Chapter 5
HIV and SIV Evolution

Brian T. Foley

Keywords Immunodeficiency • Evolution • Selection • Phylogenetics • Epidemiology • AIDS

Core Message
Human immunodeficiency virus type 1, the virus responsible for the AIDS pandemic, is just one of a large group of primate lentiviruses. Although many aspects of the evolution of these viruses is known in great detail, other aspects remain shrouded in mystery due to large timescales and a lack of fossil records. Retroviruses in general are not primarily noted for their neurological effects, but a great many, including the lentiviruses, cause neurological problems in at least a subset of infected hosts. The lentiviruses are primarily noted as the cause of immunodeficiency, but they also cause neurological damage. Despite the vast genetic distances between the groups of primate lentiviruses, many aspects of their biology remain remarkably similar.

5.1 Introduction

The primate immunodeficiency viruses (PIVs) comprise a diverse group of lentiviruses, all derived from a single common ancestor, which infect old world monkeys and apes. Many, but not all, species of old world monkeys each carry their own lineage of PIV and have apparently coevolved together for as much as several million years. The evidence suggesting such an ancient origin includes the fact that some PIV lineages recapitulate their host evolution. For example, the African green monkeys (**family**, Cercopithecidae; **genus**, *Chlorocebus*; **species**, *Ch. aethiops*, *Ch.*

B.T. Foley, PhD (✉)
Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: BTF@LANL.gov

P. Shapshak et al. (eds.), *Global Virology II - HIV and NeuroAIDS*,
DOI 10.1007/978-1-4939-7290-6_5

cynosurus, *Ch. djamdjamensis*, *Ch. pygerythrus*, *Ch. sabaeus*, and *Ch. Tantalus*) are known to carry at least three lineages of PIV (SIV-sabaeus, SIV-tantalus, and SIV-vervet) which are more closely related to each other than to lineages of PIVs found in other primates. Likewise the chimpanzees and gorillas carry PIVs that are related to each other and different than those carried by other primates [1–3] (Fig. 5.1).

The origins and evolution of retroviruses are not yet known. Mammals and many other vertebrates carry dozens of endogenous retroviruses in their germ line DNA. Nearly all endogenous retroviruses have a simple LTR-*gag*-*pol*-*env*-LTR genome, whereas the T-cell leukemia viruses and the lentiviruses have accessory genes such as *tat, rev, vif*, and *nef.* Several mammal species have now been found with endogenous retroviruses that appear to be ancestral to the lentiviruses, but they all lack most of the accessory genes [4]. Thus, the origins of lentiviruses are also somewhat unknown, but the evidence suggests that they have been infecting mammals for more than 20 million years [4–6]. In addition to primates being infected with primate immunodeficiency viruses, the felines, bovines, equines, ovines, and caprines each have lentiviruses which also infect T cells and cause immunodeficiencies and other pathologies.

Retroviruses in general tend to be quite host specific, and the lentiviruses are typical in this regard. Most primate lentiviruses cannot replicate in human cells primarily due to the host APOBEC enzymes but also due to other innate host defense mechanisms

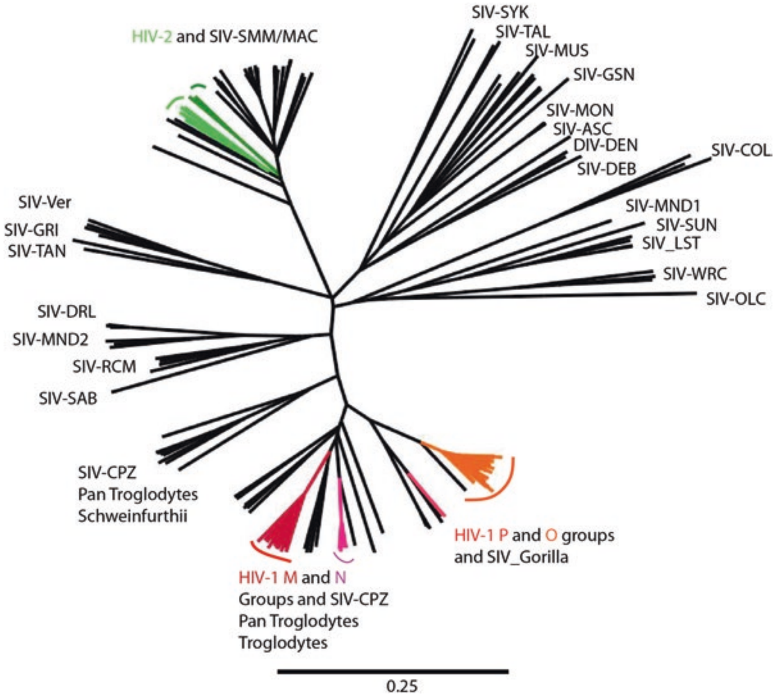

Fig. 5.1 Phylogenetic tree of primate lentiviruses. Phylogenetic tree constructed by maximum likelihood method using complete genomes of primate lentiviruses. Viruses infecting and spreading in humans are colored, shades of red and orange for HIV-1 viruses derived from chimpanzees and gorillas, and shades of green for HIV-2 viruses derived from sooty mangabeys. Several other viruses were transferred from sooty mangabeys to humans but have not been noted to spread in humans

