Population Genomics of Human Viruses

Fernando González-Candelas, Juan Ángel Patiño-Galindo, and Carlos Valiente-Mullor

Abstract Viruses, and a few RNA viruses in particular, represent one of the greatest threats for human health. High mutation rates, large population sizes, and short generation times contribute to their typically fast evolutionary rates. However, many additional processes operate on their genomes, often in opposite directions, driving their evolution and allowing them to adapt to diverse host populations and antiviral drugs. Until recently, the high levels of genetic variation of most viruses have been explored only at a few genes or genome regions. The recent advent and increasing affordability of next-generation sequencing techniques have allowed obtaining complete genome sequences of large numbers of viruses, mainly HIV, HCV, influenza A, and others associated with emerging infections, such as Zika, chikungunya, or dengue virus. This opens the possibility to explore the effects of the different processes affecting viral diversity and evolution at the genome level. Consequently, population genomics provides the conceptual and empirical tools necessary to interpret genetic variation in viruses and its dynamics and drivers and to transform these results into information that may complement the epidemiological surveillance of the virus and its disease. This chapter provides an overview of human viruses from a population genomics perspective, with a special emphasis on RNA viruses, and the potential benefits of "genomic surveillance" to establish public health policies that improve the control and monitoring of the diseases caused by these viruses.

Keywords Complete genome · Epidemiology · Genetic variation · Mutation · Next-generation sequencing · Phylogeography · Reassortment · Recombination · Secondary structure

F. González-Candelas (✉), J. Á. Patiño-Galindo · C. Valiente-Mullor
Joint Research Unit "Infection and Public Health" FISABIO-Universitat de València,
Institute for Integrative Systems Biology, I2SysBio (CSIC-UV), Valencia, Spain

CIBER in Epidemiology and Public Health, Madrid, Spain
e-mail: fernando.gonzalez@uv.es

Martin F. Polz and Om P. Rajora (eds.), *Population Genomics: Microorganisms*,
Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_31,

1 Introduction

The development of fast and efficient sequencing methodologies has brought the opportunity for obtaining complete sequences of hundreds, even thousands, of viral genomes at affordable costs. This has led to a new interest in the analysis of viral populations, which, until recently, was usually linked to outbreaks and other health emergencies. Most previous studies paid attention only to those fragments of the viral genome that were of interest from a clinical perspective, for diagnostics, surveillance, or similar applications. Furthermore, most insights into the population genetics of viral populations were drawn from markers likely under the influence of selective forces, thus leading to distorted or biased views of viral population genetics. This situation is rapidly changing, and the availability of complete genome sequences is shifting the perspective from "population genetics" to "population genomics," that is, the analysis of the processes and mechanisms that govern the population dynamics of genetic variation at the complete genome level and not only on a portion of it.

Although information on complete genomes is rapidly accumulating, there is still a huge gap between the number and diversity of viral population samples that have been analyzed in only one or a few genes and those with complete genome information. However, there is a shift of interest in using population genomics inference to better understand the intra-host and inter-host dynamics of epidemiologically and evolutionary relevant processes and to incorporate this information into surveillance systems. This shift has also benefited from recent methodological and technical advances, which have allowed the combination of different sources of information (temporal, geographical, genetic, and epidemiological) into a comprehensive framework, known as "genomic surveillance." Here, we review the current state of the art in the population genomics of human viruses and its relevance for the surveillance, monitoring, and control of the diseases they cause. Because of their rapid rate of evolution and the serious diseases they produce – AIDS, hepatitis C, Ebola, influenza, among many others – RNA viruses have received most attention until now and abundant information on their population genomics is accumulating. We will center this review on these viruses.

2 Evolutionary Processes in Viral Populations

Mutation is the ultimate source of variation in all living organisms, viruses included. However, the genetic diversity and the evolutionary rate of RNA viruses are influenced and shaped by other processes and factors apart from mutation. The action of natural selection and genetic drift, the mode of transmission, particular mechanisms for genetic exchange (such as recombination and reassortment), genome size, procedures for compressing genetic information, generation time, and population size are the most relevant such factors. In addition, we must also consider environmental factors, resulting from differences among hosts and, occasionally, from antiviral treatments (Cuypers et al. 2016; Rambaut et al. 2008; Renzette et al. 2014; Simon-Loriere et al. 2013; Snoeck et al. 2011; Wilson et al. 2016).

Deciphering the mechanisms responsible for the production of spontaneous mutations in viruses has important applications for public health and for basic science (Geller et al. 2016) due to their critical role in virus evolution and genetic diversity (Cuevas et al. 2015). One defining feature of RNA viruses is their high mutation rates, in the range from 10^{-3} to 10^{-6} mutations/nucleotide/replication round, which result from low-fidelity replication (Simon-Loriere et al. 2013; Cuevas et al. 2015; Duffy et al. 2008; Sanjuan et al. 2010). This also leads to a very high evolutionary rate of 10^{-2} to 10^{-5} substitutions/site/year. These high mutation rates can be decomposed into several factors or mechanisms with complex interactions such as the fidelity of the RNA polymerase, the capacity for error correction, the propensity of ribonucleic acid to damage, or the edition by hosts' enzymes (Geller et al. 2016; Cuevas et al. 2015). These high mutation rates might explain the small genome size of RNA viruses (ranging from 3 to 29 kb) because in larger genomes, deleterious mutations would appear at such a high frequency that they would compromise virus survival (Duffy et al. 2008; Bradwell et al. 2013).

In general, RNA viruses have short generation times and large population sizes. These features favor fast evolutionary rates, which lead to genetically very diverse populations, with a high capacity for adaptation even under very strong selective pressures (Wilson et al. 2016). However, we must consider that virus evolutionary rates are limited by the frequency of deleterious mutations since virus mutation rates are very close to the error threshold beyond which deleterious mutations are so frequent that they lead to population extinction (Holmes 2003). In addition, some mechanisms for genetic exchange, which are present in some RNA viruses, favor the generation and maintenance of diversity. Two such mechanisms are genetic reassortment and recombination (homologous and nonhomologous).

Genetic reassortment occurs in segmented viruses, whose genome is distributed in individual segments, each carrying a different portion of the genetic information. Reassortment plays a major role at the epidemiological level in the evolution of influenza A virus (Rambaut et al. 2008; Wilson et al. 2016; Steel and Lowen 2014) and in other segmented viruses (McDonald et al. 2016; Nomikou et al. 2015). Recombination is also a frequent process in many viruses. In retroviruses, such as HIV, a nonhomologous type of recombination known as copy-choice recombination is common, and it can occur when two different viral strains simultaneously infect the same cell. In this form of recombination, the RNA polymerase "jumps" between two copies of single-stranded RNA, which makes up the genome of retroviruses, while it is still attached to the newly synthesized chain. This mechanism occurs only during RNA synthesis and the parental (donor) strand is not physically transferred to the recombined strand. It is likely that secondary structures of the RNA genomes are involved in controlling the "jump" between strands (Lai 1992; Negroni and Buc 2000; Simon-Loriere and Holmes 2011).

Natural selection may deplete genetic diversity from viral populations (negative or purifying selection) or increase its levels (some forms of positive selection) and, consequently, may increase or decrease the rate of evolution. Therefore, there is a trade-off between the conservation of those regions that are essential for completing the viral cycle of replication and the genetic change and innovation that are involved in evading the immune system and responding to antiviral treatments. The former group includes genes encoding for slowly evolving enzymes and structural proteins as well as genome regions involved in the formation of secondary structures. This compromise is partially achieved through differential mutation rates along the viral genome (Geller et al. 2015, 2016).

Because of their high mutation rates, RNA viruses are under selection for small genome size. This is due to the deleterious effect on fitness of most mutations, which lead to an excessive genetic load in large genomes, which, in turn, leads to population extinction (Muller 1932). A small genome size represents a limitation for the generation of genetic diversity because (1) sequence lengths are limited and (2) using gene overlapping to compress genetic information implies an increase in the sensitivity to deleterious mutations in certain parts of the genome and, consequently, a larger role for purifying selection (Simon-Loriere et al. 2013). Although gene overlapping is present in all cellular organisms, mammals included (Veeramachaneni et al. 2004), it is very frequent only in viruses (Rogozin et al. 2002; Brandes and Linial 2016).

Another factor limiting the rate of evolution is the transmission between hosts. Each of these events represents a bottleneck that dramatically reduces the size of the viral population and, as a result, its genetic diversity (Gray et al. 2011; Grenfell et al. 2004; Joseph et al. 2015). Besides, founder viruses will generally be poorly adapted to the new environment because, in general, the specific adaptations to the immune system of the donor/source individual do not imply a higher fitness in another individual of the same species and they can even be penalized by natural selection (Kubinak et al. 2012).

Therefore, viruses mutate and may evolve very fast. It is crucial to understand the mechanisms by which they generate and maintain genetic diversity for the application of research on these organisms to health-related questions, such as the evasion of immune response, the development and spread of drug resistance mutations, virulence, species jumps, and the failure or success of vaccination campaigns (Wilson et al. 2016; Geller et al. 2016; Smyth et al. 2012). In this context, the rate of mutation should be considered not only as a mechanism generating diversity but also as a virulence factor (Cuevas et al. 2015).

3 Selective Pressures

In viruses, as in all living organisms, natural selection operates as a force that, on the one hand, may reduce genetic variability and, on the other hand, may increase genetic diversity (Snoeck et al. 2011). Hence, we start by describing the different types of natural selection that operate in viral populations.

Positive selection promotes an increase of the relative frequency of an allele or genetic variant in a population. Positively selected mutations confer higher fitness to their carriers, resulting in an increased frequency of the corresponding allele. Two paradigmatic examples of positive selection are immune escape mutations and drug resistance mutations.

