply to mutation. This point will be addressed again in the closing Chapter 10, concerning implications of clonality.

2.13 Overview and concluding remarks

All forms of genetic variation of viruses must be viewed essentially as blind processes despite preferences of nucleotide sequences or structures for mutation and recombination events: *hot spots* with higher than average rate values, and *cold spots* with lower than average rate values. Mutation originates largely in fluctuations of electronic structure that modify base-pairing properties, and from features of polymerase-template interactions, not subject to regulation, in the sense that we understand the regulation of gene expression or enzymatic activity. Absence of regulation is not incompatible with long-term evolution having shaped the molecular interactions that yield a level of mutagenesis compatible with survival and adaptability. Given the biological consequences of mutation rates, many additional studies are needed for the biological phyla, to quantity not only basal mutation rates but also the possible presence of mutator strains.

Similar arguments can be used for recombination and reassortment. The number of segments that enter a new genomic constellation may be regulated but not which variant forms of the individual segments will make up the new viral particles. It is short-term selection acting at the very center of replication and recombinant complexes that preserves some mutant and recombinant forms in detriment of others. Subsequent levels of selection occur when variant forms expand in multiple rounds of infection first within cells, then within an organism and then at the epidemiological level. The very nature of life in our planet has been built upon an inherent tendency to instruct variation in an incessant fashion, as necessary and unavoidable as the physical principles that dictate the behavior of our universe.

The net result of all mechanisms of genetic variation available to a virus is the generation of repertoires of variant genomes for random drift and selective forces to act upon. In other terms, genetic variation sets the scene for the actors of evolution to play their roles, and secure a continuous input of new forms despite subtle or catastrophic environmental perturbations. The same forces that drive general evolution have produced the dominant virus forms we see in nature, with all their nuances in the interaction with cell components. The adaptation of viruses to participate in intracellular processes with cells dictates that genetic variation of viruses has its limits to prevent deleteriousness. This is currently exemplified by the effects of amino acid substitutions in viral polymerases that either increase or decrease template-copying fidelity. Viruses have reached a compromise between the stability of core information and flexibility for adaptability. Although not yet treated in this chapter, viral population numbers are a key parameter in the evolutionary events. Next chapters address some of these questions, not only in general conceptual terms but also in the way evolution affects our daily confrontation with viral disease (see Summary Box).

Summary Box

- Mutation, recombination, and genome segment reassortment are the mechanisms of genetic variation used by DNA and RNA viruses. Mutations are due mainly to changes in the electronic distributions of the standard nucleotides, to damage of nucleotides by external influences, and by alignment alterations of the template relative to product polynucleotide chains. The effect of a mutation can range from being well-tolerated to highly detrimental or lethal.
- Mutation frequencies are only an indirect consequence of mutation rates. Their values for viruses whose replication is catalyzed by polymerases devoid of proofreading-repair activity are 10^5- to 10^6-fold higher than those displayed by replicative cellular DNA polymerases. Error-prone replication is a hallmark of RNA viruses and some DNA viruses. The larger the amount of genetic information encoded in a viral genome, the lower the mutation rate must be to maintain the genetic message.
- Several mechanisms of genetic recombination have been described for DNA and RNA viruses. The best characterized is homologous recombination whose frequency of occurrence is dependent on the replicative machinery, in particular, polymerase processivity. Genome segment reassortment is operative in segmented genomes, and it gives rise to biologically relevant changes, such as an antigenic shift in the influenza type A viruses.
- Studies with foot-and-mouth disease virus have shown that extensive evolution of an unsegmented RNA genome has the potential to undergo a recombination-mediated transition akin to genome segmentation. Therefore, segmented and unsegmented forms of RNA viruses need not be considered as completely unrelated classes of genome organization.
- Recombination and genome segment reassortment have been viewed as conservative forces to rescue viable genomes from a damaged pool, and also as a means to explore new genomic compositions that deviate from their parents. All forms of genetic variation give rise to repertoires of variant genomes on which selection and random drift act to produce the viral forms that we isolate and study.

References

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

Achaz, G., 2009. Frequency spectrum neutrality tests: one for all and all for one. Genetics 183, 249–258.

Agol, V.I., 2010. Picornaviruses as a model for studying the nature of RNA recombination. In: Ehrenfeld, E., Domingo, E., Roos, R.P. (Eds.), The Picornaviruses. ASM Press, Washington, DC, pp. 239–252.

Agol, V.I., Gmyl, A.P., 2018. Emergency services of viral RNAs: repair and remodeling. Microbiol. Mol. Biol. Rev. 82 e00067-17.

Aguilera, A., Garcia-Muse, T., 2013. Causes of genome instability. Annu. Rev. Genet. 47, 1–32.

Akashi, H., 2001. Gene expression and molecular evolution. Curr. Opin. Genet. Dev. 11, 660–666.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., et al., 2002. Molecular Biology of the Cell. Garland Science, New York, N.Y.

Alcami, A., 2003. Viral mimicry of cytokines, chemokines and their receptors. Nat. Rev. Immunol. 3, 36–50.

Alejska, M., Figlerowicz, M., Malinowska, N., Urbanowicz, A., Figlerowicz, M., 2005. A universal BMV-based RNA recombination system—how to search for general rules in RNA recombination. Nucleic Acids Res. 33, e105.

Aragones, L., Guix, S., Ribes, E., Bosch, A., Pinto, R.M., 2010. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis a virus capsid. PLoS Pathog. 6, e1000797.

Arias, A., Arnold, J.J., Sierra, M., Smidansky, E.D., Domingo, E., et al., 2008. 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. 82, 12346–12355.