[7, 8]. Other factors above the cellular level also influence the relationship of a virus to a particular host species or geographical location. One factor is the frequency of interaction between infected individuals of one species with other species. It is quite likely that increased hunting of chimpanzees and gorillas with modern weapons lead to transfers of simian immunodeficiency viruses (SIVcpz and SIVgor) from these species into humans. Another possibility is that humans have been exposed to these SIVs many times in the past, but it took higher human population densities and/or blood exposures via transfusions and needle reuse to jump-start the human epidemic [9, 10]. Iatrogenic transmission of HIV has been greatly reduced in the developed world, but unsafe medical practices continue in some developing nations [11].

After a cross-species transfer event takes place, such as transfer from chimpanzee into humans, the virus evolves from the single point source of the transfer into lineages that diversify from that single common ancestral virus to form what is known as a "star phylogeny" except when multiple viruses infect any one individual and then recombine. Inter-lineage recombination creates a network or web relationship between lineages rather than a perfect star or tree. Within any one infected individual, recombination within the viral swarm or population is a highly frequent occurrence because each virion packages two ssRNA viral genomes and the reverse transcriptase skips between the two genomes when synthesizing the first strand complementary DNA. There is some, but not a complete, blockage of multiple infections of a single cell with more than one virion, but many of the billions of infected cells in an individual will be multiply infected. The vast majority of humans infected sexually are singly infected from a single-donor sex partner. In IV drug user communities, multiple infections from more than one donor are common. Within a single community, whether sexual or IV drug user, it is more common for a set of highly related viruses within a single subtype than for multiple subtypes of virus to be circulating (Fig. 5.2).

Selection pressure on virions within each infected individual is quite extreme. Both CTL-mediated and antibody (B-cell-mediated) immune responses vastly reduce the viral load from as high as the tens of millions of virions in the first weeks of infection to a "set point," a typical reduction four orders of magnitude lower than the peak. That is, 99.99% of virions are being removed, and only 0.01% survives to reproduce in each round of replication. This can result in nearly complete replacement of one population, be it one that is sensitive to attack of a single CTL epitope or one that is sensitive to a drug, by a mutant-resistant population, in a matter of weeks [12, 13]. The host immune system attacks many virus epitopes at once, some CTL and some antibody, and also evolves over time to become more efficient at attacking the virus. Drugs on the other hand can only target one viral factor per drug and do not evolve over time. Using multiple drugs simultaneously is thus critical to success. Combination antiretroviral therapy (cART) goes by various names such as highly active retroviral therapy (HAART). The most common combinations include drugs that target two sites in the reverse transcriptase enzyme and one target in the protease enzyme.

Selection pressure driving the virus to change over time such as to evade the host immune system or to evade (become resistant to) drugs is known as positive selection. The other major selection pressure on the virus is for the virus to be "highly fit" in terms of replication rate within each individual and in terms of being transmissible to other individuals.