Positive selection may act as an evolutionary force that restricts the genetic diversity of a population (directional selection) or as a force that promotes an increase of genetic diversity (diversifying selection). Directional selection is commonly associated with selective sweeps. During selective sweeps, neutral or nearly neutral mutations increase their relative frequencies, even become fixed, in the population due to genetic linkage with positively selected variants (Maynard-Smith and Haigh 1974). The strength and scope of selective sweeps (which may act at the genome-wide level) (Rambaut et al. 2008) will also depend on the rate of recombination. For example, as detailed above, high recombination rates limit the scope of selective sweeps in HIV (Ramirez et al. 2008; Vuilleumier and Bonhoeffer 2015; Zanini et al. 2015). Additionally, the rate of emergence of adaptive mutations influences the intensity of selective sweeps.

When different adaptive mutations, which have either newly arisen or were previously present at low frequencies in a population, are selected simultaneously or nearly simultaneously, then soft sweeps will be produced as several mutations – located in different regions of the viral genome – propagate jointly. Such soft sweeps may result in clonal interference, which consists of competition between distinct lineages in the viral population carrying different adaptive mutations. Consequently, even mutations favored by natural selection may not be fixed in the population or, alternatively, they may get fixed at lower rates. Thus, clonal interference may slow down adaptation in a viral population (Miralles et al. 1999).

Soft selective sweeps have a minor impact on the loss of population genetic diversity. However, if adaptive mutations arise rarely in a population, then a single variant will increase its frequency along with its genetically linked neutral alleles. Consequently, a hard selective sweep will be produced, which implies a huge decrease in genomic diversity (Feder et al. 2016; Hermisson and Pennings 2005; Messer and Petrov 2013; Pennings et al. 2014).

The rate at which adaptive mutations emerge and become positively selected depends on the mutation rate, the population size, with small sizes resulting in strong genetic drift and reduced efficiency of natural selection, and the strength of the selective pressure. This complex process can be studied in individuals under antiviral drug therapy. Highly efficient drug treatments, consisting of a combination of antiviral drugs, reduce viral population size and the frequencies of drug-resistant alleles. Moreover, the number of permissive mutations needed for acquiring drug resistance (genetic barrier) may increase (Feder et al. 2016).

Diversifying selection occurs when selection favors different adaptive mutations over time and/or space, and it results in an increase in genetic diversity. Generally, we can observe this type of selection in viral responses to the hosts' immune systems. As a result, the genome regions coding for proteins targeted by the immune response (antigens) present much higher variability than the remainder of the genome. Antigenic drift – antigenic evolution in influenza A virus – exemplifies this phenomenon and should not be confused with genetic drift. Due to the interaction between influenza A virus and the human immune system, mutations accumulate in antigenic regions encoding surface proteins such as hemagglutinin and neuraminidase. Influenza A viruses show episodic selection, that is, positive selection over long periods interspersed with purifying selection over short time periods. This process might explain the new seasonal antigenic variants of influenza A virus (Rambaut et al. 2008; Cobey and Koelle 2008; McHardy and Adams 2009). However, it must be noted that the two sides of positive selection are linked: antigenic drift is inevitably related to periodic selective sweeps (McHardy and Adams 2009).

Negative (or purifying) selection operates by removing deleterious alleles (i.e., mutations that decrease viral fitness). Negative effects of deleterious mutations can involve a reduction in the replication rate or increased susceptibility to the host immune response or to antiviral drugs. Most mutations arising in living organisms are deleterious. For instance, nearly 60% of the spontaneous mutations in vesicular stomatitis virus are deleterious (Duffy et al. 2008). Thus, purifying selection constitutes a force acting to preserve nucleotide or amino acid sequence. Therefore, negative selection constrains genetic diversity.

Purifying selection can be prominent in the viral genome, even in viruses, such as HIV, in which positive selection and neutral evolution have an important role at the intra-host level (Snoeck et al. 2011; Zanini et al. 2015; Pybus and Rambaut 2009; Ross and Rodrigo 2002). HCV is a clear example of predominance of negative selection. Despite the high levels of genetic variability in this virus, negative selection represents the main force acting on the HCV genome: more than 80% of the nucleotide sites in the viral genome are under negative selective pressure (Cuypers et al. 2016; Geller et al. 2016; Patiño Galindo and González-Candelas 2017).

Natural selection can be studied by comparing the synonymous substitution rate per synonymous site (dS) and the non-synonymous substitution rate per non-synonymous site (dN). The ratio of both rates (ω = dN/dS) allows different types of selection to be distinguished throughout the viral genome. Under neutral evolution, all mutations are expected to have the same effect (i.e., none or negligible) on fitness, and, thus, ω will be around 1. Negative or purifying selection reduces dN – because non-synonymous substitutions lead to changes in the amino acid sequence and thus protein structure or function will likely be affected – whereas dS should not be affected. Therefore, the ratio ω will be lower than 1. In contrast, positive selection favors non-synonymous substitutions over synonymous substitutions, and, therefore, ω will be larger than 1 (Cobey and Koelle 2008; Jackowiak et al. 2014).

When interpreting the results of analyses based on this popular method for analyzing selection at the genome level, several caveats have to be considered. Firstly, the method was originally proposed to analyze selection acting over evolutionary large time scales, because it makes use of the rates of substitution, which implies the replacement and fixation of mutations in populations/species. This is not usually the case in viral populations, where we are dealing with constantly arising polymorphisms that, even when they are deleterious, will segregate in the population before selection removes them. This effect can be controlled for by considering only those mutations that can be mapped onto the internal branches of the phylogeny whereas those at the external branches are excluded from the analyses. Secondly, these tests can be misleading if recombination occurs frequently (Anisimova et al. 2003), because it may alter the estimates of dN and dS, thus leading to incorrect estimates of ω.

4 How Selective Pressures Operate on Viral Genomes

Viruses are subjected to different types of selective pressures that drive their evolution and shape their genome diversity. Distinct selective pressures can increase or constrain genome variability. These selective pressures and evolutionary trade-offs drive virus evolution. They include interactions with the host's immune system as well as the need for immune escape, the pressures exerted by antiviral drug therapies, the trade-off between high viral mutation rates and genome size, the maintenance of protein structure and function, the maintenance of RNA secondary structures, and the presence of epistatic interactions between different parts of the genome. It is necessary to take into account these, sometimes opposite, forces for understanding viral genome evolution (Snoeck et al. 2011).

4.1 Interaction with the Host Immune System

Some of the most prevalent infectious diseases are caused by RNA viruses due to their high capacity for escaping their hosts' immune system by rapid antigenic evolution (Cobey and Koelle 2008). Viruses, as well as parasites, are involved in a constant "arms race" with their hosts. The former evolve to evade the immune system of the latter, while hosts' immune systems evolve to detect, control, and efficiently eliminate pathogens [the "Red Queen hypothesis" illustrates this situation (Van Valen 1973)]. The strong selective pressures exerted by the hosts' immune systems on viruses, along with their high genomic variability, result in rapid adaptation and constant evolution in coding genome regions involved in interaction with the hosts (Snoeck et al. 2011; Duffy et al. 2008; Kubinak et al. 2012; Jackowiak et al. 2014; Alizon and Fraser 2013). Thus, mutations that allow evading the immune system usually propagate rapidly through viral populations (Zanini et al. 2015).

Genome regions or segments involved in immune escape show high evolutionary rates due to positive selective pressures exerted by the hosts (Rambaut et al. 2008). Generally, these regions encode surface or viral envelope proteins. Therefore, these proteins act as targets for viral recognition by the host's immune system. Examples include the *env* region in HIV (Cobey and Koelle 2008; Alizon and Fraser 2013), the E1 and E2 genes in HCV (Thurner et al. 2004; Campo et al. 2008), and the hemagglutinin and neuraminidase segments in influenza A virus (Rambaut et al. 2008; Cobey and Koelle 2008; Pybus and Rambaut 2009; Neverov et al. 2015).

There are three types of canonical viral targets and, consequently, “hot spots” for positive selection. These are targets of neutralizing antibodies, CD4 T-cell and CD8 T-cell epitopes (i.e., regions of viral antigen recognized by molecules of the host immune system) (Zanini et al. 2015; Jackowiak et al. 2014). However, their relevance has been questioned. For example, CD4 T-cell epitopes seem to be conserved (i.e., under negative selection) in HCV, whereas CD8 T-cell epitopes are under positive selection and, consequently, drive immune evasion in this virus (Cuypers et al. 2016; Patiño Galindo and González-Candelas 2017). Another example is represented by the mapping of positively selected sites in the HIV genome and by considering different likely targets of selection, such as epitopes recognized by immune system cells, secondary structure of protein and nucleic acids, and particular dinucleotides targeted by antiviral proteins such as APOBEC3G/F (Snoeck et al. 2011). Antibody and CD4 T-cell epitopes were found to be under positive selection. However, no positive selection was detected on CD8 T-cell epitopes. Although this observation may suggest an absence of host selective pressures acting on CD8 T-cell epitopes, the authors suggest other explanations. On the one hand, positively selected escape variants without deleterious effects will fix rapidly in the viral population, thus becoming relatively conserved. On the other hand, T-cell epitopes could be under opposite selective pressures over chronic infection.

HCV is a good example of changing host selective pressures through chronic infection at the intra-host level. HCV populations progress through different stages. Firstly, right after infection, the viral population establishes under relaxed selective pressures, before triggering the immune response. As the viral population size increases, the immune response activates. Consequently, population diversity also increases, whereas escape variants appear and become fixed under positive selection. In the last stage, purifying selection predominates. This suggests that the virus has adapted steadily to its host (Jackowiak et al. 2014).

4.2 Antiviral Drug Therapies

The evolution of pathogenic microorganisms – including viruses – and the emergence of drug resistances are major concerns for public health. Drug resistance is usually related to treatment failure and results in increasing deaths, hospitalizations, and treatment duration as well as huge economic costs (Wilson et al. 2016; McGowan 2001; WHO Scientific Working Group 1983).

