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

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

Marcos Pérez-Losada, Miguel Arenas, Juan Carlos Galán, Ferran Palero, Fernando González-Candelas

© 2014 Elsevier B.V. All rights reserved.
1. Introduction

Viruses encompass an enormous variety of genomic structures. The majority of classified viruses have RNA genomes, but many have DNA genomes, and some may even have DNA and RNA at different stages in their life cycle. Attending to genome architecture, they can be either single-stranded (ss) or double-stranded (ds) and some virus families (e.g., Hepadnaviridae) contain a double-stranded genome with single-stranded regions. Finally, several virus classes (RNA genomes and some with ssDNA genomes) are characterized by the basis of the nature and polarity of their mRNAs, depending on whether they have the translatable information in the same or complementary strand (denoted as plus-strand, +, or minus-strand, –). Several types of ssDNA (e.g., geminiviruses) and ssRNA (e.g., arenaviruses) viruses have genomes that are ambifunctional (+/–). All these genome properties are used to classify viruses into seven groups (following Baltimore, 1971): Group I, ssDNA viruses (e.g., Herpesvirus); II, ssDNA (e.g., Parvoviruses); III, dsRNA (e.g., Reoviruses); IV, (+)ssRNA (e.g., Picornaviruses); V, (–)ssRNA (e.g., Orthomyxovirus); VI, ssRNA-RT (reverse transcriptase) (e.g., Retroviruses); VII: dsDNA-RT (e.g., Hepadnaviruses). This huge biological diversity suggests multiple strategies for generating genetic diversity.

Viruses undergo genetic change by several mechanisms, including point mutation (the ultimate source of genetic variation) and recombination. In general, RNA viruses have smaller genomes than DNA viruses, probably as consequence of their higher mutation rates (Holmes, 2003). The reason for this inverse relationship between genome size and mutation rate is arguably the incapability of large RNA viruses to replicate without generating lethal mutations (Belshaw et al., 2007; Sanjuan et al., 2010). In contrast, DNA viruses generally have larger genomes because of the higher fidelity of their replication enzymes. ssDNA viruses are an exception to this rule, since the mutation rates of their genomes can be as high as those seen in ssRNA genomes (Duffy and Holmes, 2009).

Recombination occurs when at least two viral genomes co-infect the same host cell and exchange genetic segments. Different types of viral recombination are recognized based on the structure of the crossover site (Austermann-Busch and Becher, 2012; Scheel et al., 2013). Homologous recombination occurs in the same site in both parental strands, while non-homologous or illegitimate recombination (Lai, 1992) occurs at different sites of the genetic fragments involved, frequently originating aberrant structures (Galli and Buhk, 2014). A particular type of recombination, known as shuffling or reassortment, occurs in viruses with segmented genomes, which can interchange complete genome segments, giving rise to new segment combinations. Illegitimate recombination is relatively infrequent in RNA viruses, but in DNA viruses can occur at much higher frequencies than homologous recombination (Robinson et al., 2011). In addition, the exchange of genetic material between viruses is usually non-reciprocal, meaning the recipient of a genome portion does not act as donor of the replaced portion in the original source. In this respect, the term recombination does not have the same meaning in viruses that it does in diploid, sexually reproducing organisms wherein the exchange of genetic material between chromatids in the first meiosis division is reciprocal. Viral recombination could be more appropriately denoted as “gene conversion,” but the term has become so widely used that it is pointless to change it.

Recombination is a widespread phenomenon in viruses and can have a major impact on their evolution. Indeed, recombination has been associated with the expansion of viral host ranges, the emergence of new viruses, the alteration of transmission vector specificities, increases in virulence and pathogenesis, the modification of tissue tropisms, the evasion of host immunity, and the evolution of resistance to antivirals (Martin et al., 2011a; Simon-Loriere and Holmes, 2011). The frequencies of recombination vary extensively among viruses. Recombination seems highly frequent in some dsDNA viruses, such as α-Herpesviruses, where recombination is intimately linked to replication and DNA repair (Robinson et al., 2011; Thiry et al., 2005) and can prevent the progressive accumulation of harmful mutations in their genomes (i.e., prevent the mutational meltdown and increase fitness). In fact, homologous recombination seems to be the most frequent mechanism in this group of viruses, although illegitimate recombination has also been observed (Robinson et al., 2011; Thiry et al., 2005, 2006). DNA recombination can also facilitate access to evolutionary innovations that would otherwise be inaccessible by mutation alone, although the viability of DNA recombinants will depend on how severely recombination disrupts genome architecture and the interaction networks (Lefevre et al., 2009).

Among RNA viruses, recombination is particularly frequent in Retroviruses (ssRNA-RT), most notably in HIV, where the rate of recombination per nucleotide exceeds that of mutation (Jetzt et al., 2000). In contrast, recombination occurs at variable frequencies in (+)ssRNA viruses, with some families showing high rates (e.g., Picornaviridae), while others see only occasional (e.g., Flaviviridae) or nonexistent (e.g., Leviviridae) occurrence. Current data also indicate that recombination is even far less common in (–)ssRNA viruses (Holmes, 2009). Similarly, there is extensive variation in the rate of re-assortment among RNA viruses. Many segmented viruses, such as Hantaviruses, Lassa virus and Tenuiviruses, exhibit relatively low levels of re-assortment, while others show high rates (e.g., influenza A virus, rotavirus A and Cystoviruses) (Simon-Loriere and Holmes, 2011).

The evolutionary reasons for the occurrence of recombination in RNA viruses are less clear. Since RNA viruses exhibit high mutation rates and large population sizes, it is more likely that these factors, rather than recombination, drive their evolutionary fate, as they regularly produce advantageous mutations and protect themselves from the accumulation of deleterious ones. There is little compelling evidence supporting the idea that recombination enhances genome repair (except in viruses with diploid or pseudodiploid genomes, such as HIV) or has evolved as a form of sexual reproduction, allowing the more efficient purging of deleterious mutations and accelerating the generation of advantageous genetic combinations (Holmes, 2009; Simon-Loriere and Holmes, 2011). Instead, prior hypotheses contend that mechanistic constraints associated with particular genome structures and viral life histories are the major elements shaping recombination rates in RNA viruses. Consequently, RNA recombination and re-assortment can be considered as byproducts of selection for other genomic characteristics to control gene expression, rather than being favored as specific
manifestations of sexual reproduction. Nevertheless, this does not exclude the possibility that natural selection can favor specific genotypes generated by recombination (Holmes, 2009).