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

Barbezange, C., Jones, L., Blanc, H., Isakov, O., Celniker, G., et al., 2018. Seasonal genetic drift of human influenza A virus quasispecies revealed by deep sequencing. Front. Microbiol. https://doi.org/10.3389/fmicb.2018.02596.

Barrett, J.W., McFadden, G., 2008. Origin and evolution of poxviruses. In: Domingo, E., Parrish, C.R., Holland, J.J. (Eds.), Origin and Evolution of Viruses, second ed. Elsevier, Oxford, pp. 431–446.

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

Beaucourt, S., Vignuzzi, M., 2014. Ribavirin: a drug active against many viruses with multiple effects on virus replication and propagation. Molecular basis of ribavirin resistance. Curr. Opin. Virol. 8C, 10–15.

Bebenek, A., Ziuzia-Graczyk, I., 2018. Fidelity of DNA replication – a matter of proofreading. Curr. Genet. 64, 985–996.

Bentley, K., Evans, D.J., 2018. Mechanisms and consequences of positive-strand RNA virus recombination. J. Gen. Virol. 99, 1345–1356.

Bernad, A., Blanco, L., Lazaro, J.M., Martin, G., Salas, M., 1989. A conserved 3′- 5′ exonuclease active site in prokaryotic and eukaryotic DNA polymerases. Cell 59, 219–228.

Bloomfield, V.A., Crothers, D.M., Tinoco Jr., I., 2000. Nucleic Acids. Structures, Properties, and Functions. University Science Books, Sausalito, CA.

Boehmer, P.E., Nimonkar, A.V., 2003. Herpes virus replication. IUBMB Life 55, 13–22.

Borderia, A.V., Rozen-Gagnon, K., Vignuzzi, M., 2016. Fidelity variants and RNA quasispecies. Curr. Top. Microbiol. Immunol. 392, 303–322.

Botstein, D., 1980. A theory of modular evolution for bacteriophages. Ann. NY Acad. Sci. 354, 484–491.

Botstein, D., 1981. A modular theory of virus evolution. In: Fields, B.N., Jaenisch, R., Fox, C.F. (Eds.), Animal Virus Genetics. Academic Press, New York, pp. 363–384.

Bradwell, K., Combe, M., Domingo-Calap, P., Sanjuan, R., 2013. Correlation between mutation rate and genome size in riboviruses: mutation rate of bacteriophage Qbeta. Genetics 195, 243–251.

Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., et al., 2011. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat. Genet. 43, 242–245.

Brule, C.E., Grayhack, E.J., 2017. Synonymous codons: choose wisely for expression. Trends Genet. 33, 283–297.

Bujarski, J.J., 2013. Genetic recombination in plant-infecting messenger-sense RNA viruses: overview and research perspectives. Front. Plant Sci. 4, 68.

Bushman, F., 2002. Lateral DNA Transfer. Mechanisms and Consequences. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Campagnola, G., McDonald, S., Beaucourt, S., Vignuzzi, M., Peersen, O.B., 2015. Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. J. Virol. 89, 275–286.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., Brussow, H., 2003. Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. 6, 417–424.

Castro, C., Arnold, J.J., Cameron, C.E., 2005. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. Virus Res. 107, 141–149.

Cattaneo, R., Billeter, M.A., 1992. Mutations and A/I hypermutations in measles virus persistent infections. Curr. Top. Microbiol. Immunol. 176, 63–74.

Chao, L., 1988. Evolution of sex in RNA viruses. J. Theor. Biol. 133, 99–112.

Chetverin, A.B., Kopein, D.S., Chetverina, H.V., Demidenko, A.A., Ugarov, V.I., 2005. Viral RNA-directed RNA polymerases use diverse mechanisms to promote recombination between RNA molecules. J. Biol. Chem. 280, 8748–8755.

Cheynier, R., Kils-Hutten, L., Meyerhans, A., Wain-Hobson, S., 2001. Insertion/deletion frequencies match those of point mutations in the hypervariable regions of the simian immunodeficiency virus surface envelope gene. J. Gen. Virol. 82, 1613–1619.

Coffey, L.L., Beeharry, Y., Borderia, A.V., Blanc, H., Vignuzzi, M., 2011. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. Proc. Natl. Acad. Sci. U.S.A. 108, 16038–16043.

Coffin, J.M., 1990. Genetic variation in retroviruses. Appl. Virol. Res. 2, 11–33.

Combe, M., Sanjuan, R., 2014. Variation in RNA virus mutation rates across host cells. PLoS Pathog. 10, e1003855.

Crandall, K.A., Kelsey, C.R., Imamichi, H., Lane, H.C., Salzman, N.P., 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Mol. Biol. Evol. 16, 372–382.

Czarnecki, M.W., Traktman, P., 2017. The vaccinia virus DNA polymerase and its processivity factor. Virus Res. 234, 193–206.

Czech, A., Fedyunin, I., Zhang, G., Ignatova, Z., 2010. Silent mutations in sight: co-variations in tRNA abundance as a key to unravel consequences of silent mutations. Mol. Biosyst. 6, 1767–1772.

de Vega, M., Lazaro, J.M., Mencia, M., Blanco, L., Salas, M., 2010. Improvement of phi29 DNA polymerase amplification performance by fusion of DNA binding motifs. Proc. Natl. Acad. Sci. U.S.A. 107, 16506–16511.

Denny, S.K., Greenleaf, W.J., 2018. Linking RNA sequence, structure, and function on massively parallel high-throughput sequences. Cold Spring Harb. Perspect. Biol. https://doi.org/10.1101/cshperspect.a 032300.

Dimmock, N.J., Easton, A.J., 2014. Defective interfering influenza virus RNAs: time to reevaluate their clinical potential as broad-spectrum antivirals? J. Virol. 88, 5217–5227.