Some features of RNA viruses, as with immune escape, allow them to adapt rapidly in response to the strong selective pressures exerted by antiviral treatments. These features (including high mutation rates, large population sizes, and recombination or reassortment) facilitate the emergence of de novo resistance mutations. In the absence of drug-selective pressures, resistance mutations may be deleterious or, occasionally, neutral, which implies that their evolution will be governed mainly by genetic drift. For this reason, in the absence of treatment, drug-resistant variants are usually found as minority variants that increase their relative frequency in the population only in the presence of antiviral drugs. Hence, the possibility of transmission of resistance mutations between hosts must be taken into account in order to predict the effectiveness of a particular antiviral therapy. Next-generation sequencing is necessary to detect resistance variants at low frequencies prior to the start of treatment. The development of drug resistance may depend on the presence of various permissive mutations in the same haplotype in order to decrease the genetic barrier (Wilson et al. 2016; Pybus and Rambaut 2009; Chabria et al. 2014).

It is expected that strong and directional positive selection, which is restricted to periods of time when a patient is undergoing antiviral treatment, will increase the relative frequency of resistance alleles, whereas the genetic variability of those regions close to selected loci will decrease due to selective sweeps (Renzette et al. 2014; Murrell et al. 2012). The evolution of HIV since the introduction of early antiretroviral therapies is a good example of this process. Modern treatments – highly active antiretroviral therapy (HAART) – are more effective than single drug-based early therapies. HAART consists in a customized combination of drugs. Therefore, several resistance mutations are necessary to develop simultaneous resistance against every drug included in the treatment. In contrast, early, single drug-based therapies were prone to the rapid emergence of drug resistance (Smyth et al. 2012; Martin et al. 2008). Due to the high efficiency of treatments consisting of different drugs, resistance mutations are uncommon and emerge rarely. Thus, positive selection results in strong selective sweeps that reduce genetic diversity and slow down virus evolution (Feder et al. 2016). The opposite situation was found in influenza A virus resistance to oseltamivir. One of several resistance mutations to oseltamivir (H274Y) underwent rapid and global spread during the influenza seasons between 2007 and 2009. However, the rapid increase in H274Y frequency did not substantially alter the viral genomic diversity. It is perhaps a consequence of emergence of different mutations conferring resistance to oseltamivir (Renzette et al. 2014).

4.3 Secondary RNA Structures: Protein Structure and Function

The presence of structural elements at the nucleotide and amino acid levels is of major significance for viral genome evolution because they contribute to increasing genome stability, controlling viral replication, and avoiding genome recognition by RNAses and innate antiviral defenses (Baird et al. 2006; Watts et al. 2009). Structural elements are often highly conserved. Mutations that disrupt RNA secondary structures or protein domains may have strong deleterious effects (Thurner et al. 2004; Simmonds et al. 2004).

Coding regions are under strong purifying selection, and, therefore, they are highly conserved at the amino acid level, particularly those involved in the maintenance of protein secondary structure and function (Snoeck et al. 2011). This is true for genes or segments that code for RNA polymerase in different viruses (Rambaut et al. 2008; Zanini et al. 2015; Rothenberger et al. 2016).

RNA secondary structures are frequent in viral genomes, particularly in those of single-stranded RNA viruses. RNA secondary structures may be relevant for replication and transmission of the virus as well as for drug resistance and host interaction (Cuypers et al. 2016; Simon-Loriere et al. 2013; Thurner et al. 2004; Simmonds et al. 2004; Sanjuán and Bordería 2011). In this case, purifying selection operates at the nucleotide sequence level. As nucleotide changes driven by positive selection might disrupt RNA secondary structures, this will result in conflict between purifying selection and positive selection acting on coding regions (Snoeck et al. 2011; Sanjuán and Bordería 2011). In other words, the maintenance of RNA secondary structures may restrict protein evolution, and, in turn, selection at the protein level may restrict the pairing of nucleotides that maintain RNA secondary structures. The disruption of RNA secondary structures produced by amino acid changes could explain the fitness decrease in drug-resistant viruses in the absence of selective pressure by antiviral therapies (Sanjuán and Bordería 2011).

The case of HIV illustrates this situation. Although HIV evolution is largely driven by positive selection, more than 60% of its amino acid sites are strongly conserved. RNA secondary structures and α-helix domains mainly determine conservation in the HIV genome (Snoeck et al. 2011).

4.4 Genome Size and Gene Overlapping

Due to their high mutation rates, RNA viruses are under selective pressures favoring small genome sizes. Because most spontaneous mutations are deleterious, high mutation rates in large genomes result in excessive mutational load that may lead to population extinction. More deleterious and even lethal mutations emerge in large genomes per replication cycle than in small genomes although mutation rates can be similar. Moreover, a trend toward small genome size may also be influenced by the rate of replication, because selection favoring rapid replication will, in turn, favor viruses with minimal genome sizes (Simon-Loriere et al. 2013; Duffy et al. 2008; Bradwell et al. 2013).

Small genome size implies two problems for viral evolution: firstly, the need for storing all the genetic information in a limited space and, secondly, the need for generating genetic novelty while maintaining a small genome size. Consequently, RNA viruses often use gene overlapping in order to compress genetic information and avoid the aforementioned problems without increasing their genome size. However, gene overlapping leads to hypersensitivity to deleterious mutations (i.e., an increase in the deleterious effects of mutations in overlapping genome regions) as they affect more than one gene. Therefore, strong purifying selection operates in these regions, resulting in a reduced evolutionary rate and adaptation in RNA viruses. Despite this, the negative effects of gene overlapping on evolutionary rate depend on the type of overlapping where internal overlapping (i.e., a single gene that contains another gene within its nucleotide sequence) is associated with stronger negative selection (Simon-Loriere et al. 2013).

In conclusion, small genome sizes limit the generation of genetic diversity as the nucleotide sequence space is limited and the use of gene overlapping as a mechanism of genome compression leads to hypersensitivity to deleterious mutations in certain regions of the genome, thus resulting in stronger purifying selection.

4.5 Epistasis

Epistasis has been described as an evolutionary phenomenon in which the fitness of a mutation depends on its genetic background (Phillips 2008). In other words, different loci along the viral genome interact with each other and determine fitness. Consequently, the phenotypic effects of a mutation may change in the presence or absence of certain genetic elements. Therefore, epistasis can significantly influence how certain mutations navigate the adaptive landscape (Wilson et al. 2016; Cobey and Koelle 2008; Assis 2014). Epistasis can be relevant in the fitness effects of RNA secondary structures, drug resistance mutations, and recombination or reassortment events. Thus, epistasis must be taken into account in order to predict the success of mutations in a viral population.

A simple form of epistasis occurs in the secondary structures of RNA viruses. The maintenance of these structures depends on base pairing between sites located on a single-stranded RNA genome. Nucleotide pairings usually follow the classical Watson-Crick model (guanine-cytosine [G-C] and adenine-uracil [A-U]). As expected, any mutation disrupting Watson-Crick pairs will alter highly conserved RNA secondary structures. Thus, they are often deleterious, and we expect that strong purifying selection operates on Watson-Crick sites, resulting in a reduced rate of evolution. This pattern has been observed in HIV, HCV, and influenza A virus. However, G-U pairs are also stable and they can maintain RNA structures. Although G-U pairs usually show fewer effects on fitness than Watson-Crick pairs, the fitness difference is relatively small. G-U pairs can operate as intermediates between adaptive peaks (i.e., G-C and A-U pairs), thus relaxing negative selective pressures on Watson-Crick sites. Moreover, G-U can remain in the population because, after all, G-U pairs show higher fitness than unpaired nucleotides (Assis 2014).

Epistasis is also relevant for the emergence of drug resistance. The fate of new drug resistance mutations depends on their efficiency in avoiding antiviral drugs effects and on their deleterious effects, mainly on viral replication. However, a permissive mutation can interact epistatically with drug resistance mutations in order to increase their fitness and, therefore, their relative frequencies in the viral population (Wilson et al. 2016; Chabria et al. 2014). The emergence of oseltamivir resistance in influenza A virus during the influenza seasons of 2007–2009 illustrates this phenomenon (see Sect. 4.2). Highly deleterious effects were predicted for the H274Y drug resistance mutation. However, H274Y spread rapidly and globally, thanks to two permissive mutations that made the mutant fitness equal to that of the non-mutated genotype in the absence of oseltamivir (Neverov et al. 2015; Duan et al. 2014; Kryazhimskiy et al. 2011).

Influenza A virus can also be used as an example to highlight the relevance of genetic background for genetic exchange between different strains. Most segment combinations resulting from genetic reassortment are probably deleterious due to epistatic interactions (Rambaut et al. 2008; Renzette et al. 2014; Sobel Leonard et al. 2017).

In conclusion, epistatic interactions must be taken into account in order to predict virus evolution and, specifically, the epidemiological consequences of drug resistance mutations. Complete genome sequencing can be used in this context to detect epistatic interactions between distant genome regions (Rambaut et al. 2008; Wilson et al. 2016).

5 Mutation Rate and Natural Selection

Mutation is a key factor in the generation of genetic variability. In addition, the rate of mutation is a viral character evolving under natural selection. Natural selection favors high mutation rates in viruses as they increase their adaptive capacity, particularly regarding infection, host adaptation, and immune escape. In this light, viral mutation rates might be considered a virulence factor. The presence of local RNA secondary structures in the viral genome may operate as a mechanism of modulation for genome variability. RNA secondary structures flank hypervariable regions, which are prone to low-fidelity replication because they are usually located in single-stranded segments, thus focusing higher mutation rates in genomic regions involved in immune escape (Geller et al. 2016; Cuevas et al. 2015; Duffy et al. 2008; Sanjuán and Bordería 2011).