Although it has been postulated that recombination facilitates cross-species transmission, its actual relevance is quite uncertain in some RNA viruses (Simon-Loriere and Holmes, 2011). Indeed, recombination and re-assortment are powerful ways for viruses to acquire new antigenic combinations that may assist in the process of entering a new host. However, in most cases the emergence of a specific RNA virus cannot be attributed directly to its ability to recombine. Notable exceptions are the Western equine encephalitiis virus (Weaver, 2006) and the turkey coronavirus (Jackwood et al., 2010). We are still far from fully understanding the evolution of recombination in RNA viruses. Hopefully, the development of next-generation sequencing methods will provide new insights into the causes and consequences of recombination in this viral group (Simon-Loriere and Holmes, 2011).

Given the frequency, relevance and impact of recombination in viral evolution, it is not surprising that many bioinformatic approaches have been developed to accurately detect and estimate its occurrence in viral genomes. Some of these tools already take advantage of the genomic data generated using high-throughput sequencing. Nonetheless, the new opportunities that these technologies offer also come with some serious analytical challenges that must be considered. In the following sections we briefly review (i) the different mechanisms of recombination in DNA and RNA viruses, (ii) the different methods for detecting and estimating recombination, and (iii) the impact of recombination on the inference of evolutionary forces acting on viral populations. For more comprehensive reviews on these topics, we refer the reader to the specific literature cited in each section.

2. Mechanisms of recombination in viruses

Genetic recombination is the molecular process by which new genetic combinations are generated from the crossover of two nucleic acid strands. Depending on the origin of the parental strands, we can distinguish between intra-genomic recombination (wherein the recombinating strands belong to the same genetic unit) and inter-genomic recombination (wherein the fragments have different origins). Intra-genomic recombination is related to the regulation of gene expression (Johansson, 2013). This is a particular feature of Herpesviruses, such as human Herpesvirus-1 (HSV-1), probably the best studied DNA virus. Inter-genomic recombination represents a strategy to generate genetic diversity and, paradoxically, to maintain chromosomal integrity. Inter-genomic recombination is the most relevant form for both DNA and RNA viruses because it is capable of generating new variants with different biological features, and in this section we will focus on its different mechanisms.

Several genetic, biological and epidemiological factors can affect the probability of recombination between different strains, including genetic homology, population size in the host, co-circulation in the same geographical area, prevalence in the population, and coinfection rate. For instance, the most closely related members of human α-Herpesviruses, such as HSV-1 and HSV-2 (75% similarity), are capable of recombining between themselves under laboratory conditions (Halliburton et al., 1977), but recombinant forms have not been found among other members of this subfamily with lower similarity. Similar results were obtained with HIV-1 recombination (Galli et al., 2010). The complexity of recombination patterns was proportional to the levels of similarity between the HIV-1 subtypes, as a result of decreased opportunities for recombination. In many cases, viral recombination is replication-dependent, so that high viral loads increase the chances of exchanging strands (De Candia et al., 2010). Among the epidemiological factors, the extensive co-circulation of different variants and high prevalence of infections are opportunities for coinfection (i.e., simultaneous infection by different viruses of the same individual host), an essential requirement for inter-genomic recombination (Bowden et al., 2004). A good example has been the epidemic of HIV-1 in Africa, where its elevated prevalence has favoured high coinfection rates (Redd et al., 2011), giving rise to many new recombinant forms (http://www.unaids.org/GlobalReport/). The impact of coinfections on recombination might be dependent on the time interval between first and second viral infections. Meurens et al. (2004) demonstrated in HSV that the time interval between successive infections was inversely proportional to the number of recombinant forms detected, because superinfection can be prevented when viral proteins are expressed (Wildum et al., 2006). The superinfection exclusion phenomenon has also been detected in RNA viruses such as HCV (Schaller et al., 2007). However, as the HIV epidemic has demonstrated, this barrier is not totally efficient (Piantadosi et al., 2007; Redd et al., 2013; Koning et al., 2013) because superinfection can occur several years after the primary infection (Pernas et al., 2006) with – as is the case with HCV – an increasing number of inter- and intrasubtype recombinant strains being detected (González-Candelas et al., 2011).

Different mechanisms of recombination have been described in DNA and RNA viruses, but only a few examples have been characterized in detail. Detailed information about the specific mechanisms of the process will be presented in this section (i.e., HSV-1 among DNA viruses and, influenza and HIV, as examples of RNA viruses and retroviruses). The α-Herpesviruses example is probably the best characterized one, as these viruses represent an important model for the study of eukaryotic DNA replication (Weller and Coen, 2012). The first evidence of recombination in HSV-1 was obtained in 1955 thanks to Wildy’s (1955) work using two temperature-sensitive HSV-1 mutants. Two mechanisms of recombination are generally accepted for RNA viruses: replicative and non-replicative. The replicative model requires the transfer of the replication complex from a RNA template to a different template, whereas the non-replicative model requires a previous break in the RNA molecule facilitating the formation of hybrid genomes (Galli and Bukh, 2014).

2.1. Recombination in DNA viruses

Herpesviruses share a mechanistic model with bacteriophages (involving concatamers, two-component recombinases, and commandeering cellular proteins) in which recombination and DNA replication are interconnected (Weller and Sawitzke, 2014), suggesting an ancestral mechanism. Moreover, all Herpesviruses have a common replicative complex characterized by highly conserved groups of recombination and replication proteins (Rennekamp and Lieberman, 2010). In fact, many of these proteins can be interchanged between different Herpesviruses; for example, Epstein-Barr virus proteins have been shown to replicate human cytomegalovirus (Sarisky and Hayward, 1996). In DNA viruses, recombination also plays an important role in the repair of DNA damage, especially in the latency stage (Wilkinson and Weller, 2004) but, paradoxically to previous observations, the analysis of complete genomes of many Herpesviruses has shown that the repair-initiating DNA sequences differ between α- and γ-Herpesviruses, thus suggesting possible differences in the recombination mechanisms between these two subfamilies (Brown, 2014).