Domingo, E., 2005. Virus entry into error catastrophe as a new antiviral strategy. Virus Res. 107 (Special Issue), 115–228.

Domingo, E., Perales, C., 2019. Viral Quasispecies. PLoS Genetics in press.

Domingo, E., Sabo, D., Taniguchi, T., Weissmann, C., 1978. Nucleotide sequence heterogeneity of an RNA phage population. Cell 13, 735–744.

Domingo, E., Mateu, M.G., Martínez, M.A., Dopazo, J., Moya, A., et al., 1990. Genetic variability and antigenic diversity of foot-and-mouth disease virus. In: Kurkstak, E., Marusyk, R.G., Murphy, S.A., Van-Regenmortel, M.H.V. (Eds.), Applied Virology Research. Plenum Publishing Co., New York, pp. 233–266.

Domingo, E., Holland, J.J., Biebricher, C., Eigen, M., 1995. Quasispecies: the concept and the word. In: Gibbs, A., Calisher, C., García-Arenal, F. (Eds.), Molecular Evolution of the Viruses. Cambridge University Press, Cambridge, pp. 171–180.

Domingo, E., Biebricher, C., Eigen, M., Holland, J.J., 2001. Quasispecies and RNA Virus Evolution: Principles and Consequences. Landes Bioscience, Austin.

Domingo, E., Escarmis, C., Baranowski, E., Ruiz-Jarabo, C.M., Carrillo, E., et al., 2003. Evolution of foot-and-mouth disease virus. Virus Res. 91, 47–63.

Domingo, E., Escarmís, C., Menéndez-Arias, L., Perales, C., Herrera, M., et al., 2008. Viral quasispecies: dynamics, interactions and pathogenesis. In: Domingo, E., Parrish, C., Holland, J.J. (Eds.), Origin and Evolution of Viruses. Elsevier, Oxford, pp. 87–118.

Drake, J.W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U.S.A. 88, 7160–7164.

Drake, J.W., Holland, J.J., 1999. Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. U.S.A. 96, 13910–13913.

Drolet, B.S., Chiou, P.P., Heidel, J., Leong, J.A., 1995. Detenction of truncated virus particles in a persistent RNA virus infection *in vivo*. J. Virol. 69, 2140–2147.

Earl, D.J., Deem, M.W., 2004. Evolvability is a selectable trait. Proc. Natl. Acad. Sci. U.S.A. 101, 11531–11536.

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

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

Eigen, M., 2002. Error catastrophe and antiviral strategy. Proc. Natl. Acad. Sci. U.S.A. 99, 13374–13376.

Eigen, M., 2013. From Strange Simplicity to Complex Familiarity. Oxford University Press, Oxford.

Eigen, M., Biebricher, C.K., 1988. Sequence space and quasispecies distribution. In: Domingo, E., Ahlquist, P., Holland, J.J. (Eds.), RNA Genetics. CRC Press Inc, Boca Raton, FL, pp. 211–245.
Elena, S.F., Sanjuan, R., 2005. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. J. Virol. 79, 11555–11558.
Escarmís, C., Dávila, M., Charpentier, N., Bracho, A., Moya, A., et al., 1996. Genetic lesions associated with Muller's ratchet in an RNA virus. J. Mol. Biol. 264, 255–267.
Escarmís, C., Dávila, M., Domingo, E., 1999. Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. J. Mol. Biol. 285, 495–505.
Fan, X., Di Bisceglie, A.M., 2010. RT-PCR amplification and cloning of large viral sequences. Methods Mol. Biol. 630, 139–149.
Faure, G., Ogurtsov, A.Y., Shabalina, S.A., Koonin, E.V., 2017. Adaptation of mRNA structure to control protein folding. RNA Biol. 14, 1649–1654.
Ferrer-Orta, C., Arias, A., Escarmis, C., Verdaguer, N., 2006. A comparison of viral RNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. 16, 27–34.
Ferrer-Orta, C., Agudo, R., Domingo, E., Verdaguer, N., 2009. Structural insights into replication initiation and elongation processes by the FMDV RNA-dependent RNA polymerase. Curr. Opin. Struct. Biol. 19, 752–758.
Forloni, M., Liu, A.Y., Wajapeyee, N., 2018. Random mutagenesis using error-prone DNA polymerases. Cold Spring Harb. Protoc. https://doi.org/10.1101/pdb.prot097741.
Friedberg, E.C., Wagner, R., Radman, M., 2002. Specialized DNA polymerases, cellular survival, and the genesis of mutations. Science 296, 1627–1630.
Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., et al., 2006. DNA Repair and Mutagenesis. American Society for Microbiology, Washington, DC.
Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147, 915–925.
Gago, S., Elena, S.F., Flores, R., Sanjuan, R., 2009. Extremely high mutation rate of a hammerhead viroid. Science 323, 1308.
Gallei, A., Pankraz, A., Thiel, H.J., Becher, P., 2004. RNA recombination in vivo in the absence of viral replication. J. Virol. 78, 6271–6281.
Galli, A., Bukh, J., 2014. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. Trends Microbiol. 22, 354–364.
Ganai, R.A., Johansson, E., 2016. DNA replication-a matter of fidelity. Mol. Cell 62, 745–755.
García-Arriaza, J., Manrubia, S.C., Toja, M., Domingo, E., Escarmís, C., 2004. Evolutionary transition toward defective RNAs that are infectious by complementation. J. Virol. 78, 11678–11685.
Gatenby, R.A., Frieden, B.R., 2004. Information dynamics in carcinogenesis and tumor growth. Mutat. Res. 568, 259–273.
Gavrilin, G.V., Cherkasova, E.A., Lipskaya, G.Y., Kew, O.M., Agol, V.I., 2000. Evolution of circulating wild poliovirus and of vaccine-derived poliovirus in an immunodeficient patient: a unifying model. J. Virol. 74, 7381–7390.
Gerhardt, M., Mloka, D., Tovanabutra, S., Sanders-Buell, E., Hoffmann, O., et al., 2005. In-depth, longitudinal analysis of viral quasispecies from an individual triply infected with late-stage human immunodeficiency virus type 1, using a multiple PCR primer approach. J. Virol. 79, 8249–8261.
Ghedin, E., Sengamalay, N.A., Shumway, M., Zaborsky, J., Feldblyum, T., et al., 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 437, 1162–1166.
Gibbs, A.J., Calisher, C.H., García-Arenal, F. (Eds.), 1995. Molecular Basis of Virus Evolution. Cambridge University Press, Cambridge.
Gmyl, A.P., Pilipenko, E.V., Maslova, S.V., Belov, G.A., Agol, V.I., 1993. Functional and genetic plasticities of the poliovirus genome: quasi- infectious RNAs modified in the 5′-untranslated region yield a variety of pseudorevertants. J. Virol. 67, 6309–6316.
Gmyl, A.P., Korshenko, S.A., Belousov, E.V., Khitrina, E.V., Agol, V.I., 2003. Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? RNA 9, 1221–1231.
Gnadig, N.F., Beaucourt, S., Campagnola, G., Borderia, A.V., Sanz-Ramos, M., et al., 2012. Coxsackievirus B3 mutator strains are attenuated in vivo. Proc. Natl. Acad. Sci. U.S.A. 109, E2294–E2303.
Graci, J.D., Cameron, C.E., 2004. Challenges for the development of ribonucleoside analogues as inducers of error catastrophe. Antivir. Chem. Chemother. 15, 1–13.
Hamano, T., Matsuo, K., Hibi, Y., Victoriano, A.F., Takahashi, N., et al., 2007. A single-nucleotide synonymous mutation in the gag gene controlling human immunodeficiency virus type 1 virion production. J. Virol. 81, 1528–1533.
Hauser, P.M., Cushion, M.T., 2018. Is sex necessary for the proliferation and transmission of *Pneumocystis*? PLoS Pathog. 14, e1007409.
Heitman, J., 2010. Evolution of eukaryotic microbial pathogens via covert sexual reproduction. Cell Host Microbe 8, 86–99.
Hirose, Y., Onuki, M., Tenjimbayashi, Y., Mori, S., Ishii, Y., et al., 2018. Within-host variations of human papillomaviruses reveal APOBEC signature mutagenesis in the viral genome. J. Virol. 92, e00017–18.
Hirsch, N., Birnbaum, R.Y., 2015. Dual function of DNA sequences: protein-coding sequences function as transcriptional enhancers. Perspect. Biol. Med. 58, 182–195.