However, variability in viral genomes has an upper limit. Mutation rates are often close to the error threshold. Beyond the error threshold, deleterious mutations emerge too frequently, resulting in population extinction (error catastrophe). Therefore, purifying selection purges variants exceeding certain mutation rates. In this context, it must be noted that viral genome hypermutation exerted by host deaminases constitutes a potential mechanism against viral infection. This is apparently the case in HIV infection (Snoeck et al. 2011; Cuevas et al. 2015; Duffy et al. 2008; Holmes 2003; Neogi et al. 2013; Noguera-Julian et al. 2016).

6 Within and Among Patient Diversification

The evolutionary dynamics of genetic diversity in RNA viruses can differ markedly between levels of biological organization, within individuals (intra-host), and at the epidemiological level (inter-hosts). This prominent feature has been analyzed in depth in some viruses that produce chronic or persistent infections, such as HIV or hepatitis C virus (HCV). However, it is also possible to analyze the genetic changes at the intra- and inter-hosts levels in viruses that produce acute infections, such as influenza A virus. Viral evolution during chronic infection occurs simultaneously in different parts of the genome and, depending on the virus, independently in the segments. Hence, it is important to analyze genetic diversity in complete genomes, because different genome regions can be under distinct, even opposed, selective pressures (Pybus and Rambaut 2009; Holmes 2004; Luciani and Alizon 2009; Lythgoe and Fraser 2012; Sobel Leonard et al. 2016).

Viral infections usually start by a founder virus or a population of a few viral units with very similar genomes (Joseph et al. 2015; Jackowiak et al. 2014; Sobel Leonard et al. 2016). It is unlikely that there is only one genome sequence in the founder population shared by all the viruses. However, among the many variants present in the source individual, the fittest phenotypes for transmission will be more represented in the infecting population. Shortly after the infection, the process known as clonal expansion starts. This process results from the rapid replication of the virus that leads to an increasingly diverse population in which new mutations accumulate from the initial sequence. This genetically diverse population is usually known as a viral quasispecies (Eigen 1996), a set of highly diverse, evolutionarily close, nonidentical haplotypes (because they derive from the same virus or a reduced population) undergoing diversification, competition, and selection (Chabria et al. 2014; Domingo et al. 2012; Khiabanian et al. 2014). In later stages of infection, the initially homogeneous viral population will be more diverse. This indicates that, during transmission, there are several bottlenecks that reduce diversity at the inter-host level (Gray et al. 2011; Joseph et al. 2015).

Many pathogens produce chronic infections that evolve so rapidly that late variants in the infection are very different from the genetic variants in the founders (Luciani and Alizon 2009; Vrancken et al. 2015). During the early stages of chronic infection by RNA viruses, such as HIV, mutations that contribute to evade the host's immune system may appear and increase in frequency (Goonetilleke et al. 2009; Kearney et al. 2009; Liu et al. 2011). Hence, chronically infecting viral populations become adapted to their hosts and this may compromise their capacity for transmission (Wright et al. 2010; Brockman et al. 2010).

During infection, viral populations explore the adaptive landscape – the set of variants close to a given genotype that might increase the fitness of the population – around the founder virus. This is supported by the fact that the same reversions are observed in unrelated individuals. In HIV, some nucleotide substitutions produced during intra-host evolution are reversions to that global consensus sequence (Zanini et al. 2015; Li et al. 2007). This trend suggests that in chronic infections, directional natural selection is the main evolutionary force determining the diversity of the viral population. But most mutations are neutral or reduce rather than increase fitness. Nevertheless, in populations with high recombination rates, such as in HIV, adaptation to the host may be concurrent with a sustained exploration of the adaptive landscape. This trend is more evident for globally conserved genome positions, and it can also be observed in viruses producing acute infections (Zanini et al. 2015; Sobel Leonard et al. 2016; Wang et al. 2014; Gire et al. 2014). However, the number of positions under directional positive selection in the HIV genome is limited. Most of the genome is under purifying selection or accumulates neutral mutations. The action of diversifying selection, which acts in an opposite sense to directional selection, and the emergence of neutral mutations may disguise the convergence toward a global consensus sequence in positively selected positions (Snoeck et al. 2011; Ross and Rodrigo 2002).

Selective pressures acting on a viral population can differ intra- or inter-host and can often have opposing effects, leading to a trade-off. At the intra-host level, natural selection favors fast replicating variants, those that can evade the immune response, and, if the patient is being treated, those with resistance mutations against the corresponding drugs. At the inter-host level, natural selection will favor variants that can propagate rapidly in the host population, that is, those that are more easily transmitted from one host to another (Alizon and Fraser 2013).

One of the most remarkable differences between intra- and inter-host dynamics is the faster evolutionary rate associated with intra-host differentiation compared to the inter-host rate of evolution (Alizon and Fraser 2013; Lythgoe and Fraser 2012; Khiabanian et al. 2014). Intra-host evolutionary rates can be from two to six times higher than those among hosts (Lythgoe and Fraser 2012). Viral evolutionary rates show a trend to slow down in the long term. This trend is reinforced by the bottlenecks and selective pressures operating at transmission events (Zanini et al. 2015). Due to their dependence on infecting other hosts, inter-host evolutionary rates are also dependent on the transmission rate (Gray et al. 2011).

The difference in intra- and inter-host evolutionary rates means that, in chronic infections, viral populations are not homogeneous in their capacity for transmission to another host. If this were the case, we would not observe such different values between the corresponding rates (Alizon and Fraser 2013; Lythgoe and Fraser 2012). To explain this difference, we should also take into account that the viral population needs to adapt to the immune system of a specific host after each transmission. Therefore, intra-host evolution is governed by strong, continuous selective pressures leading to fast evolutionary dynamics with high evolutionary rates. Furthermore, the heterogeneity of the viral population and the different lineages that can coinfect an individual may affect the action of the immune system and, in consequence, the viral evolutionary dynamics (Grenfell et al. 2004).

However, although the intra-host rate of evolution is generally higher throughout the genome of these viruses, the pattern of evolution and the intra- and inter-host differences vary among genomic regions (Alizon and Fraser 2013). In some viruses, different genome regions can evolve independently due to recombination, such as in HIV (Zanini et al. 2015), thus minimizing the effect of selective sweeps (see below). For instance, some genes encoding for viral proteins targeted by the immune response show a faster intra-host evolution, with high levels of positive selection as a consequence of the selective pressures by the host's immune system (Gray et al. 2011; Sobel Leonard et al. 2016).

The reasons for the differences between intra- and inter-host evolutionary rates are not fully understood. Among potential alternatives, we can mention the following: (a) preferential transmission of slow-evolving lineages, (b) reduced intra-host rate of evolution over time, (c) reversion to genotypes similar to the founder virus that are likely better adapted to infecting other hosts, and (d) changes in selective pressures over the course of infection (Gray et al. 2011; Pybus and Rambaut 2009; Lythgoe and Fraser 2012). In HCV, it has been shown that the large differences between intra- and inter-host evolutionary rates in genome regions related to evasion from the immune system can be explained by reversions of host-specific adaptations to genotypes similar to those of the founder virus. The hypothesis of a preferential transmission of slow-evolving lineages seems to be quite unlikely, at least for HCV (Gray et al. 2011). In other viruses, such as HIV, the contribution of reversions to evolution has not been studied in detail (Zanini et al. 2015). Another contributing factor is that inter-host evolution is shaped by many bottlenecks produced in every transmission event (Gray et al. 2011; Joseph et al. 2015), which act reducing the evolutionary rate. As a consequence, phylogenies including isolates serially sampled within patients usually present long external but short internal branches, the latter corresponding to evolutionary changes occurring among patients.

7 Conflict Between Selective Pressures Within and Among Hosts

Intra- and inter-host selective pressures can be in conflict because mutations favoring adaptations to exploit the host, that is, those that are favored at the intra-host level (including immune system evasion and resistance mutations), are unlikely to also increase transmissibility to other hosts. Consequently, such mutations will be neutral, or selection at the inter-host level may act against them. The viral population evolves at the intra-host level during infection, becoming adapted to each new host. However, genotypes carrying host-specific adaptations do not seem to be the most efficient in being transmitted to new hosts (Alizon and Fraser 2013). The study of this conflict, known as “short-sighted evolution,” was initiated in the last decade of the past century and applied to different pathogens (Levin and Bull 1994). Although until recently this conflict had been studied at the genomic scale only in HIV, its presence in other viruses such as HCV or Marburg virus has led to question whether this is a common feature of RNA viruses (Gray et al. 2011). Would it be possible then that less fit variants, presumably purged by natural selection or belonging to minority classes, persist and be transmitted in a population?

Several mechanisms have been proposed to explain the transmission of those less fit variants (intra-host) to new hosts. For instance, HIV populations “archive” resistance variants in latent T-cells, which act as reservoirs of variants that can be transmitted later. Alternatively, mutations reverting to the founder virus, the one initially infecting the host and presumably fitter for transmission (Joseph et al. 2015;

Zanini et al. 2015; Jackowiak et al. 2014; Alizon and Fraser 2013; Chabria et al. 2014), might be transmitted preferentially to variants better adapted to the current host. These mechanisms might help to explain the persistence and transmission of resistance mutations to drugs in untreated hosts because, in an analogous way, resistance mutations usually reduce viral fitness in the host in the absence of selective pressure by drugs (Chabria et al. 2014).