Considerable controversy exists about the onset of replication once cell infection by HSV has been completed. At present, it is not known whether viral DNA is circularized before the start of replication, although several pieces of indirect evidence (such as loss of free ends in a short time after infection and bidirectional
replication) suggest that a circular DNA is formed. Therefore, we accept this as the most plausible model at this moment. When linear viral DNA reaches the cell nucleus, it is rapidly converted into a circular form by human DNA ligase IV (Fig. 1). Immediately, the replicative complex or replisome, constituted by a core set of seven highly conserved proteins (UL9/UL29; UL5/UL8/UL52 and UL30/UL42), begins to replicate the DNA. This is a coordinated process with several well-defined steps. DNA replication initiates when the UL9 protein (the so-called “origin-binding protein”) recognizes the origins of replication (ori) and disrupts the dsDNA structure. There are three ori in the HSV genome (two oriS in US fragment and one oriL in UL fragment). Mutants containing only one origin frequently undergo recombination resulting in the duplication of a second site, suggesting an evolutionary benefit in having multiple origins of replication in this virus. In the next step, the UL29 protein (also known as “single strand binding protein”, ICP8) is recruited, forming an UL9-UL29 complex, and induces conformational changes (recombinogenic hairpin structure). UL29 has been reported to contact several viral (UL9, UL12, ICP4 or ICP27) and cellular proteins (Ku86 or Rad50) to carry out its multiple functions, such as viral DNA synthesis or control of viral gene expression. Once the origin is activated, the HSV-1 helicase/primase complex (UL5/UL8/UL52) is now recruited to unwind, in the presence of UL29, the duplex DNA and synthesize short RNA primers. These are necessary to initiate DNA replication by the action of DNA polymerase (UL30/UL42) joining simultaneously to UL29 and U5, completing the replicative complex. A bidirectional theta-type replication model has been proposed in the initial amplification of circular DNA, which then switches to a rolling-circle replication model to generate concatemers (sequences larger-than-unit length) of the viral genome. Electronic microscopy analyses have shown that the accumulation of concatemers is highly branched, suggesting that recombination events occur as soon as newly replicated DNA is detected (Severini et al., 1996). The coordinate action of UL12 (viral exonuclease) and UL29 cleaves the newly synthesized DNA into effective units prior to packaging (Muylaert et al., 2011). Concatemers of HSV-1 genomes contain genomic inversions suggestive of strand-transfer events (Bataille and Epstein, 1994). During DNA replication, multiple random double-strand breaks are produced and these breaks are efficient initiators of homologous recombination events between inverted regions, giving rise to new reassortments of isomers. These inversion events occur during DNA replication as a consequence of the recognition of multiple origins of replication by the UL9/UL29 complex. The UL29 protein has also recombinase activity, promoting strand invasion and, together with the helicase/primase complex (UL5/UL8 and UL52), it promotes strand exchange during replication (Makhov and Griffith, 2006). Moreover, the UL12 protein stimulates recombination through a single-strand annealing mechanism between

Fig. 1. Replication and recombination steps in herpes simplex virus. Both processes are intimately linked. Although the re-circularization of linear DNA has not demonstrated experimentally, several lines of evidence suggest that it must occur. Three ori sites are the distributed along the DNA genome (2 oriS and 1 oriL), but in this figure we only show the interaction of the replication complex with one site. In the oriS (45 bp) there are two fragments with high affinity to UL9 (box1 and box2) characterized by the sequence TTGCAC, whereas in oriL (144 bp) two box1 are present. It has been suggested that these origins may function differently during lytic and latent infection.
concatamers at the end of the replication process (Schumacher et al., 2012).

The number of isomers is dependent on the number of internal inverted repeats. For instance, in HSV-1 there are 4 inverted repeats and consequently 4 isomers (called P, IS, IL and ILS); in the VZV genome, only two inverted repeats have been described, hence, only two isomers are detected (Thiry et al., 2005). These reassortments play a major role in viral evolution because they lead to differences in inter-genomic recombination rates between viruses. They explain why recombination occurs frequently in HSV-1 (all genomes contain mosaic patterns) but rarely in VZV (Norberg et al., 2011; Szpara et al., 2014).

2.2. Recombination in RNA viruses

Recombination rates are very different in the major categories of RNA viruses. In general, (−)ssRNA viruses show the lowest rates of recombination (McVean et al., 2002), whereas (+)ssRNA viruses frequently show high recombination rates. However, there are several relevant exceptions to this general pattern. For instance, recombination is rarely observed in Flaviviruses (Taucher et al., 2013), whereas recombination in orthomyxoviridae such as influenza A virus (Rabadan et al., 2008), which belong to the (−)ssRNA viruses, with rates 10 times higher than previously suspected in natural populations (Marshall et al., 2013). What biological, mechanistic or evolutionary processes determine the frequency of recombination in different viral species?

Two different mechanisms can generate recombinant genomes in RNA viruses. Reassortment only occurs in segmented RNA viruses, whereas recombination stricto sensu occurs in virtually all RNA viruses. The formation of a hybrid RNA sequence after inter-molecular exchange of genetic information between two nucleotide sequences results specifically from the latter. The best known model for this case is HIV-1 (Galetto and Negroni, 2005). Reassortment occurs when two or more viruses co-infect a single host cell and exchange discrete RNA segments by means of the packaging of segments from different origins in some progeny viruses. The most characteristic model is antigenic shift in influenza A virus (Nelson et al., 2008).

The international alarm as a consequence of the pandemic caused by influenza A(H1N1)pdm 2009 virus reinforced the efforts in surveillance and characterization of reasortments among viral segments. Nowadays, it is well-established that A(H1N1)pdm 2009 was the result of a genomic reorganization between two A(H1N1) swine viruses with at least four previously reassorted gene segments from avian, human and swine-adapted viruses (Zimmer and Burke, 2009). Moreover, the new highly-pathogenic influenza viruses from avian origin, H5N1 (identified in Hong Kong) and H7N9 (initially detected in Shanghai), are also the result of triple reassortments. The H5N1 virus infecting humans was the result of a reassortment among a quail H9N2 strain with segment-4 (hemagglutinin, HA) from goose H5N1 and with segment-6 (neuraminidase, NA) from teal H6N1 viruses (Guand et al., 2002). The H7N9 virus also resulted from a reassortment of avian influenza viruses where HA and NA had different origins; H7N3 was recovered from ducks and H7N9 from wild birds, respectively, whereas the internal genes were closely related to H9N2, which comes from chickens (To et al., 2013). These are excellent examples of how recombination can impact the evolution of segmented viruses, whereas homologous recombination seems to play little or no role in the evolution of influenza viruses in natural settings (Han and Worobey, 2011). An analysis of 8,307 full-length sequences of influenza A segments identified only two single recombiant sequences (Boni et al., 2010).