Holland, J.J., Villarreal, L.P., 1974. Persistent noncytocidal vesicular stomatitis virus infections mediated by defective T particles that suppress virion transcriptase. Proc. Natl. Acad. Sci. U.S.A. 71, 2956–2960.

Holland, J.J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., et al., 1982. Rapid evolution of RNA genomes. Science 215, 1577–1585.

Holland, J.J., de la Torre, J.C., Steinhauer, D.A., Clarke, D., Duarte, E., et al., 1989. Virus mutation frequencies can be greatly underestimated by monoclonal antibody neutralization of virions. J. Virol. 63, 5030–5036.

Jin, Z., Leveque, V., Ma, H., Johnson, K.A., Klumpp, K., 2013. NTP-mediated nucleotide excision activity of hepatitis C virus RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 110, E348–E357.

Joyce, G.F., 2004. Directed evolution of nucleic acid enzymes. Annu. Rev. Biochem. 73, 791–836.

Ke, R., Aaskov, J., Holmes, E.C., Lloyd-Smith, J.O., 2013. Phylodynamic analysis of the emergence and epidemiological impact of transmissible defective dengue viruses. PLoS Pathog. 9, e1003193.

Kew, O., Morris-Glasgow, V., Landaverde, M., Burns, C., Shaw, J., et al., 2002. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. Science 296, 356–359.

Kim, K.S., Tracy, S., Tapprich, W., Bailey, J., Lee, C.K., et al., 2005. 5′-Terminal deletions occur in coxsackievirus B3 during replication in murine hearts and cardiac myocyte cultures and correlate with encapsidation of negative-strand viral RNA. J. Virol. 79, 7024–7041.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK.

Kimura, M., 1989. The neutral theory of molecular evolution and the world view of the neutralists. Genome 31, 24–31.

King, A.M.Q., 1988. Genetic recombination in positive strand RNA viruses. In: Domingo, E., Holland, J.J., Ahlquist, P. (Eds.), RNA Genetics. CRC Press Inc, Boca Raton, FL, pp. 149–165.

King, J.L., Jukes, T.H., 1969. Non-Darwinian evolution. Science 164, 788–798.

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

Kolakofsky, D., Roux, L., Garcin, D., Ruigrok, R.W., 2005. Paramyxovirus mRNA editing, the "rule of six" and error catastrophe: a hypothesis. J. Gen. Virol. 86, 1869–1877.

Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. Microbiol. Mol. Biol. Rev. 78, 278–303.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22, 3096–3098.

Kryazhimskiy, S., Plotkin, J.B., 2008. The population genetics of dN/dS. PLoS Genet. 4, e1000304.

Kunkel, T.A., Erie, D.A., 2005. DNA mismatch repair. Annu. Rev. Biochem. 74, 681–710.

Lafforgue, G., Martinez, F., Sardanyes, J., de la Iglesia, F., Niu, Q.W., et al., 2011. Tempo and mode of plant RNA virus escape from RNA interference-mediated resistance. J. Virol. 85, 9686–9695.