When studying the virus rate of replication, we find a trade-off that represents a nice example of the conflict between selection pressures at the intra- and inter-host levels. The rate of replication of the founder virus is an important factor for the epidemiological success of the disease as well as for the natural history of the viral population in the infected individual. The rate of replication influences the interaction between the viral population and the immune system of the host, which is a key factor determining the outcome of the infection (acute or chronic). The rate of replication is a quantitative trait that also evolves throughout an infection. There are observations of groups of variants in subpopulations, both within and among hosts, with different RNA polymerase activity. Hence, diverse variants with different ranges in their rates of replication can coexist in the same individual (Luciani and Alizon 2009). High rates of growth lead to a stronger immune response against the virus. Consequently, at the inter-host level, the prevalence of slow-replicating variants is favored by natural selection, because it allows a longer time of infection in the host and, as a result, maximizes the reproductive number (R_0) of the infection. In epidemiology, R_0 is defined as the number of new infections caused by an infected individual in a susceptible population and is very closely related to the intrinsic rate of growth of a population in ecological models. However, at the intra-host level, variants with a high rate of replication are favored, because they allow a faster exploitation of the host’s resources.

Therefore, it seems likely that this trade-off results in variants with intermediate rates of replication, which maximize the number of infected individuals from a single host and the exploitation of resources, being favored by natural selection (Luciani and Alizon 2009; Alizon et al. 2009).

Studying the intra- and inter-host dynamics and variation provides relevant information about the transmission and epidemiology of infectious diseases. This is highly relevant in the case of outbreaks, because groups of patients that share similar and even identical viral genotypes usually also show patterns of transmission coincident in time and suggest links that can help to determine the origin or the routes of transmission of the outbreak (Gire et al. 2014). In addition, understanding the evolution and diversity of viruses and their intra- and inter-host dynamics is relevant at the clinical level. The viral diversity and its dynamics are crucial for the design of vaccines (Cuypers et al. 2016; Gaschen et al. 2002) and for determining whether an infection leads to a chronic or acute disease or the chances of success of the antiviral therapy (Gray et al. 2011; Chabria et al. 2014).

The recent advances in sequencing technologies, more specifically in high-throughput sequencing (HTS), have led to significant improvements for the analysis of viral diversity and how it affects intra- and inter-host dynamics. The development of ultra-deep sequencing has been very important for research on chronic viral infections, which can show high levels of intra-host diversity such as HIV and HCV. Its higher sensitivity compared to traditional Sanger sequencing allows a deeper analysis of viral diversity, identifying minority variants and rare polymorphisms that, on the one hand, are invisible for classical techniques, which usually involve reconstructing consensus sequences, and, on the other hand, can be very relevant for basic and applied research (Chabria et al. 2014; Khiabanian et al. 2014). Furthermore, the capability of HTS to sequence a large number of molecules in parallel allows obtaining large datasets, which also help in reducing the economic costs of sequencing (Hall 2007; Churko et al. 2013).

The efforts to investigate evolutionary dynamics at the genome level have focused mainly on RNA viruses causing chronic infections, for which the study of changes in genomic diversity at the intra-host level is more relevant. Among these, HIV and HCV have received most attention due to their evident clinical and epidemiological relevance for humans. In addition, there have also been studies at the genome level aimed at relating intra- and inter-host dynamics in acute disease-causing viruses such as influenza A (Sobel Leonard et al. 2016). Hence, lack of representative data for some viruses is still a major obstacle for studying their population evolution and dynamics.

8 Spatial Distribution of Viruses

The spatial distribution of rapidly evolving viruses depends on ecological and evolutionary processes that interact with each other. In RNA viruses, ecological processes, such as spatial spread, and epidemiological processes occur in a similar time scale to that of evolutionary processes, as a result of their high mutation and evolutionary rates (Holmes 2008). This makes them very appropriate model organisms to study the dynamics of microevolutionary changes, because these can be observed “in real time.” In addition, there is a bias toward studying RNA viruses rather than those with a DNA genome that derives not only from their fast evolution (Duffy et al. 2008) but because, in general, they are more relevant in epidemics and emerging diseases (Holmes 2004; WHO 2017).

Avise (2000) defined phylogeography as the field of study concerned with the principles and processes governing the distribution of geographical lineages at the intraspecific level as well as the interspecific level for related species. In other words, from a more applied perspective, phylogeography includes studies using phylogenetic trees to combine genetic data with spatial information and analyze the spatial patterns suggested in these trees (Holmes 2004; Pybus et al. 2015). Holmes (2004) used a wider definition in which phylogeography incorporates spatial and temporal patterns as well as their interactions. The rapid evolution of viruses can generate enough genetic variation, even at the intra-host level, in just a few days to perform phylogenetic analyses at the infected individual level. This allows applying phylogenetic methods to emerging diseases and to build highly resolved phylogenetic trees (Holmes 2004; Avise 2000; Pybus et al. 2015). The most basic way to integrate spatial and genetic information consists of localizing cases of infections and associating them to different variants (subtypes, genotypes, etc.) of the disease-causing virus (Pybus et al. 2015).

Phylogeographic methods are a powerful tool to infer migration and transmission routes and to reconstruct the evolutionary history of a lineage from genetic data. When applied to viruses, these methods are useful to track the origin of outbreaks and the source of emerging diseases and to reconstruct transmission histories not only between individual hosts but also among social groups of the hosts, among host species, and even their dispersion within body compartments within an individual (De Maio et al. 2015; Alcala et al. 2016).

Due to the coincidence of time scales between molecular evolution and ecological processes that shape their diversity, virus phylogenies provide not only spatial information (i.e., lineages that cluster in geographically defined clades) but also temporal information (i.e., lineages ordered according to sampling times). The molecular clock is a statistical model that establishes a relationship between time and genetic distances in nucleotide sequences. If samples are identified with known dates, then the branching events and the common ancestor in a phylogeny can be placed in a temporal scale. This information can be integrated with spatial information to reconstruct the dispersal history of a virus, linking each branch of the phylogeny with its geographic location. Therefore, with models based on the molecular clock, it is possible to analyze the spread of an epidemic (in months or years) complementing the phylogeny of the isolates with a time scale (Pybus and Rambaut 2009; Pybus et al. 2015). The simplest models for the molecular clock, also known as "strict clock" models, assume a single, constant evolutionary rate for all the lineages. However, more complex, "relaxed clock" models have incorporated variation in the evolutionary rate among lineages or through time (Drummond et al. 2006).

However, the application of phylogeographic tools is valuable only if the spatial epidemiology leaves a signal in the viral genome. This depends both on the rate of molecular evolution and on the rate of transmission in space. If the genome accrues diversity too quickly compared to the rate of spatial spread, then the information provided by phylogeographic analyses is lost as a result of mutation saturation at informative positions (Emmett et al. 2015; Pybus et al. 2015).

Using specific genes or regions to build phylogenetic trees is still a current and complementary approach to analyzing complete genomes (Shen et al. 2016), especially when these genome regions are important sources of predictive information because they encode antigenic proteins (McHardy and Adams 2009). However, the analysis at the genome level is very important to obtain a more complete and unbiased information. Mechanisms such as recombination and reassortment may generate genomes in which different portions thereof have different evolutionary histories (Rambaut et al. 2008; McHardy and Adams 2009; Holmes 2004; Pybus et al. 2015), and this has to be considered when analyzing complete genomes. Next-generation sequencing methods have advanced to the "subnucleotide" level in the analysis of viral sequences. This implies considering the infected individuals as viral populations rather than repetitive collections of the same consensus genome and, additionally, detecting variability within individuals, even very low-frequency variants (subclonal variants). Studying the intra-host and subclonal variability can improve the resolution of phylogenetic analyses and, when combined with epidemiological information, provide a very valuable information to track transmission chains during an outbreak, especially when the transmission rate is very fast, even higher than the viral evolutionary rate (Emmett et al. 2015).

9 Transmission Dynamics

In order to study and understand the dynamics of viral epidemics, we need an approach combining the methods and theories of evolutionary biology, epidemiology, and human geography.

For obligate parasites, such as viruses, which are usually unable to survive for a long time outside their hosts, the mobility and movement patterns of the host are crucial for understanding their transmission dynamics (Alcala et al. 2016; Hufnagel et al. 2004; Pybus et al. 2015). This is closely related to the density and communication between susceptible populations because for virus transmission, a certain proximity between hosts or hosts and vectors is necessary. The smaller the population size of the host, the less likely transmission will be and, consequently, the more difficult to be sustained long enough to cause acute infections. However, large, dense host populations can easily sustain a virus that causes short, virulent infections. In this context, the analysis of the basic reproductive number (R_0) is highly relevant. This number depends on several factors, such as the number of contacts with susceptible individuals, the probability of transmission, and the length of the infectious period (Dietz 1993). This value is very useful to estimate the speed of propagation of an infection in a susceptible population (Ridenhour et al. 2014). The interest in estimating this parameter and its application to the analysis of outbreaks and epidemics and the design of public health strategies gained momentum during the influenza A pandemics of 2009 (Fraser et al. 2009; Ridenhour et al. 2014).

Therefore, the spatial distribution of human viruses will reflect, at least partially, the spatial distribution of human populations, which will also influence the virulence of the disease. However, we must also consider whether the virus can infect other animal species or whether they represent a reservoir for human infections (zoonoses). This is the case for some viruses, such as Ebola virus, with reservoirs in animal species but also capable of being transmitted from person to person. Furthermore, even in RNA viruses well-adapted to humans, there is the possibility of relatively frequent zoonotic contacts, such as in influenza A and MERS-CoV, which are usually associated with the emergence of epidemics and pandemics as a result of genetic exchanges between strains from different species. For vector-borne viruses, we must consider not only human geography but also the geographic distribution of the corresponding vectors, such as different mosquitos of the genera *Aedes* and *Culex*, which are vectors for Zika, dengue, or chikungunya viruses. Spatial distribution analyses should also include ecological features, life history, or migration potential of the vectors (Holmes 2004; Faria et al. 2017; Shen et al. 2016; Bullivant and Martinou 2017; Cunha and Opal 2014).