Reassortment does not require physical proximity of the parental genomes during replication. Nevertheless, two models for the assembly and packaging of RNA in budding virions have been proposed: the random incorporation model, which assumes that the virus does not differentiate among the different segments, and the selective packaging model, that results from reassortant viruses having a selective incorporation through the so-called packaging signal (Simon-Loriere and Holmes, 2011). There is increasing evidence supporting the theory of selective incorporation, because the number of reassortant genotypes, in experimental and natural conditions, is lower than predicted from theoretical considerations. Therefore, this model suggests that some type of constrictions at the protein or genomic level must exist. However little is known about the rules underlying this process. In an approach to define the most relevant genetic fragments for the packaging of influenza segments, Marsh et al. (2007) found that 45 and 80 nucleotides in the 3′ and 5′ ends respectively of the hemagglutinin (HA) gene along with an internal fragment of 15 nucleotide in the same segment must be conserved for reassortment to occur. Synonymous mutations introduced in these regions of HA resulted in a 3-fold reduction of packaging. Moreover, in a recent work aimed at understanding the exchange of gene segments between two different influenza A subtypes, human H3N2 and avian H5N2, the incorporation of avian HA in a human background required the co-selection of the avian M segment or five silent mutations in the human M segment (Essere et al., 2013). This suggests that sequence-specific interactions between RNA segments drive their selective incorporation during virion assembly (Essere et al., 2013) and that such RNA–RNA interactions would limit reassortment between divergent viruses.

The paradigmatic model of homologous recombination in Retroviruses is HIV. This virus presents one of the highest known recombination rates, with estimates ranging from 1.38 × 10⁻⁴ to 1.4 × 10⁻³ s/s/y (Shriner et al., 2004); however, these rates are not a common feature of all Retroviruses. For instance, the murine leukaemia virus, a γ-Retrovirus, has an estimated recombination rate 10–100 times lower than HIV, suggesting the existence of particular traits in HIV that explain its high recombination rate.

The HIV epidemic is caused by two different viruses, HIV-1 and HIV-2, with a genetic identity of around 50%. Group M includes the predominant circulating variants of HIV-1. It has been divided into 9 subtypes (with genetic distance between them at about 10–25%) and more than 60 circulating recombinant forms (CRFs) resulting from recombination events among different subtypes. Currently, HIV infections worldwide caused by CRFs represent around 20% of the total number (Hemelaar et al., 2011). The efficiency of recombination depends upon the length and identity of the templates (Zhang and Temin, 1994). For instance, recombination rates between two variants of different subtypes with similarities of 85% and 70% showed a 2-fold and 5-fold reduction, respectively, compared to those between variants of the same subtype (Chin et al., 2005). However, among the known CRFs, the recombinant forms between subtype B and F (BFs) disproportionately represent the highest diversity of different recombination events, even though these subtypes share higher homologies with other subtypes, suggesting that recombination among different subtypes is not random. There is also an inverse relationship between the divergence rate in the palindromic sequences and their recombination rate (Hussein et al., 2010), suggesting that similarly to HSV-1, the secondary structure of the RNA may play an important role in the recombination in HIV-1 (Negroni and Buc, 2000).

Recombination in HIV is intimately linked to replication (Delviks-Frankenberg et al., 2011), because the reverse transcription involves at least two template switching events (minus-strand DNA transfer and plus-strand DNA transfer). The process of reverse transcription in HIV is very complicated (Coffin et al., 1997), and
recombination likely takes place during minus-strand DNA synthesis (Onafuwa-Nuga and Telesnitsky, 2009). The model proposed to explain retroviral recombination is the dynamic copy choice model (Hwang et al., 2001), which proposes that a balance between the polymerase and RNase H activity of retrotranscriptase (RT) influences template switching. When polymerase activity is low (and displays high RNase activity) the dissociation of RT from the nascent strand of DNA increases and, consequently, the likelihood of exchanging template strands increases. Conversely, when the RNase activity is low (maintaining high RT activity), opportunities for recombination are low because the RT is linked to the growing strand. It is estimated that, during reverse transcription, the RT is dissociated from the template at least 8 times per genome (Pathak and Hu, 1997). This dissociation is responsible for 3–12 template switches per genome and replication cycle (Zhuang et al., 2002). Interestingly, some RT mutations involved in antiretroviral resistance, such as K65R in RT (which confers resistance to tenofovir), can also increase template switching and, subsequently, viral evolution (Nikolenko et al., 2004).

Recombination breakpoints are not distributed at random along the HIV-1 genome (Archer et al., 2008), suggesting that recombination in HIV-1 may be relatively predictable. Two hypotheses have been suggested to explain this. First, the dominance of purifying selection in some areas of the HIV-1 genome define regions where functional contractions appear to be particularly strong (Simon-Loriere et al., 2009). Under the second hypothesis, which does not exclude the former, there might be a direct link between enhanced recombination rates and the formation of secondary RNA structure, because these RNA structures induce stops, stalling or delays in DNA synthesis, thus favouring strand invasion (Simon-Loriere and Holmes, 2011). Interestingly, the presence or absence of the viral chaperone nucleocapsid modulates RNA secondary structure, revealing an evolutionary role for these proteins (Negroni and Buc, 2000); this effect is related to HIV-1 nucleocapsid promoting dimeric G-quartet formation (Shen et al., 2011).

3. Estimation of recombination

The evolutionary relevance of recombination in viruses leads to a need for accurate inferences of recombination rates and breakpoints (Robertson et al., 1995). The inference of recombination events can be applied to understand evolutionary processes such as genome dynamics – since recombinant genomes have been associated with changes in phenotype and fitness (Monjane et al., 2011; Springman et al., 2012) – to carry out genetic mapping (Anderson et al., 2000), or to perform ‘realistic’ phylogenetic analysis that accommodate recombination (Schierup and Hein, 2000a; Arenas and Posada, 2010c; Arenas, 2013b).

Currently, a variety of analytical tools exist to estimate recombination rates. Despite differences in their performance (Posada and Crandall, 2001; Wiuf et al., 2001), the choice of an appropriate tool still remains a difficult task from a practical point of view. The choice depends on different considerations, such as the type of recombination analyses, computational costs or genetic markers (specific tools exist for specific markers). A few articles have provided comprehensive reviews of computer tools that infer recombination (Posada et al., 2002; Awadalla, 2003; Martin et al., 2011a,b), and David Robertson’s group (http://bioinf.man.ac.uk/recombination/programs.shtml) has compiled a list of programs for this task. In this section, we present an overview of the main frameworks to analyze recombination in viruses.