Lai, M.M.C., 1992. Genetic recombination in RNA viruses. Curr. Top. Microbiol. Immunol. 176, 21–32.

Lenski, R.E., Winkworth, C.L., Riley, M.A., 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. J. Mol. Evol. 56, 498–508.

Levy, D.N., Aldrovandi, G.M., Kutsch, O., Shaw, G.M., 2004. Dynamics of HIV-1 recombination in its natural target cells. Proc. Natl. Acad. Sci. U.S.A. 101, 4204–4209.

Li, D., Lott, W.B., Lowry, K., Jones, A., Thu, H.M., Aaskov, J., 2011. Defective interfering viral particles in acute dengue infections. PLoS One 6, e19447.

Ling, J., Reynolds, N., Ibba, M., 2009. Aminoacyl-tRNA synthesis and translational quality control. Annu. Rev. Microbiol. 63, 61–78.

Liu, D., Keijzers, G., Rasmussen, L.J., 2017. DNA mismatch repair and its many roles in eukaryotic cells. Mutat. Res. 773, 174–187.

Liu, W., Shi, X., Gong, P., 2018. A unique intra-molecular fidelity-modulating mechanism identified in a viral RNA-dependent RNA polymerase. Nucleic Acids Res. 46, 10840–10854.

López-Ferber, M., Simon, O., Williams, T., Caballero, P., 2003. Defective or effective? Mutualistic interactions between virus genotypes. Proc. R. Soc. Lond. B Biol. Sci. 270, 2249–2255.

Lowry, K., Woodman, A., Cook, J., Evans, D.J., 2014. Recombination in enteroviruses is a biphasic replicative process involving the generation of greater-than genome length 'imprecise' intermediates. PLoS Pathog. 10, e1004191.

Maas, S., Rich, A., Nishikura, K., 2003. A-to-I RNA editing: recent news and residual mysteries. J. Biol. Chem. 278, 1391–1394.

Makhoul, C.H., Trifonov, E.N., 2002. Distribution of rare triplets along mRNA and their relation to protein folding. J. Biomol. Struct. Dyn. 20, 413–420.

Makino, S., Keck, J.G., Stohlman, S.A., Lai, M.M., 1986. High-frequency RNA recombination of murine coronaviruses. J. Virol. 57, 729–737.

Malpica, J.M., Fraile, A., Moreno, I., Obies, C.I., Drake, J.W., et al., 2002. The rate and character of spontaneous mutation in an RNA virus. Genetics 162, 1505–1511.

Mansky, L.M., Temin, H.M., 1995. Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. 69, 5087–5094.

Mansky, L.M., Le Rouzic, E., Benichou, S., Gajary, L.C., 2003. Influence of reverse transcriptase variants, drugs, and Vpr on human immunodeficiency virus type 1 mutant frequencies. J. Virol. 77, 2071–2080.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260−262.

Martínez, M.A., Jordan-Paiz, A., Franco, S., Nevot, M., 2016. Synonymous viral genome recoding as a tool to impact viral fitness. Trends Microbiol. 24, 134−147.

McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassorment in segmented RNA viruses: mechanisms and outcomes. Nat. Rev. Microbiol. 14, 448−460.

McFadden, G., 2005. Poxvirus tropism. Nat. Rev. Microbiol. 3, 201−213.

McGeoch, D.J., Davison, A.J., Dolan, A., Gatherer, D., Sevilla-Reyes, E.E., 2008. Molecular evolution of the *Herpesvirales*. In: Domingo, E., Parrish, C.R., Holland, J.J. (Eds.), Origin and Evolution of Viruses, second ed. Elsevier, Oxford, pp. 447−476.

Menéndez-Arias, L., 2002. Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases. Prog. Nucleic Acid Res. Mol. Biol. 71, 91−147.

Meng, T., Kwang, J., 2014. Attenuation of human enterovirus 71 high-replication-fidelity variants in AG129 mice. J. Virol. 88, 5803−5815.

Methot, S.P., Di Noia, J.M., 2017. Molecular mechanisms of somatic hypermutation and class switch recombination. Adv. Immunol. 133, 37−87.

Meyer, P.R., Matsuura, S.E., So, A.G., Scott, W.A., 1998. Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. Proc. Natl. Acad. Sci. U.S.A. 95, 13471−13476.

Meyerhans, A., Vartanian, J.-P., 1999. The fidelity of cellular and viral polymerases and its manipulation for hypermutagenesis. In: Domingo, E., Webster, R.G., Holland, J.J. (Eds.), Origin and Evolution of Viruses. Academic Press, San Diego, pp. 87−114.

Miller, R.V., Day, M.J., 2004. Microbial Evolution. Gene Establishment, Survival and Exchange. ASM Press, Washington, D.C.

Mills, D.R., Peterson, R.L., Spiegelman, S., 1967. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. Proc. Natl. Acad. Sci. U.S.A. 58, 217−224.

Minskaia, E., Hertzig, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., et al., 2006. Discovery of an RNA virus 3′→5′ exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc. Natl. Acad. Sci. U.S.A. 103, 5108−5113.

Mittal, P., Brindle, J., Stephen, J., Plotkin, J.B., Kudla, G., 2018. Codon usage influences fitness through RNA toxicity. Proc. Natl. Acad. Sci. U.S.A. 115, 8639−8644.

Moreno, E., Ojosnegros, S., Garcia-Arriaza, J., Escarmis, C., Domingo, E., et al., 2014. Exploration of sequence space as the basis of viral RNA genome segmentation. Proc. Natl. Acad. Sci. U.S.A. 111, 6678−6683.