In the study of the mobility and geographic distribution of humans for understanding the distribution and spread of human viruses, it is necessary to take into account social factors such as international trade and air traffic. The global communications and interrelationships of human populations are growing continuously and represent new opportunities for the transmission, propagation, and colonization of new regions by viruses and their vectors. These can move viruses across geographic barriers and bring into contact with previously isolated populations. This process has contributed to the emergence and reemergence of viral epidemics such as Zika, dengue, and chikungunya. However, we are just starting to understand the effects of global mobility of people and goods on the genetic diversity and evolution of viruses (Alcala et al. 2016; Pybus et al. 2015). To better control epidemics and to understand the evolution and ecology of viruses, it will be necessary to integrate spatial and genomic information along with information about human mobility in a single mathematical framework (Pybus et al. 2015). One example in this direction is BEAST, a framework for Bayesian statistical analysis that allows inference of phylogeographic relationships including spatial and temporal dynamics of migration (Lemey et al. 2009; Drummond and Rambaut 2007).

Clear examples of the relevance of this approach are the analyses of emerging viral epidemics such as SARS or Zika virus. The international spread of Zika virus is likely due to a global increase in air traffic. Specifically, using phylogeographic methods, the origin of the epidemics has been traced to Brazil, where it was detected in 2015, dating its origin in this country between 2013 and 2014 (Worobey 2017). These dates were coincident with several events that brought an important flow of international air traffic to Brazil, such as the 2014 FIFA World Cup (June–July 2014) and the 2013 FIFA Confederations Cup (June 2013) of football (Faria et al. 2016). This highlights the importance of integrating genomic and epidemiologic information about the global movement of persons when surveillance systems are implemented. The large-scale patterns of people's movements can suggest useful hypotheses to study the introduction of viruses and the emergence of epidemics (Faria et al. 2016, 2017; Shen et al. 2016).

From a population genomics perspective, how does this increase in international trade and movements impact the spread of infectious diseases, the population dynamics of viruses, and their genetic diversity?

Isolation and subsequent secondary contact of viral populations are common in natural host populations and can occur at short time scales. These events, facilitated by a higher mobility and contact among human populations, are usually associated to epidemics and pandemics. This has been observed in viruses such as influenza A virus, HIV, and human cytomegalovirus. Furthermore, these processes are important for understanding the evolutionary trajectory of zoonotic viruses, such as Ebola virus.

While they are isolated, viral populations from the same species diverge and adapt to the specific features of their host populations. Hence, during this period, natural selection and demographic changes, such as expansions and bottlenecks, acting on either the viral or the host population will affect the evolution of the virus. After the viral populations are connected again, gene flow, recombination, or reassortment will influence the evolution of the virus, leading to a "mixture" at the genome level. Although selection and demographic changes still act during the reconnection, the other processes act more intensely and rapidly. This mixture impacts on diversity at the genome level: isolated populations have evolved independently, diverging and adapting to the specific conditions of their host populations. After reconnection, the diversity that has accumulated separately increases, which also leads to higher adaptive potential since recombination and reassortment allow the combination of polymorphisms selected in different environments into the same genome. If these polymorphisms are compatible in that particular genomic context, this opens the opportunity for the development of new features which might have not developed (or do so only after very long periods) by just mutation and selection. The second consequence of this shared genetic diversity is a progressive trend toward the homogenization of the populations. Due to the increase in human mobility, these events are expected to be more frequent in the future (Alcala et al. 2016).

10 Epidemiological Surveillance and Genomic Surveillance

Phylogenetic and phylogeographic analyses complement each other, and both are used in epidemiological surveillance systems to control infectious diseases. Phylogeographic information can be used to confirm the source(s) of epidemic outbreaks, and it can also provide valuable information when surveillance is not well implemented or the data it generates are uncertain, unavailable, or insufficient to reconstruct or predict the propagation of the virus (Faria et al. 2017; Pybus et al. 2015). It is even possible to talk about "genomic surveillance" (Emmett et al. 2015) in which the sequencing and analysis of complete genomes contribute to tracking evolution at the genome level as the disease spreads. On the other hand, phylogenetic analysis combined with epidemiological information is useful to study the routes of infection in human populations or the number of introductions that have caused an epidemic (Blackley et al. 2016; Gire et al. 2014; Shen et al. 2016; Drummond et al. 2006; Emmett et al. 2015; Faria et al. 2016).

Another goal of virus phylogeography is to ascertain the future propagation of the organisms and the potential for epidemics by asking which variants are more likely to become predominant and which places are more likely to be colonized and through which ways. This implies building a predictive framework integrating social and environmental factors associated to virus movement and transmission along with genomic and epidemiological information (McHardy and Adams 2009; Pybus et al. 2015).

Influenza A is a good example of how a well-established, global epidemiological surveillance system provides useful information for disease control and vaccine design. It also facilitates the collection of genome sequences at temporal and spatial scales that can be used in evolutionary and phylogeographic analyses. Conversely, at the beginning of the Zika virus epidemics in Brazil in 2015, the country lacked a surveillance system for this virus, and, 1 year later, this task still rested on the passive diagnostics of the disease. This problem, along with the added difficulties for the diagnosis of Zika due to its coexistence with dengue and chikungunya virus, has been a major hurdle in the epidemiological study of the disease and the gathering of abundant genomic information. The example of Zika reinforces the relevance of epidemiological surveillance for the phylogeographic analysis of the virus (Faria et al. 2016; Worobey 2017; Metsky et al. 2017).

The phylogenetic analysis of a virus can help in evaluating the efficiency of surveillance systems. Estimating the most recent common ancestor of a group of sequences can inform about the delay in the detection and notification of the pathogen with respect to the moment of its introduction in the population (Pybus et al. 2015).

One of the limitations in the phylogeographic analysis of viruses is the choice of the correct model. A wrong model selection can lead to erroneous inferences about the transmission history of the pathogen. As epidemiological investigations rely increasingly on genome sequencing to study the origin and spread of infections, the use of accurate phylogeographic methods will be crucial to stop their propagation and design public health preventive measures. De Maio et al. (2015) review different models used to infer transmission rates and spread patterns for viruses, and they illustrate a trade-off between computational costs and speed, on the one hand, and the reliability of the conclusions, on the other hand. The more reliable approaches (continuous models) are, in general, the slowest and most costly with regard to computational resources.

Recently, and partly to fulfill the need for a fast response in cases of outbreaks and emerging epidemics, the so-called discrete character models have gained popularity (Gire et al. 2014). These models treat locations as if they were discrete traits evolving as alleles in a locus. This approach allows a much faster analysis; however, its results are not reliable. They are very sensitive to sampling bias and not robust to scarce genetic data. Different models can yield very different results for the same dataset and, in general, lead to very different and wrong biological interpretations when applied to the study of virus transmission (rather than to the evolution of discrete traits, their original target). De Maio et al. (2015) suggested a model for phylogeographic analysis that combines the advantages of both approaches, discrete and continuous (reliability and precision along with speed and computational efficiency). This model has been used recently in the study of emerging epidemics, such as Zika in Brazil (Faria et al. 2017).

Another limitation for phylogeographic analyses is the public availability of sequences. This depends, in part, on the relevance of the disease caused by the virus, the implementation of an efficient epidemiological surveillance, and the stage of the epidemics. For instance, in 2016 the number of genome sequences for Zika virus available in GenBank was very limited (Shen et al. 2016) as a result of being a recent epidemic and inefficient surveillance. On the contrary, the availability in the public domain of influenza A virus sequences is much higher (Pybus et al. 2015). The rapid publication of genome sequences during emerging epidemics is important to improve genomic and epidemiological surveillance and to monitor the spread of the disease and the adaptive processes in the virus (Gire et al. 2014).

11 Conclusions

Viruses, especially those with RNA genomes, have high mutation rates, short generation times, and large population sizes and are under strong selective pressures. These factors make these viruses organisms with fast evolutionary rates, high genetic variability, and great adaptive capacity.

Understanding the mechanisms that allow human viruses to generate and maintain genetic diversity and to adapt to the host's selective pressure is fundamental for human health. A better knowledge of the evolution of human viruses at the genome level can shed light on questions such as the evasion of immune response, the development and transmission of resistance mutations, vaccine design, the evolution and virulence of the disease or the control of outbreaks, epidemics, and emerging diseases.

In general, viruses, as any other pathogen, are under strong selective pressure by the immune system of their hosts. In addition, human viruses are usually under the additional pressure of antiviral drugs and treatments. These pressures result in high mutation rates in those genome regions involved in the interaction with the host and in those that encode the targets of antiviral drugs. This leads to the development of drug resistance and of mechanisms to evade the immune system.

The typically high mutation rates of RNA viruses are, most likely, another consequence of these selective pressures because in a stable environment (very different from a host infected by the virus), natural selection will favor a low mutation rate (Kamp et al. 2002). This common feature of RNA viruses is a key factor to explain their adaptation, and, simultaneously, it keeps viral populations at the extinction threshold by accumulating an excessive number of deleterious mutations. Recently, it has been observed that the human immune system might take advantage of this feature to fight viral infections by forcing hypermutation in the viral genome.

The genomic diversity is also limited by different constraints: the need to keep a small genome size, RNA secondary structures at the genome level, structural domains of proteins to sustain their function, and gene overlapping. In some viruses, such as HCV, these negative selection pressures might be the main factor driving evolution. In others, such as HIV, positive selection has a more relevant role.

In this context, it is important to consider how the interactions between genome positions can affect the "displacement" of different mutations through the adaptive landscape. Mutations that could be considered as deleterious, such as some resistance mutations or those that disrupt secondary structures, can be retained in a population and even spread rapidly depending on the genome context where they appear.