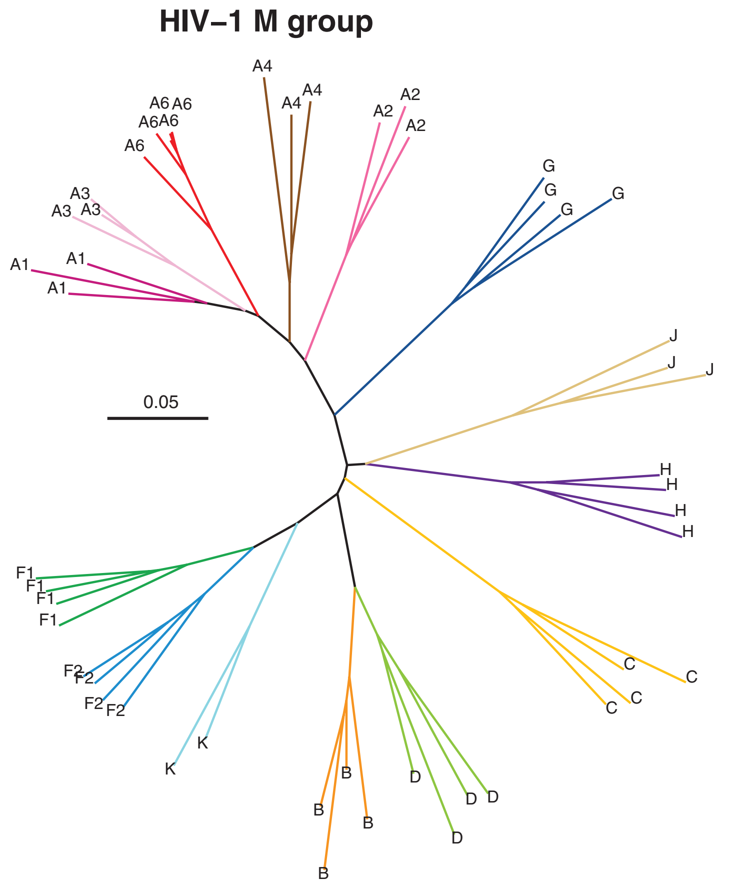

Fig. 5.2 HIV-1 M group phylogenetic tree. Phylogenetic tree constructed by maximum likelihood method using complete genomes of HIV-1 M group major subtypes. CRF26_AU is included because the authors defined the A-like regions as subsubtype A5. It is noteworthy that subtypes B and D are related to each other, and recent standards of nomenclature and classification would be subsubtypes (as are F1–F2, A1–A6) rather than full subtypes, but they were classified very early in the study of HIV-1 as subtypes

5.2 Mutations

5.2.1 Bias

Retroviruses are observed to evolve at a rate nearly ten million-fold faster than mammalian nuclear genes. The exact rate of evolution depends on the region of the genome, with the envelope gene evolving faster than the polymerase gene, for example, due to differences in selection pressures. Moreover, the rate of evolution is a function of both the mutation rate and the selection pressures on the genes under study. It is generally accepted that mutations occur at random and selection creates differences in the observed patterns of change over time [14], but there are several biases in the mutation step of the process as well. Transition mutations (A <-> G and C <-> T) occur much more frequently than transversion mutations (A <-> T, A <-> C, G <-> T, and G <-> C) for a number of reasons, and for retroviruses, the G to A mutation rate is far higher than any other rate including A to G. Also, the context of bases surrounding a given base can influence the mutation rate. One well-known example is that C followed by G is prone to mutation to T because of DNA methyltransferases. Methylcytosine can deaminate to uracil, which then pairs with adenosine so that the cytosine is replaced by thymidine in the next round of replication. Most mammalian DNA is depleted of CpG dinucleotides except for regions known as CpG islands [15], and many viruses also have a lack of CpG.

With human infections we can almost never know the exact sequence of the infecting virus, and we therefore always observe changes over time that are at least partially influenced by selection as well as the underlying mutations. In experimental infections of macaques, a single infectious molecular clone with a known genome sequence can be used, but even in this case, lethal mutations are eliminated, and with sampling over time, we observe evolution that is influenced by selection. Other ways to study the fidelity of HIV reverse transcriptase and thus mutations without selection include using single-round replication vectors in vitro or in cell cultures and sequencing cDNA made from RNA templates of known sequence after a single round of reverse transcription.

The base composition of lentiviruses is A-rich and C-poor with the A:C ratio close to 2:1. The base composition bias is fairly uniform across the genome and even more uniform between different viruses. The same regions that are less A-rich in one primate virus are also less A-rich in others. The result of this is that many so-called “silent” sites in the genome, third codon positions where any of the four bases can be used to encode the same amino acid when translated, appear to be under strong purifying selection. Many phylogenetic analysis programs can be set to assume that silent sites are neutral and freer to change than nonsilent sites in protein coding regions.

For most purposes, there is no need to separate mutation rate from evolution rate, and in fact it is counterproductive to do so when the observed virus sequences are the product of both mutation and selection over time. Likewise, selection processes vary slightly from one patient to another or even within a single patient over time, but the average behavior over many sites in the genome and over larger timescales can be remarkably consistent. Conversely the study of very small regions of the genome and/or using very small or not well-chosen data sets (not randomly selected from the population under study) can result in very poor estimations of rates and/or patterns of evolution [16–19].

5.2.2 Hypermutation

A large source of mutation in retroviruses is the activity of APOBEC enzymes. The APOBEC enzymes recognize DNA/RNA heteroduplex molecules in the cytoplasm and deaminate cytidines in the RNA strand. The result is that many stop codons are introduced into the genes, and functional proteins can no longer be produced by the mutated viral genome [8, 20]. Because the retrovirus brings some reverse transcriptase and integrase protein into the cell, the mutated nonfunctional provirus can be integrated into the host genome, and it is not uncommon to find hypermutated viral sequences when proviral DNA is amplified and sequenced. Although dead or nonfunctional genomes cannot continue to replicate and evolve, each one is just one replication round away from its parental virus and so can contribute valid information about the host virus population.

5.2.3 Insertions/Deletions

Duplications of motifs such as NK-kappa-B binding site, PSAPP/PTAPP motif in gag. Duplications of nearby DNA RNY in env. Length of Env loops. Vpr/Vpx duplication.