Moreno, E., Gallego, I., Gregori, J., Lucía-Sanz, A., Soria, M.E., et al., 2017. Internal disequilibria and phenotypic diversification during replication of hepatitis C virus in a noncoevolving celular environment. J. Virol. 91, e02505−e02516.

Morse, S.S. (Ed.), 1994. The Evolutionary Biology of Viruses. Raven Press, New York.

Muslin, C., Joffret, M.-L., Pelletier, I., Blondel, B., Delpeyroux, F., 2015. Evolution and emergence of enteroviruses through intra- and inter-species recombination: plasticity and phenotypic impact of modular genetic exchanges in the 5′untranslated region. PLoS Pathog. 11, e1005266.

Naegeli, H., 1997. Mechanisms of DNA Damage Recognition in Mammalian Cells. Landes Bioscience, Austin, Texas.

Nagy, P.D., Simon, A.E., 1997. New insights into the mechanisms of RNA recombination. Virology 235, 1−9.

Nee, S., 1987. The evolution of multicompartmental genomes in viruses. J. Mol. Evol. 25, 277−281.

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

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418−426.

Nevot, M., Martrus, G., Clotet, B., Martinez, M.A., 2011. RNA interference as a tool for exploring HIV-1 robustness. J. Mol. Biol. 413, 84−96.

Nevot, M., Jordan-Paiz, A., Martrus, G., Andrés, C., García-Cehic, D., et al., 2018. HIV-1 protease evolvability is affected by synonymous nucleotide recoding. J. Virol. 92 e00777-18.

Nimonkar, A.V., Boehmer, P.E., 2003. Reconstitution of recombination-dependent DNA synthesis in herpes simplex virus 1. Proc. Natl. Acad. Sci. U.S.A. 100, 10201−10206.

Nishikura, K., 2010. Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. 79, 321−349.

Nishimura, I., Kurokawa, M., Liu, L., Ying, B.W., 2017. Coordinated changes in mutation and growth rates induced by genome reduction. mBio 8 e00676-17.

Novella, I.S., 2004. Negative effect of genetic bottlenecks on the adaptability of vesicular stomatitis virus. J. Mol. Biol. 336, 61−67.

Novella, I.S., Zarate, S., Metzgar, D., Ebendick-Corpus, B.E., 2004. Positive selection of synonymous mutations in vesicular stomatitis virus. J. Mol. Biol. 342, 1415−1421.

Nüesch, J.P.F., De Chastonay, J., Siegl, G., 1989. Detection of defective genomes in hepatitis A virus particles present in clinical specimens. J. Gen. Virol. 70, 3475−3480.

Oberste, M.S., Maher, K., Pallansch, M.A., 2004. Evidence for frequent recombination within species human enterovirus B based on complete genomic sequences of all thirty-seven serotypes. J. Virol. 78, 855−867.

Ochman, H., Elwyn, S., Moran, N.A., 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. U.S.A. 96, 12638–12643.

Ojosnegros, S., Garcia-Arriaza, J., Escarmis, C., Manrubia, S.C., Perales, C., et al., 2011. Viral genome segmentation can result from a trade-off between genetic content and particle stability. PLoS Genet. 7, e1001344.

Palma, E.L., Huang, A.S., 1974. Cyclic production of vesicular stomatitis virus caused by defective interfering particles. J. Infect. Dis. 126, 402–410.

Parmley, J.L., Chamary, J.V., Hurst, L.D., 2006. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol. Biol. Evol. 23, 301–309.

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

Perales, C., Moreno, E., Domingo, E., 2015. Clonality and intracellular polyploidy in virus evolution and pathogenesis. Proc. Natl. Acad. Sci. U.S.A. 112, 8887–8892.

Pérez-Losada, M., Arenas, M., Galán, J.C., Palero, F., González-Candelas, F., 2015. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 30, 296–307.

Pfaller, C.K., Donohue, R.C., Nersisyan, S., Brodsky, L., Cattaneo, R., 2018. Extensive editing of cellular and viral double-stranded RNA structures accounts for innate immunity suppression and the proviral activity of ADAR1^{p150}. PLoS Biol. 16, e2006577.

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

Pfeiffer, J.K., Kirkegaard, K., 2005. Ribavirin resistance in hepatitis C virus replicon-containing cell lines conferred by changes in the cell line or mutations in the replicon RNA. J. Virol. 79, 2346–2355.

Pfeiffer, J.K., Topping, R.S., Shin, N.H., Telesnitsky, A., 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription. J. Virol. 73, 8441–8447.

Plyusnin, A., Kukkonen, S.K., Plyusnina, A., Vapalahti, O., Vaheri, A., 2002. Transfection-mediated generation of functionally competent Tula hantavirus with recombinant S RNA segment. EMBO J. 21, 1497–1503.

Poirier, E.Z., Mounce, B.C., Rozen-Gagnon, K., Hooikaas, P.J., Stapleford, K.A., et al., 2015. Low-fidelity polymerases of alphaviruses recombine at higher rates to overproduce defective interfering particles. J. Virol. 90, 2446–2454.

Preston, B.D., Dougherty, J.P., 1996. Mechanisms of retroviral mutation. Trends Microbiol. 4, 16–21.

Ramirez, J.D., Llewellyn, M.S., 2014. Reproductive clonality in protozoan pathogens – truth or artefact? Mol. Ecol. 23, 4195–4202.

Rawson, J.M.O., Nicolaitchik, O.A., Keele, B.F., Pathak, V.K., Hu, W.S., 2018. Recombination is required for efficient HIV-1 replication and the maintenance of the viral genome integrity. Nucleic Acids Res. 46, 10535–10545.