The population dynamics of RNA viruses are different depending on the level of biological organization at which they are analyzed. Selective pressures acting at the intra-host and inter-hosts levels can differ and often act in opposite directions. Frequently, these selective pressures conflict between the need to adapt to the host and the ability for transmission to other hosts. Those variants that are favored by selection within hosts – mutations for evading the immune system and drug resistance – may diminish the capacity for transmission of the virus and, in consequence, will be selected against at the inter-host level. In addition, every transmission event represents a bottleneck that reduces drastically the population size of the virus and, consequently, also its genetic diversity. This leads to slower evolutionary rates at the inter-host level. For instance, in HIV there seems to be an inverse relationship between transmission and evolutionary rate (Berry et al. 2007). We must also consider how and by which means is the virus transmitted. Transmission rates are higher in air-transmitted virus, such as influenza A, than in those that use the sexual route. Similarly, those viruses that use arthropod species as vectors have lower rates of evolution, a cost associated to their need for replication in different hosts (Holmes 2004; Woelk and Holmes 2002).

Another consequence of the high rates of evolution is that ecological and evolutionary processes acting on viral populations occur at similar time scales. Their interaction affects their spatial distribution. The combination of complete viral genomes and phylogeographic methods is very useful for tracking the origin of epidemic outbreaks, locating reservoirs that may act as sources of infection for humans or of new potentially virulent strains (such as influenza A), to reconstruct transmission histories and to monitor the spread of an epidemics. These applications are very relevant nowadays, in an increasingly connected planet in which trade and air traffic bring geographically distant populations close and erase natural barriers for the transmission of diseases. Furthermore, human impacts on previously intact ecosystems are helping the emergence and global spread of new infectious, as illustrated by the recent epidemics of Zika and Ebola viruses.

12 Future Perspectives

The development of population genomics is closely linked to advances in sequencing technologies. Standard techniques, based on deriving consensus sequences, miss the presence of minor or subclonal variants (low-frequency polymorphisms) which might be important to understand the dynamics of viral populations as well as the evolution and spread of the disease. Next-generation sequencing techniques allow the detection of rare polymorphisms and minor variants and lead to consideration of infected hosts as viral populations rather than "collections" of the same consensus genome. Consequently, these methods provide a better view of viral diversity, which enables an improvement in the study of the epidemiology and evolution of human viruses. A more widespread use of these technologies to characterize genome variation will provide increased information about the intra-host dynamics and the relationship between viral diversity and infection outcome (Liu et al. 2012; Farci et al. 2000), the inter-host transmission and dynamics (reservoirs for better-transmitted variants), the development of resistance and the failure of antiviral treatments, and the building of highly resolved phylogenies and transmission histories during epidemic outbreaks. In addition, advances in sequencing technologies have also allowed the fast and in-depth analysis of complete genomes. The evolution and accumulation of genetic variation occur differently and simultaneously throughout the genome. Separate regions of the same genome can interact with each other (epistasis) and, even, evolve independently and show different phylogenetic histories. Hence, the possibility of analyzing complete genomes – as opposed to the analysis of individual loci or isolated genome regions – provides a more complete, resolved, and less biased view of genomic variation, the phylogeny and population dynamics of the virus.

Finally, an important limitation in the population genomic study of virus populations is the availability of genomic information for many viruses. This is intimately related to the clinical and epidemiological relevance of the disease caused by most viruses. Human diseases with high prevalence and important consequences such as HIV, hepatitis C, or influenza receive much attention in the public health realm and have a more efficient surveillance. This translates in higher availability of viral genomes and epidemiological information, which are necessary for the evolutionary analysis of virus populations.

The evolutionary analysis of viral genomes and epidemiological surveillance are, in consequence, necessarily complementary. Implementing a "genomic surveillance" can contribute to control and monitor the spread of infectious diseases and to design better public health strategies to achieve these goals.

Acknowledgments This work was supported by projects BFU2014-58656R and BFU2017-89594R from MINECO (Spanish Government) and PROMETEO2016-0122 from Generalitat Valenciana.

References

Alcala N, Jensen JD, Telenti A, Vuilleumier S. The genomic signature of population reconnection following isolation: from theory to HIV. G3 (Bethesda). 2016;6(1):107–20.

Alizon S, Fraser C. Within-host and between-host evolutionary rates across the HIV-1 genome. Retrovirology. 2013;10(1):49.

Alizon S, Hurford A, Mideo N, van Baalen M. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. J Evol Biol. 2009;22(2):245–59.

Anisimova M, Nielsen R, Yang Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics. 2003;164(3):1229–36.

Assis R. Strong epistatic selection on the RNA secondary structure of HIV. PLoS Pathog. 2014;10 (9):e1004363.

Avise JC. Phylogeography. The history and formation of species. 1st ed. Cambridge: Harvard University Press; 2000.

Baird HA, Galetto R, Gao Y, Simon-Loriere E, Abreha M, Archer J, et al. Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. Nucleic Acids Res. 2006;34 (18):5203–16.

Berry IM, Ribeiro R, Kothari M, Athreya G, Daniels M, Lee HY, et al. Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. J Virol. 2007;81(19):10625–35.

Blackley DJ, Wiley MR, Ladner JT, Fallah M, Lo T, Gilbert ML, et al. Reduced evolutionary rate in reemerged Ebola virus transmission chains. Sci Adv. 2016;2(4):e1600378.

Bradwell K, Combe M, Domingo-Calap P, Sanjuán R. Correlation between mutation rate and genome size in riboviruses: mutation rate of bacteriophage Qb. Genetics. 2013;195(1):243–51.

Brandes N, Linial M. Gene overlapping and size constraints in the viral world. Biol Direct. 2016;11 (1):1–15.

Brockman MA, Brumme ZL, Brumme CJ, Miura T, Sela J, Rosato PC, et al. Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated for in chronic infection. J Virol. 2010;84(22):11937–49.

Bullivant G, Martinou AF. Ascension Island: a survey to assess the presence of Zika virus vectors. J R Army Med Corps. 2017;163(5):347–54.

Campo DS, Dimitrova Z, Mitchell RJ, Lara J, Khudyakov Y. Coordinated evolution of the hepatitis C virus. Proc Natl Acad Sci U S A. 2008;105(28):9685–90.

Chabria SB, Gupta S, Kozal MJ. Deep sequencing of HIV: clinical and research applications. Annu Rev Genomics Hum Genet. 2014;15(1):295–325.

Churko JM, Mantalas GL, Snyder MP, Wu JC. Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases. Circ Res. 2013;112(12):1613–23.

Cobey S, Koelle K. Capturing escape in infectious disease dynamics. Trends Ecol Evol. 2008;23 (10):572–7.

Cuevas JM, Geller R, Garijo R, Lòpez-Aldeguer J, Sanjuán R. Extremely high mutation rate of HIV-1 *in vivo*. PLoS Biol. 2015;13(9):e1002251.

Cunha CB, Opal SM. Middle East respiratory syndrome (MERS): a new zoonotic viral pneumonia. Virulence. 2014;5(6):650–4.

Cuypers L, Li G, Neumann-Haefelin C, Piampongsant S, Libin P, Van Laethem K, et al. Mapping the genomic diversity of HCV subtypes 1a and 1b: implications of structural and immunological constraints for vaccine and drug development. Virus Evol. 2016;2(2):vew024.

De Maio N, Wu CH, O'Reilly KM, Wilson D. New routes to phylogeography: a Bayesian structured coalescent approximation. PLoS Genet. 2015;11(8):e1005421.

Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res. 1993;2(1):23–41.

Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiol Mol Biol Rev. 2012;76 (2):159–216.

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4(5):e88.

Duan S, Govorkova EA, Bahl J, Zaraket H, Baranovich T, Seiler P, et al. Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. Nat Commun. 2014;5:5029.

Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet. 2008;9(4):267–76.

Eigen M. On the nature of virus quasispecies. Trends Microbiol. 1996;4(6):216–8.

Emmett KJ, Lee A, Khiabanian H, Rabadan R. High-resolution genomic surveillance of 2014 Ebolavirus using shared subclonal variants. PLoS Curr. Outbreaks 2015;7.

Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, et al. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science. 2000;288:339–44.

Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. Genome Med. 2016;8(1):97.

Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017;546:406–10.

Feder AF, Rhee SY, Holmes SP, Shafer RW, Petrov DA, Pennings PS. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. Elife. 2016;5:e10670.

Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. Science. 2009;324 (5934):1557–61.

Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, et al. Diversity considerations in HIV-1 vaccine selection. Science. 2002;296(5577):2354–60.

Geller R, Domingo-Calap P, Cuevas JM, Rossolillo P, Negroni M, Sanjuan R. The external domains of the HIV-1 envelope are a mutational cold spot. Nat Commun. 2015;6:8571.

Geller R, Estada Ú, Peris JB, Andreu I, Bou JV, Garijo R, et al. Highly heterogeneous mutation rates in the hepatitis C virus genome. Nat Microbiol. 2016;1(7):16045.

Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345 (6202):1369–72.

Goonetilleke N, Liu MKP, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J Exp Med. 2009;206(6):1253–72.

Gray R, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus O. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC Evol Biol. 2011;11(1):131.

Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303(5656):327–32.

Hall N. Advanced sequencing technologies and their wider impact in microbiology. J Exp Biol. 2007;210(9):1518–25.

Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 2005;169(4):2335–52.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, et al. Phylogeography's past, present, and future: 10 years after. Mol Phylogenet Evol. 2010;54 (1):291–301.

Holmes EC. Error thresholds and the constraints to RNA virus evolution. Trends Microbiol. 2003;11(12):543–6.

Holmes EC. The phylogeography of human viruses. Mol Ecol. 2004;13(4):745–56.

Holmes EC. Evolutionary history and phylogeography of human viruses. Annu Rev Microbiol. 2008;62(1):307–28.

Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci U S A. 2004;101(42):15124–9.