3.1. Statistical tests to detect recombination

A variety of methods allow us to detect the presence of recombination (Posada et al., 2002; Wiuf et al., 2001). However, their performances vary depending on the amount of recombination and the genetic diversity of the dataset. In general, these methods are more accurate with increasing recombination rates and at high levels of diversity. In particular, a minimum nucleotide diversity of 5% and several recombination events are required to obtain enough statistical power for analysis (Posada et al., 2002). Viruses are characteristically capable of such features (Bowen et al., 2000; Monjane et al., 2011; Kapusinszky et al., 2012). Additionally, rate variation among sites, which can be easily observed in virus data (Worobey, 2001), can also affect the detection of recombination by increasing false positives (Worobey, 2001; Posada, 2002). As a consequence, methods that consider substitution patterns (Stephens, 1985; Gibbs et al., 2000) and rate variation among sites are more powerful than simplistic, topology-based methods that do not consider these factors (Husmeier and Wright, 2001). Homoplasy tests are recommended to analyze low diversity datasets (Bruen et al., 2006), whereas methods such as maximum chi-square (Smith, 1992) or RDP (Martin et al., 2010) are robust to high levels of diversity (Posada, 2002). Additionally, the pairwise homoplasy index (PHI), implemented in the PhiPack package (Bruen et al., 2006), can accurately distinguish recurrent mutation from recombination in a wide variety of circumstances.

Recombination results in the presence of fragments with different most recent common ancestors on the same hereditary molecule. The molecular phylogenies derived from each of these fragments will necessarily be different. Consequently, the ultimate test for recombination should be based on revealing a statistically significant lack of congruence between the phylogenies of the genome arrays involved. This can be tested explicitly (Coscoillá and González-Candelas, 2007; Sentandreu et al., 2008) using phylogenetic incongruence tests (Shimodaira and Hasegawa, 1999; Strimmer and Rambaut, 2002) once the putative recombinant fragments have been identified using several of the methods detailed above. However, as previously noted, phylogenetic inference must consider the many different intricacies that the actual evolutionary processes pose (González-Candelas et al., 2011). In conclusion, the use of several recombination tests followed by explicit testing of phylogenetic incongruence is recommended to generate a more robust detection of recombination (Posada, 2002).

3.2. Inference of recombination breakpoints

The identification of recombination breakpoints in viruses is useful to detect circulating recombinant forms (Minin et al., 2007; Archer et al., 2008), to study adaptive recombination (Monjane et al., 2011), to analyze recombinant data (Kosakovsky Pond et al., 2006a; Arenas and Posada, 2010c; Arenas, 2013b), and to infer the underlying mechanisms of recombination. Recombination is generally non-homogeneous along genomes (Minin et al., 2007; Monjane et al., 2011). A number of programs exist to detect recombination breakpoints (Martin et al., 2011a,b), but some of them are more commonly used because of their high accuracy and user-friendly environment. For example, the Datamonkey web server (http://www.datamonkey.org/) (Kosakovsky Pond and Frost, 2005) and the Hyphy package (Kosakovsky Pond et al., 2005) implement the GARD algorithm (Kosakovsky Pond et al., 2006b), which is based on phylogenetic discordance. The program RDP3 can also detect recombination breakpoints accurately and presents a friendly graphical interface (Martin et al., 2010). Finally, the SCUET tool implements a phylogenetic algorithm for mapping the location of breakpoints and assigning parental sequences in recombinant strains (Kosakovsky Pond et al., 2009). This method also allows for automatically subtyping viral sequences (Yabar et al., 2012).
3.3. Estimating recombination rates

A number of recombination rate estimators have been developed (Posada et al., 2002; Martin et al., 2011a,b). Here, we highlight the widely used computer programs OmegaMap (Wilson and McVean, 2006), Lamarc (Kuhner, 2006) and LDHat (McVean et al., 2004). The OmegaMap program is based on the product of approximating conditionals (PAC) likelihood (Li and Stephens, 2003) and infers population recombination rate and selection (by the nonsynonymous to synonymous rates ratio, dN/dS), allowing both estimates to vary along the sequence. Lamarc relies on a Markov chain Monte Carlo (MCMC) coalescent genealogy sampler and estimates a variety of population genetic parameters such as recombination, migration and exponential growth rates. Finally, LDHat computes the likelihood function under a coalescent-based model to estimate recombination rates. Overall, the accuracy of these analytical tools depends on the amount of genetic diversity. In general, most methods are less robust under low levels of genetic diversity, with Lamarc showing the best performance in such conditions (Lopes et al., 2014).

An alternative and emerging methodology to estimate recombination rates uses the approximate Bayesian computation (ABC) approach (Beaumont, 2010). ABC is based on extensive computer simulations (Arenas, 2012, 2013a; Hoban et al., 2012) and can consider the influences of different evolutionary processes on parameter estimation. Importantly, ABC methods make several important assumptions, such as a total dependence on the user-specified prior distributions, the need for extensive and realistic computer simulations or informative summary statistics, which may hamper their application. An interesting example was performed by Wilson et al. (2009), in which they applied ABC to estimate recombination rates among other evolutionary parameters (such as mutation and dN/dS). This methodology is likely to be used in next year for the analysis of highly recombinant viruses, such as those recently described by Lopes et al. (2014) with HIV-1 data.

3.4. The impact of next-generation sequencing (NGS) on the estimation of recombination

NGS technologies are currently used to address a number of biological problems by enabling the comprehensive analysis of genomes and transcriptomes (Shendure and Ji, 2008). These technologies generate many short reads with associated quality values (Metzker, 2010), which are later assembled into genomes or contigs. However, the assembly of fragments may produce artificial recombination signals with false polymorphisms (Martin et al., 2011a,b; Prosperi et al., 2011). A methodological alternative to avoid this source of error was proposed by Johnson and Slatkin (2009). They presented a two-site composite likelihood estimator of the population recombination rate that operates on genomic assemblies, in which each sequenced fragment derives from a different individual. More recently, a variety of new strategies to accurately estimate recombination from NGS data, based on both novel experimental methodologies (Grossmann et al., 2011; Beerenwinkel et al., 2012) and Bayesian approaches have emerged (Castillo-Ramirez et al., 2012; Zhang, 2013). These strategies are expected to overcome some of the initial problems and to generate more accurate estimates of recombination rates across the genome. An example was provided by Schlub et al. (2010), who accurately measured recombination in HIV-1 genomes.