Resch, A.M., Carmel, L., Marino-Ramirez, L., Ogurtsov, A.Y., Shabalina, S.A., et al., 2007. Widespread positive selection in synonymous sites of mammalian genes. Mol. Biol. Evol. 24, 1821–1831.

Rezelj, V.V., Levi, L.I., Vignuzzi, M., 2018. The defective component of viral populations. Curr. Opin. Virol. 33, 74–80.

Richmond, R.C., 1970. Non-Darwinian evolution: a critique. Nature 225, 1025–1028.

Rocha, E.P., 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14, 2279–2286.

Rosario, D., Perez, M., de la Torre, J.C., 2005. Functional characterization of the genomic promoter of borna disease virus (BDV): implications of 3′-terminal sequence heterogeneity for BDV persistence. J. Virol. 79, 6544–6550.

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

Rozen-Gagnon, K., Stapleford, K.A., Mongelli, V., Blanc, H., Failloux, A.B., et al., 2014. Alphavirus mutator variants present host-specific defects and attenuation in mammalian and insect models. PLoS Pathog. 10, e1003877.

Sabo, D.L., Domingo, E., Bandle, E.F., Flavell, R.A., Weissmann, C., 1977. A guanosine to adenosine transition in the 3′ terminal extracistronic region of bacteriophage Qβ RNA leading to loss of infectivity. J. Mol. Biol. 112, 235–252.

Sachs, J.L., Bull, J.J., 2005. Experimental evolution of conflict mediation between genomes. Proc. Natl. Acad. Sci. U.S.A. 102, 390–395.

Saira, K., Lin, X., DePasse, J.V., Halpin, R., Twaddle, A., et al., 2013. Sequence analysis of *in vivo* defective interfering-like RNA of influenza A H1N1 pandemic virus. J. Virol. 87, 8064–8074.

Salemi, M., Vandamme, A.M., 2004. The Phylogenetic Handbook. A Practical Approach to DNA and Protein Phylogeny. Cambridge University Press, Cambridge.

Sanjuan, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R., 2010. Viral mutation rates. J. Virol. 84, 9733–9748.

Santiago, M.L., Greene, W.C., 2008. The role of the APOBEC3 family of cytidine deaminase in innate immunity, G-to-A hypermutation, and evolution of retroviruses. In: Domingo, E., Webster, R.G., Holland, J.J. (Eds.), Origin and Evolution of Viruses, second ed. Elsevier, Oxford, pp. 183–206.

Sarisky, R.T., Nguyen, T.T., Duffy, K.E., Wittrock, R.J., Leary, J.J., 2000. Difference in incidence of spontaneous mutations between Herpes simplex virus types 1 and 2. Antimicrob. Agents Chemother. 44, 1524–1529.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., Malim, M.H., 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418, 646–650.

Shriner, D., Rodrigo, A.G., Nickle, D.C., Mullins, J.I., 2004. Pervasive genomic recombination of HIV-1 in vivo. Genetics 167, 1573–1583.

Sicard, A., Yvon, M., Timchenko, T., Gronenborn, B., Michalakis, Y., et al., 2013. Gene copy number is differentially regulated in a multipartite virus. Nat. Commun. 4, 2248.

Simmonds, P., 2010. Recombination in the evolution of picornaviruses. In: Ehrenfeld, E., Domingo, E., Roos, R.P. (Eds.), The Picornaviruses. ASM Press, Whashington, DC, pp. 229–237.

Smith, E.C., Denison, M.R., 2013. Coronaviruses as DNA wannabes: a new model for the regulation of RNA virus replication fidelity. PLoS Pathog. 9, e1003760.

Smith, D.B., Inglis, S.C., 1987. The mutation rate and variability of eukaryotic viruses: an analytical review. J. Gen. Virol. 68, 2729–2740.

Smith, P.J., Jones, C.J., 1999. DNA Recombination and Repair. Oxfird University Press, Oxford.

Smith, E.C., Blanc, H., Vignuzzi, M., Denison, M.R., 2013. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. PLoS Pathog. 9, e1003565.

Stadler, P.F., 2016. Evolution of RNA-based networks. Curr. Top. Microbiol. Immunol. 392, 43–59.

Stavrou, S., Crawford, D., Blouch, K., Browne, E.P., Kohli, R.M., et al., 2014. Different modes of retrovirus restriction by human APOBEC3A and APOBEC3G in vivo. PLoS Pathog. 10, e1004145.

Stedman, K., 2013. Mechanisms for RNA capture by ssDNA viruses: grand theft RNA. J. Mol. Evol. 76, 359–364.

Steinhauer, D.A., Holland, J.J., 1986. Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. J. Virol. 57, 219–228.

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

Steitz, T.A., 1999. DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. 274, 17395–17398.

Stratton, M.R., Campbell, P.J., Futreal, P.A., 2009. The cancer genome. Nature 458, 719–724.

Suarez, P., Valcárcel, J., Ortín, J., 1992. Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator mutants. J. Virol. 66, 2491–2494.

Supek, F., 2016. The code of silence: widespread associations between synonymous codon biases and gene function. J. Mol. Evol. 82, 65–73.

Suzuki, T., Moriyama, K., Otsuka, C., Loakes, D., Negishi, K., 2006. Template properties of mutagenic cytosine analogues in reverse transcription. Nucleic Acids Res. 34, 6438–6449.

Svarovskaia, E.S., Delviks, K.A., Hwang, C.K., Pathak, V.K., 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. J. Virol. 74, 7171–7178.

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

Szathmary, E., 1992. Viral sex, levels of selection, and the origin of life. J. Theor. Biol. 159, 99–109.

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

Tanaka, T., Eusebio-Cope, A., Sun, L., Suzuki, N., 2012. Mycoreovirus genome alterations: similarities to and differences from rearrangements reported for other reoviruses. Front. Microbiol. 3, 186.