Jackowiak P, Kuls K, Budzko L, Mania A, Figlerowicz M, Figlerowicz M. Phylogeny and molecular evolution of the hepatitis C virus. Infect Genet Evol. 2014;21(1):67–82.

Joseph SB, Swanstrom R, Kashuba AD, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. Nat Rev Microbiol. 2015;13(7):414–25.

Kamp C, Wilke CO, Adami C, Bornholdt S. Viral evolution under the pressure of an adaptive immune system: optimal mutation rates for viral escape. Complexity. 2002;8(2):28–33.

Kearney M, Maldarelli F, Shao W, Margolick JB, Daar ES, Mellors JW, et al. Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. J Virol. 2009;83(6):2715–27.

Khiabanian H, Carpenter Z, Kugelman J, Chan J, Trifonov V, Nagle E, et al. Viral diversity and clonal evolution from unphased genomic data. BMC Genomics. 2014;15(6):S17.

Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. 2011;7(2):e1001301.

Kubinak JL, Ruff JS, Hyzer CW, Slev PR, Potts WK. Experimental viral evolution to specific host MHC genotypes reveals fitness and virulence trade-offs in alternative MHC types. Proc Natl Acad Sci U S A. 2012;109(9):3422–7.

Lai MM. RNA recombination in animal and plant viruses. Microbiol Rev. 1992;51(1):61–79.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5(9):e1000520.

Levin BR, Bull JJ. Short-sighted evolution and the virulence of pathogenic microorganisms. Trends Microbiol. 1994;2(3):76–81.

Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, et al. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. J Virol. 2007;81(1):193–201.

Liu Y, McNevin JP, Holte S, McElrath MJ, Mullins JI. Dynamics of viral evolution and CTL responses in HIV-1 infection. PLoS One. 2011;6(1):e15639.

Liu L, Fisher BE, Thomas DL, Cox AL, Ray SC. Spontaneous clearance of primary acute hepatitis C virus infection correlated with high initial viral RNA level and rapid HVR1 evolution. Hepatology. 2012;55(6):1684–91.

Luciani F, Alizon S. The evolutionary dynamics of a rapidly mutating virus within and between hosts: the case of hepatitis C virus. PLoS Comput Biol. 2009;5(11):e1000565.

Lythgoe KA, Fraser C. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. Proc R Soc B. 2012;279(1741):3367–75.

Martin M, Del Cacho E, Codina C, Tuset M, De Lazzari E, Mallolas J, et al. Relationship between adherence level, type of the antiretroviral regimen, and plasma HIV type 1 RNA viral load: a prospective cohort study. AIDS Res Human Retrovir. 2008;24(10):1263–8.

Maynard-Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974;23 (1):23–35.

McDonald SM, Nelson MI, Turner PE, Patton JT. Reassortment in segmented RNA viruses: mechanisms and outcomes. Nat Rev Microbiol. 2016;14(7):448–60.

McGowan JE Jr. Economic impact of antimicrobial resistance. Emerg Infect Dis. 2001;7(2):286.

McHardy AC, Adams B. The role of genomics in tracking the evolution of influenza A virus. PLoS Pathog. 2009;5(10):e1000566.

Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol. 2013;28(11):659–69.

Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, et al. Zika virus evolution and spread in the Americas. Nature. 2017;546:411–5.

Miralles R, Gerrish PJ, Moya A, Elena SF. Clonal interference and the evolution of RNA viruses. Science. 1999;285:1745–7.

Muller HJ. Some genetic aspects of sex. Am Nat. 1932;66:118–38.

Murrell B, De Oliveira T, Seebregts C, Kosakovsky Pond SL, Scheffler K, Southern African Treatment and Resistance Network (SATuRN) Consortium. Modeling HIV-1 drug resistance as episodic directional selection. PLoS Comput Biol. 2012;8(5):e1002507.

Negroni M, Buc H. Copy-choice recombination by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones. Proc Natl Acad Sci U S A. 2000;97(12):6385–90.

Neogi U, Shet A, Sahoo PN, Bontell I, Ekstrand ML, Banerjea AC, Sonnerborg A. Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals. J Int AIDS Soc. 2013;16(1):18472.

Neverov AD, Kryazhimskiy S, Plotkin JB, Bazykin GA. Coordinated evolution of influenza A surface proteins. PLoS Genet. 2015;11(8):e1005404.

Noguera-Julian M, Cozzi-Lepri A, Di Giallonardo F, Schuurman R, Däumer M, Aitken S, et al. Contribution of APOBEC3G/F activity to the development of low-abundance drug-resistant human immunodeficiency virus type 1 variants. Clin Microbiol Infect. 2016;22(2):191–200.

Nomikou K, Hughes J, Wash R, Kellam P, Breard E, Zientara S, et al. Widespread reassortment shapes the evolution and epidemiology of bluetongue virus following European invasion. PLoS Pathog. 2015;11(8):e1005056.

Patiño Galindo JA, González-Candelas F. Comparative analysis of variation and selection in the HCV genome. Infect Genet Evol. 2017;49:104–10.

Pennings PS, Kryazhimskiy S, Wakeley J. Loss and recovery of genetic diversity in adapting populations of HIV. PLoS Genet. 2014;10(1):e1004000.

Phillips PC. Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9(11):855–67.

Pybus OG, Rambaut A. Modelling: evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet. 2009;10:540–50.

Pybus OG, Tatem AJ, Lemey P. Virus evolution and transmission in an ever more connected world. Proc Biol Sci. 2015;282(1821):20142878.

Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008;453(7195):615.

Ramirez BC, Simon-Loriere E, Galetto R, Negroni M. Implications of recombination for HIV diversity. Virus Res. 2008;134(1):64–73.

Renzette N, Caffrey DR, Zeldovich KB, Liu P, Gallagher GR, Aiello D, et al. Evolution of the influenza A virus genome during development of oseltamivir resistance in vitro. J Virol. 2014;88(1):272–81.

Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: considerations for public health applications. Am J Public Health. 2014;104(2):e32–41.

Rogozin I, Spiridonov A, Sorokin A, Wolf Y, Jordan I, Tatusov R, et al. Purifying and directional selection in overlapping prokaryotic genes. Trends Genet. 2002;18(5):228–32.

Ross HA, Rodrigo AG. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J Virol. 2002;76(22):11715–20.

Rothenberger S, Torriani G, Johansson MU, Kunz S, Engler O. Conserved endonuclease function of hantavirus L polymerase. Viruses. 2016;8(5):108.

Sanjuán R, Bordería AV. Interplay between RNA structure and protein evolution in HIV-1. Mol Biol Evol. 2011;28(4):1333–8.

Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010;84 (19):9733–48.

Shen S, Shi J, Wang J, Tang S, Wang H, Hu Z, Deng F. Phylogenetic analysis revealed the central roles of two African countries in the evolution and worldwide spread of Zika virus. Virol Sin. 2016;31(2):118–30.

Simmonds P, Tuplin A, Evans DJ. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. RNA. 2004;10(9):1337–51.

Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? Nat Rev Microbiol. 2011;9 (8):617–26.

Simon-Loriere E, Holmes EC, Pagán I. The effect of gene overlapping on the rate of RNA virus evolution. Mol Biol Evol. 2013;30(8):1916–28.

Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. Virus Res. 2012;169 (2):415–29.

Snoeck J, Fellay J, Bartha I, Douek D, Telenti A. Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. Retrovirology. 2011;8(1):87.

Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, Lin X, et al. Deep sequencing of influenza A virus from a human challenge study reveals a selective bottleneck and only limited intrahost genetic diversification. J Virol. 2016;90(24):11247–58.

Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, Lin X, et al. The effective rate of influenza reassortment is limited during human infection. PLoS Pathog. 2017;13(2): e1006203.

Steel J, Lowen AC. Influenza A virus reassortment. In: Influenza pathogenesis and control – volume I. Cham: Springer; 2014. p. 377–401.

Thurner C, Witwer C, Hofacker IL, Stadler PF. Conserved RNA secondary structures in Flaviviridae genomes. J Gen Virol. 2004;85(5):1113–24.

Van Valen L. A new evolutionary law. Evol Theory. 1973;1:1–30.

Veeramachaneni V, Makalowski W, Galdzicki M, Sood R, Makalowska I. Mammalian overlapping genes: the comparative perspective. Genome Res. 2004;14(2):280–6.

Vrancken B, Baele G, Vandamme AM, Van Laethem K, Suchard MA, Lemey P. Disentangling the impact of within-host evolution and transmission dynamics on the tempo of HIV-1 evolution. AIDS. 2015;29(12):1549–56.

Vuilleumier S, Bonhoeffer S. Contribution of recombination to the evolutionary history of HIV. Curr Opin HIV AIDS. 2015;10(2):84–9.

Wang W, Zhang X, Xu Y, Weinstock GM, Di Bisceglie AM, Fan X. High-resolution quantification of hepatitis C virus genome-wide mutation load and its correlation with the outcome of peginterferon-alpha2a and ribavirin combination therapy. PLoS One. 2014;9(6):e100131.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009;460(7256):711–6.

WHO. Emerging zoonoses. 2017. http://www.who.int/zoonoses/emerging_zoonoses/en.

WHO Scientific Working Group. Antimicrobial resistance. Bull World Health Organ. 1983;61 (3):383–94.

Wilson BA, Garud NR, Feder AF, Assaf ZJ, Pennings PS. The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens. Mol Ecol. 2016;25(1):42–66.

Woelk CH, Holmes EC. Reduced positive selection in vector-borne RNA viruses. Mol Biol Evol. 2002;19(12):2333–6.

Worobey M. Molecular mapping of Zika spread. Nature. 2017;546:355–7.

Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, Brumme CJ, et al. Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. J Virol. 2010;84(20):10820–31.

Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, et al. Population genomics of intrapatient HIV-1 evolution. Elife. 2015;4:e11282.