4. The impact of recombination

Recombination creates new genetic combinations that would be hardly attainable by mere mutational processes, even for fast evolving, large populations of viruses. These offer new possibilities for exploring the phenotypic and adaptive landscapes with consequences at different levels and scales, from the purely biological to methodological problems arising when the origin of these recombining variants is not considered. In this last section, we provide a brief overview of some of the consequences of recombination in viruses.

4.1. The role of recombination with respect to mutation as a source of genetic diversity

Mutation and recombination are the two main sources of genetic diversity. Mutations introduce changes in nucleotide states, and therefore new variants. Recombination allows the movement of variants across genomes to produce new haplotypes. Therefore, recombination does not create new mutations (at the nucleotide level) but introduces new combinations of the existing ones. As a result, several mutations previously incorporated in a genomic region can be transferred simultaneously to another region by a single recombination event. Interestingly, at the codon level, recombination can also alter the codon state by intra-codon recombination events (Arenas and Posada, 2010a), a process observed in viruses (Behura and Severson, 2013). Nevertheless, the fate of these genetic changes will be ultimately determined by natural selection and genetic drift.

There are plenty of virus-based studies of large recombination rates that increase genetic diversity and create novel lineages and strains (He et al., 2010; Combelas et al., 2011; González-Candelas et al., 2011) or increase phenotypic variation (Combelas et al., 2011). Indeed, recombination has promoted the emergence of new pathogenic recombinant circulating viruses (Combelas et al., 2011; González-Candelas et al., 2011).

4.2. The influence of recombination on selection estimation

Evolutionary analyses of highly recombinant viruses usually include the estimation of dN/dS as a measure of molecular adaptation (Pérez-Losada et al., 2009, 2011; Chen and Sun, 2011) and it is useful to understand the role of particular amino acids in the protein activity (Chen and Sun, 2011).

However, recombination may bias the estimation of this parameter at the codon level (Anisimova et al., 2003; Kosakovskyy Pond et al., 2006a; Arenas and Posada, 2010a) as a consequence of ignoring recombination in phylogenetic analysis (Schierup and Hein, 2000a, 2000b; Posada and Crandall, 2002). In particular, recombination increases the number of false positively selected sites (PSS) (Anisimova et al., 2003; Kosakovskyy Pond et al., 2006a; Arenas and Posada, 2010a). By contrast, this effect was not observed in the estimation of dN/dS at the global level (entire sequence) even when intra-codon recombination was present (Arenas and Posada, 2010a, 2012). Consequently, recombination should be taken into account when estimating local dN/dS. A proposed methodology to perform this task – implemented in Datamonkey and Hyphy – starts with the detection of recombination breakpoints, followed by the reconstruction of a phylogenetic tree for each recombinant fragment and, finally, the estimation of dN/dS by using the appropriate tree for each region (Kosakovskyy Pond et al., 2006a) (Fig. 2). This methodology is commonly used to analyze viral data (Pérez-Losada et al., 2009, 2011).

4.3. The role of recombination in the design of centralized viral genes

A common strategy for developing vaccines for viruses with high mutation and recombination rates is the application of centralized genes (Doria-Rose et al., 2005; McBurney and Ross, 2007; Arenas and Posada, 2010b). The goal of this strategy is to
minimize the genetic distance to the target circulating viral strains and thus consider their immunogenic features.

Three types of centralized genes have been commonly used in the design of HIV-1 vaccine candidates, namely (i) the consensus sequence \[\text{CON}\] (e.g., Ellenberger et al., 2002; Novitsky et al., 2002), based on the most abundant states, (ii) the ancestral sequence \[\text{ANC}\] (Doria-Rose et al., 2005; Kothe et al., 2006), the common ancestral sequence in a phylogeny, and (iii) the center-of-tree sequence \[\text{COT}\] (Rolland et al., 2007), the sequence of a point in the phylogeny where the evolutionary distance to each tip node is minimized. The three methods produced promising induction of T-cell responses (Frahm et al., 2008), but all were modest and partially effective in combating HIV-1 diversity. Factors like the diversity level of the viral strains (Vidal et al., 2000), sampling procedure (samples should be as representative as possible of the viral population), or the computational design of the centralized genes (Arenas and Posada, 2010b) seem to impact the performance of the centralized genes strategy.

Recombination has been often ignored in the estimation of centralized sequences. The estimation of ANC and COT genes is based on phylogenetic inference, which can be severely affected by recombination (Schierup and Hein, 2000a; Posada and Crandall, 2002). Moreover, Arenas and Posada (2010c) have shown that recombination incorporates errors in the estimation of ANC sequences and such errors may influence the prediction of ancestral epitopes and N-glycosylation sites. These authors also showed that recombination affects the estimation of COT sequences, although its impact is less severe in the estimation of ANC sequences (Arenas and Posada, 2010b). Consequently, recombination should be always considered when estimating centralized sequences. A proposed methodology consists of estimating recombination breakpoints followed by an independent phylogenetic analysis and ancestral sequence reconstruction for each recombinant fragment (Arenas and Posada, 2010b,c).

### 4.4. Population structure and recombination

Viral genetic variants are not distributed randomly in space (García-Arenal et al., 2001; Holmes, 2004). Usually, there is a geographic range within which individual particles are more closely related to one another than those randomly selected from the general population. This is described as the extent to which the viral population is genetically structured. This can be assessed quantitatively through population differentiation coefficients (Meirmans and Hedrick, 2011). If viral particles are sampled at random from a population where viruses freely disperse among hosts (no population structure), then two sampled copies are expected to follow predictable proportions which only depend on the overall allele frequencies of the population (Gillespie, 2004). However, if viruses are distributed in patches (population structure), the particles found in different areas will follow a different proportion than that predicted by the global allele frequencies. The main effect of population structuring would be an apparent excess of samplings with two exact copies of the same RNA genome and a deficit of cases with different RNA genome copies. This relationship between the departure from random combinations of genetic variants and population structure is known as the Wahlund effect (Hartl and Clark, 2007).

One important consequence of population structuring is the generation of linkage disequilibrium (Garnier-Géré and Chikhi,
2013). When the alleles present at one locus are independent of those present at a second locus, the viral population is said to be at linkage equilibrium. However, if a particular allele at one locus is found on the same chromosome with a specific allele at a second locus more (or less) often than expected under independent segregation then these loci are in linkage disequilibrium (LD). When real examples are considered, different viral haplotypes (combinations of alleles or genetic markers) are rarely distributed as conveniently as expected under random mixing of alleles (Verschoor et al., 2004). In many cases, some viral haplotypes appear more often than expected under linkage equilibrium, so that accounting for LD becomes of great importance when making inferences about recombination in natural populations.