Thomson, M.M., Perez-Alvarez, L., Najera, R., 2002. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. Lancet Infect. Dis. 2, 461–471.

Tibayrenc, M., Ayala, F.J., 2012. Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. Proc. Natl. Acad. Sci. U.S.A. 109, E3305–E3313.

Tibayrenc, M., Ayala, F.J., 2014. Cryptosporidium, Giardia, Cryptococcus, Pnemocystis genetic variability: cryptic biological species or clonal near-clades? PLoS Pathog. 10, e1003908.

Upton, D.C., Gregory, B.L., Arya, R., Unniraman, S., 2011. AID: a riddle wrapped in a mystery inside an enigma. Immunol. Res. 49, 14–24.

Urbanowicz, A., Alejska, M., Formanowicz, P., Blazewicz, J., Figlerowicz, M., et al., 2005. Homologous crossovers among molecules of brome mosaic bromovirus RNA1 or RNA2 segments in vivo. J. Virol. 79, 5732–5742.

Valcarcel, J., Ortin, J., 1989. Phenotypic hiding: the carryover of mutations in RNA viruses as shown by detection of *mar* mutants in influenza virus. J. Virol. 63, 4107–4109.

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

Varela-Echavarria, A., Garvey, N., Preston, B.D., Dougherty, J.P., 1992. Comparison of Moloney murine leukemia virus mutation rate with the fidelity of its reverse transcriptase in vitro. J. Biol. Chem. 267, 24681–24688.

Varsani, A., Lefeuvre, P., Roumagnac, P., Martin, D., 2018. Notes on recombination and reassortment in multipartite/segmented viruses. Curr. Opin. Virol. 33, 156–166.

Vartanian, J.P., Henry, M., Wain-Hobson, S., 2001. Simulating pseudogene evolution in vitro: determining the true number of mutations in a lineage. Proc. Natl. Acad. Sci. U.S.A. 98, 13172–13176.

Velicer, G.J., Raddatz, G., Keller, H., Deiss, S., Lanz, C., et al., 2006. Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. Proc. Natl. Acad. Sci. U.S.A. 103, 8107–8112.

Venkatesan, S., Rosenthal, R., Kanu, N., McGranahan, N., Barket, J., et al., 2018. Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. Ann. Oncol. 29, 563–572.

Vignuzzi, M., Andino, R., 2010. Biological implications of picornavirus fidelity mutants. In: Ehrenfeld, E., Domingo, E., Roos, R.P. (Eds.), The Picornaviruses. ASM Press, Washington DC, pp. 213–228.

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

Viguera, E., Canceill, D., Ehrlich, S.D., 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. 20, 2587–2595.

Vijg, J., Dong, X., Milholland, B., Zhang, L., 2017. Genome instability: a conserved mechanism of ageing? Essays Biochem. 61, 305–315.

Villarreal, L.P., 2005. Viruses and the Evolution of Life. ASM Press, Washington, D.C.

Satellites and defective viral RNAs. In: Vogt, P.K., Jackson, A.O. (Eds.), 1999. Current Topics in Microbiol. and Immunol., vol. 239. Springer, Berlin.

Wainberg, M.A., Drosopoulos, W.C., Salomon, H., Hsu, M., Borkow, G., et al., 1996. Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase. Science 271, 1282–1285.

Wang, Y., Schmitt, K., Guo, K., Santiago, M.L., Stephens, E.B., 2016. Role of the single deaminase domain APOBEC3A in virus restriction, retrotranscription, DNA damage and cancer. J. Gen. Virol. 97, 1–17.

Ward, C.D., Flanegan, J.B., 1992. Determination of the poliovirus RNA polymerase error frequency at eight sites in the viral genome. J. Virol. 66, 3784–3793.

Ward, C.D., Stokes, M.A., Flanegan, J.B., 1988. Direct measurement of the poliovirus RNA polymerase error frequency in vitro. J. Virol. 62, 558–562.

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

Webster, B., Ott, M., Greene, W.C., 2013. Evasion of superinfection exclusion and elimination of primary viral RNA by an adapted strain of hepatitis C virus. J. Virol. 87, 13354–13369.

Westrich, J.A., Warren, C.J., Klausner, M.J., Guo, K., Liu, C.W., et al., 2018. Human papilloma 16 E7 stabilizes APOBEC3A protein by inhibiting cullin 2-dependent protein degradation. J. Virol. 92 e01318-17.

Winters, M.A., Merigan, T.C., 2005. Insertions in the human immunodeficiency virus type 1 protease and reverse transcriptase genes: clinical impact and molecular mechanisms. Antimicrob. Agents Chemother. 49, 2575–2582.

Worobey, M., Holmes, E.C., 1999. Evolutionary aspects of recombination in RNA viruses. J. Gen. Virol. 80 (Pt 10), 2535–2543.

Wu, J., Gong, P., 2018. Visualizing the nucleotide addition cycle of viral RNA-dependent RNA polymerase. Viruses 10, E24.

Yamaguchi, T., Yamashita, Y., Kasamo, K., Inoue, M., Sakaoka, H., et al., 1998. Genomic heterogeneity maps to tandem repeat sequences in the herpes simplex virus type 2 UL region. Virus Res. 55, 221–231.

Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15, 496–503.

Zaher, H.S., Green, R., 2009. Fidelity at the molecular level: lessons from protein synthesis. Cell 136, 746–762.

Zimmern, D., 1988. Evolution of RNA viruses. In: Domingo, E., Holland, J.J., Ahlquist, P. (Eds.), RNA Genetics. CRC Press Inc, Boca Raton, FL, pp. 211–240.