4.5. Interaction of population-level processes and recombination

Linkage disequilibrium is automatically created when a new mutation occurs on a chromosome that carries a particular allele at a nearby locus, and these allelic associations are broken down by recombination. In the absence of other processes, the degree of association between alleles in a sample of chromosomes is simply a function of the age of the mutation and the recombination rate, so that the faster the decay in LD, the larger the recombination rate (Fig. 3). Therefore, the speed in the decay of LD values within a viral population can be used to make inferences on the population-level recombination rate.

In real populations, the observed levels of LD are influenced by the interaction of recombination rates with a number of other biological factors, including viral population sizes, natural selection and population structure (limited dispersal). An understanding of how different biological processes can affect LD is essential in the population genetics analysis of viruses (Simon-Loriere and Holmes, 2011). The effective recombination rate in natural populations is dependent on the effective population size and population stochasticity may impact the efficiency of recombination (McVean et al., 2002). Indeed, Frost et al. (2000) found that differences in the rate of evolution of lamivudine resistance between infected individuals arose through genetic drift affecting the relative frequency of different genetic variants prior to therapy. Their results showed that stochastic effects can be significant in HIV evolution, even when there are large fitness differences between mutant and wild-type variants. Epistatic interactions between genetic variants in different loci may also counteract the effect of recombination over allele combinations. In fact, epistasis between alleles has been detected in the HA and NA genes of influenza (Kryazhimskiy et al., 2011).

The presence of population structure can be as important as natural selection in creating disequilibrium against the effect of recombination. To illustrate this point, we will assume a simple model where two viral strains expand through a linear space, with viral particles being able to migrate only between nearby populations (Kimura and Weiss, 1964). The populations located in the extremes of the distribution are fixed with different strains (haplotypes) so that the population in one end is formed by 100% viral particles of haplotype AB and the one in the other end only presents haplotype ab. It is expected that once viral particles coming from the extreme populations appear in the same population, new recombinant haplotypes (Ab and aB) will appear by recombination.

Under this simple model, there exists a direct relationship between the amount of LD generated by migration and the differences in gene frequencies for each locus in the parent populations, the recombination rate between loci and the dispersal rate which can be used to make inferences on the population recombination rate in a spatially-structured population.

Given that gene flow contributes to increased LD and simultaneously disequilibrium modifies the effect of natural selection over nearby loci, population structure and natural selection may interact in counterintuitive ways. In the case that some recombinant forms present lower fitness than others, the outcome of secondary contact between both viral populations will vary from a stable co-existence of two strains in intermediate populations to the replacement of the original strains by a new recombinant virus (Fig. 4). The exact spatial distribution of different viral particles (recombinant or non-recombinant) will depend on the recombination rate.

Suarez et al. (2004) provide an example of recombination resulting in a strong virulence shift in populations of avian influenza virus. Sequence analysis revealed minor differences between the viruses, except at the hemagglutinin (HA) cleavage site. The insertion, likely occurring by recombination between the HA and nucleoprotein genes of the LPAI virus, resulted in a virulence shift. While the virus without the insert caused no illness or death in experimentally infected birds, viruses with the insert caused severe disease and death (Suarez et al., 2004). The increase in virulence between viruses with and without the insert in this outbreak was readily apparent and could have been triggered by increased dispersal of the virus within Chile.

5. Conclusions

The extent and distribution of recombination is a topic of great interest. The analysis of linkage disequilibrium can be very revealing about the evolution of viral populations. As pointed out in this review, phylogenetic and population genetic studies may enable us to learn more about the interactions of recombination, migration and natural selection. A firm knowledge of the principles of recombination in population genetics is an important component of
understanding and interpreting results from genetic studies of viruses. Virus studies are increasingly characterized by large amounts of genetic data, whereas viral populations have often evolved under recombination. Hence, the application of population genetics models and advanced statistical analysis accounting for recombination is likely to become a cornerstone in the dissection of genetic variation in viral populations.

Acknowledgements

We thank Michel Tibayrenc for his invitation and encouragement to write this review and two anonymous reviewers for their comments. We also thank Rebecca G. Simmonds for her kind permission to write this review and two anonymous reviewers for their comments.

References

Anderson, T.J.C., Haubold, B., Williams, J.T., Estrada-Franco, Richardson, L., Molina-Rivera, R., Mukti, J., Bharathkura, S., French, N., Weinhold, J., Velez, I.D., Brockman, A.H., Nosten, F., Ferreira, M.U., Day, K.P., 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Mol. Biol. Evol. 17, 1467–1462.

Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164, 1229–1236.

Acker, J., Pinney, J.W., Fan, J., Simon-Loriere, E., Arts, E.J., Negroni, M., Robertson, D.L., 2008. Identifying the important HIV-1 recombination breakpoints. PLoS Comput. Biol. 4, e1000178.

Arenas, M., 2012. Simulation of molecular data under diverse evolutionary scenarios. PLoS Comput. Biol. 8, e1002495.

Arenas, M., 2013a. Computer programs and methodologies for the simulation of DNA sequence data with recombination. Front. Genet. 4, 9.

Arenas, M., 2013b. The importance and application of the ancestral recombination graph. Front. Genet. 4, 206.

Arenas, M., Posada, D., 2010a. Coalescent simulation of intracodon recombination. Genetics 184, 429–437.

Arenas, M., Posada, D., 2010b. Computational design of centralized HIV-1 genes.Curr. HIV Res. 8, 613–621.

Arenas, M., Posada, D., 2010c. The effect of recombination on the reconstruction of ancestral sequences. Genetics 184, 1133–1139.

Arenas, M., Posada, D., 2012. Simulation of coding sequence evolution. In: Cannarozzi, G.M., Schneider, A. (Eds.). Oxford University Press, Oxford, pp. 126–132.

Austermann-Busch, S., Becher, P., 2012. RNA structural elements determine ancestral sequences. Genetics 184, 1133–1139.

Awadalla, P., 2003. The evolutionary genomics of pathogen recombination. Nat. Rev. Genet. 4, 50–60.

Baltimore, D., 1971. Expression of animal virus genomes. Bacteriol. Rev. 35, 235–262.

Baltimore, D., 1971. Expression of animal virus genomes. Bacteriol. Rev. 35, 235–262.

Bowden, R., Sakaoaka, H., Donnelly, P., Ward, R., 2004. High recombination rate in herpes simplex virus type 1 natural populations suggests significant co-infection. Infect. Genet. Evol. 4, 115–123.

Brennan, V.K., Hu, W.S., 2010. Patterns of human immunodeficiency virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. J. Virol. 84, 7561–7566.

García-Arenal, F., Fraile, A., Malpica, J.M., 2001. Variability and genetic structure of hepatitis C virus. Trends Microbiol. 22, 354–364.

Garnier-Géré, P., Chikhi, L., 2013. Population subdivision, Hardy–Weinberg equilibrium and genomic novelty in RNA viruses. Genome Res. 17, 1496–1504.

Galetto, R., Negroni, M., 2005. Mechanistic features of recombination in HIV. AIDS 19, 447–456.

Gallo, A., Kearney, D., Moudarré, F., Coquin, J.M., Pathak, V.K., Hu, W.S., 2005. Patterns of human immunodeficiency virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. J. Virol. 84, 7561–7566.

García-Arenal, F., Fraile, A., Malpica, J.M., 2001. Variability and genetic structure of hepatitis C virus. Trends Microbiol. 22, 354–364.
Martin, D.P., Lemey, P., Posada, D., 2011. Recombination hotspots in human immunodeficiency virus type 1 genomic RNA. Mol. Ecol. Resour. 11, 943–955.

McBurney, S.P., Ross, T.M., 2007. Developing broadly reactive HIV-1/AIDS vaccines: a review of polyvalent and centralized HIV-1 vaccines. Curr. Pharm. Des. 13, 1957–1964.

McVean, G., Awadalla, P., Fearnhead, P., 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 162, 1279–1299.

McVean, G., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., Donnelly, P., 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304, 581–584.

Meirmans, P.G., Hedrick, P.W., 2011. Assessing population structure: FST and related measures. Mol. Ecol. Resour. 11, 5–18.

Metzker, M.L., 2010. Sequencing technologies – the next generation. Nat. Rev. Genet. 11, 31–46.

Meursens, F., Schyns, F., Keil, G., Muyilken, B., Vanderplaschen, A., Galilgo, T., E., 2004. Superinfection prevention recombinant of the alpha/theta herpesvirus bovine herpesvirus 1. J. Virol. 78, 3872–3879.

Minin, V.N., Dorman, K.S., Fang, F., Suchard, M.A., 2007. Phylogenetic mapping of recombination hotspots in human immunodeficiency virus 1 via spatially smoothed change-point processes. Genetics 175, 1773–1785.

Monjane, A.L., van der Walt, E., Varsani, A., Rybicki, E.F., Martin, D.P., 2011. Recombination hotspots and host susceptibility modulate the adaptive value of recombination during mumps virus strain evolution. BMC Evol. Biol. 11, 350.

Muyaart, L., Tang, K.W., Elias, P., 2011. Recombination and replication of herpes simplex virus DNA. J. Biol. Chem. 286, 15619–15624.

Negrouk, C.B., Baric, R., 2000. Frequency of reverse transcription by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones. Proc. Natl. Acad. Sci. USA 97, 6385–6390.

Nelson, M.L., Viboud, C., Simonsen, L., Bennett, R.T., Griesemer, S.B., George, K., Taubenberger, J.K., Spor, D., Litovitz, T.L., Holmes, E.C., 2008. Multiple recombination events in the evolutionary history of H1N1 influenza virus since 1918. PLoS Pathog. 4, e1000122.

Nikolenko, G.N., Svarovskaia, E.S., Delviks, K.A., Pathak, V.K., 2004. Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency. J. Virol. 78, 8761–8770.

Norberg, P., Tyler, S., Severini, A., WHitley, R., Liiijequist, J.A., Bergstrom, T., 2011. A genome-wide comparative evolutionary analysis of herpes simplex virus type 1 and varicella zoster virus. PLoS One 6, e22527.

Novitsky, V., Smith, U.R., Gilbert, M., McLane, M.F., Chigwedere, P., Williamson, C., Ndung’u, T., Klein, I., Chang, S.Y., Peter, T., Thor, I., Foley, B.T., Galewsky, S., R, Oshie, R., Faden, R., Gasser, G., 2011. Combinatorial analysis and algorithms for phylogeny estimation. J. Mol. Evol. 54, 396–402.

pathak, V.K., Hu, W.S., 1997. “Might as well jump!” Template switching by reverse transcriptase increase template-switching frequency. J. Virol. 71, 3872–3879.

Peeling, R.W., Berman, P.W., 2011. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. PLoS One 6, e16902.

Perez-Losada, M., Oborowicz, M., Staszewski, E., Berman, P.W., 2011. Phylogenetics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. PLoS One 6, e16902.

Perrin, M., Bello, P., Posada, D., 2011. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. PLoS Comput. Biol. 7, e1002161.

Peters, M., Sarlis, M., Delarue, C., Pannetier, C., 2011. Combinatorial analysis and algorithms for phylogeny estimation. J. Mol. Evol. 54, 396–402.

Phylogenetics of HIV-1 subtype B after 12 years after preeinfection. AIDS 24, 12–12.

Piantadosi, A., Chohan, B., Chovan, V., McClelland, R.S., Overbaugh, J., 2007. Chronic HIV-1 infection frequency fails to protect against superinfection. J. Virol. 81, 1377–1382.

Posada, D., 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol. Biol. Evol. 19, 708–717.

Posada, D., Crandall, K.A., 2002. The effect of recombination on the accuracy of phylogenetic analysis. Semin. Virol. 13, 141–150.

Posada, D., Crandall, K.A., 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. 36, 75–97.

Prosperi, M., Prosperi, L., Bruselles, A., Abbate, I., Rozera, G., Vincenti, D., Solmone, M.C., Capobianchi, M.R., Ulivi, G., 2011. Combinatorial analysis and algorithms for phylogeny estimation. J. Mol. Evol. 54, 396–402.

Redd, A.D., Colinson-Streng, A., Martens, C., Rockles, S., Mullis, C.E., Manucci, J., Tobias, A.A., Selig, E.J., Laeender, O., Sambandam, N., 2011. Identification of HIV superinfection in seronegative couples in Rajkia, Uganda, by use of next-generation deep sequencing. J. Clin. Microbiol. 49, 2859–2867.

Redd, A.D., Quinn, T.C., Tobias, A.A., 2013. Frequency and implications of HIV superinfection. Lancet Infect. Dis. 13, 622–628.

Rennekamp, A.J., Lieberman, P.M., 2010. Initiation of lytic DNA replication in Epstein-Barr virus: search for a common family mechanism. Future Virol. 5, 65–83.