In addition to point mutations, insertions and deletions contribute to the evolution of DNA. Insertions and deletions are usually not counted in phylogenetic analyses of evolution because there are many different types of insertions and deletion events, and it is not possible to directly compare them to rates of single base mutations. One common type of insertion and deletion event is known as variable numbers of tandem repeats, where a simple repetitive element increases or decreases in copy number. Gene duplication events are also relatively common in most organisms, but in retroviruses there are constraints on genome size, which would limit the viability of most such events. One gene duplication event in the primate lentiviruses is hypothesized to have created the Vpx plus Vpr gene pair in some lineages, while other lineages have Vpr plus Vpu genes [21, 22].

5.3 Selection

There are many different types of selection forces acting on viral genome sequences, usually with overlapping and either conflicting or supporting roles. The most obvious selection forces on protein coding regions of the virus are conserving the amino acid codons needed for a given protein function and changing the surface of the envelope protein to evade the host antibody immune responses [23].

5.3.1 Positive by CTL

Cytotoxic T lymphocytes possess HLA (system?) that cleaves viral proteins into primarily 9-mer peptides which are then presented on the cell surface of infected cells by xxx and recognized by killer T cells. Viral protein cleavage is not random, but dozens of 9-mers per viral protein can be presented by the average infected cell. Not all epitopes are equally effective at eliminating virus-infected cells, so there are dominant epitopes responsible for the majority of viral reduction plus many more weak epitopes that are much less effective. Different human HLA genotypes tend to target different viral epitopes such that the dominant epitope in one individual is not dominant in most others.

5.3.2 Positive by Antibody

Infected individuals produce antibody responses to nearly all viral proteins, with a general trend in the order of appearance of strong antibody responses appearing in the first weeks of infection. Most antibodies do not neutralize the virus and have little impact on viral load and thus have little impact on viral evolution. Neutralizing antibodies, which bind to the viral envelope glycoprotein, prevent the virus from binding to cell surface receptors on uninfected cells (CD4 and either CCR5 or CXCR4 coreceptor). Neutralizing antibodies are often highly specific for only a single lineage of virus. In relatively rare cases, a single antibody can bind to and neutralize a wide variety of lineages, and these are called broadly neutralizing antibodies (BNabs).

Elimination of virions that are bound by neutralizing antibodies drives the evolution of escape mutants. The most common mechanism of escape is addition and/or subtraction of glycosylation sites, which are highly prevalent and variable on the surface of the envelope glycoprotein. The glycosylation of envelope is referred to as “the glycan shield” [24].

5.3.3 Positive by Drug

In contrast to host immune responses, which are variable between infected individuals and even variable over time within each individual, an antiretroviral drug is always the same and thus exerts the same selection pressure in all individuals who take the drug. Detection of immune system escape mutations is complex and requires many viral sequences and immune reaction tests for each individual. Detection of drug resistance mutations can be done with as little as one viral sequence per individual and rather simple tests for viral replication rate in the presence or absence of each drug.

For one example, the change of wild-type methionine to valine at amino acid 184 (M184V) in the reverse transcriptase protein results in nearly complete resistance of the virus to azidothymidine (AZT) and some related nucleoside analog reverse transcriptase inhibitors. The M184V mutation also partially cripples the reverse transcriptase enzyme such that although the virus is resistant to any level of AZT, it replicates slower than wild-type virus. Maintaining AZT therapy in the presence of M184V mutant virus can be of benefit to infected people because the drug still suppresses all of the wild-type viruses; thus viral load is reduced. Also, each drug resistance mutation first occurs in a single virion, so that populations of drug-resistant virions, which descend from that mutant, have low diversity and greater chance for control by the individual's immune responses.

A major concern with drug resistance mutations is the transmission of drug-resistant virus from one infected person to others such that whole populations of people are infected with drug-resistant viruses, and the effectiveness of treatment and especially the ability to prevent mother-to-child transmissions with simple drug regimens proven to be safe for fetuses and infants is reduced. Surveillance for transmitted drug resistance (TDR) is thus an important public health issue as well as being beneficial to individuals who will get a report of which drug combinations are most suitable for treating the particular virus they are infected with.

5.3.4 Negative by Replication Rate

Although Darwin's theory of evolution is most often said to be "survival of the fittest," it is actually more accurately described as "elimination of the weakest" for most living things. It is not typically the single most fit individual that survives each generation and huge advancements in fitness are rarely realized in a single round of replication. Highly unfit or lethal mutations on the other hand occur very frequently and are eliminated from the gene pool nearly instantly. The dominant population of any group of organisms is generally referred to as "wild type," and the vast majority of mutants are observed to be less fit than wild type. When nonlethal but slightly less fit mutants arise, they can be eliminated from the population over time either by being outcompeted by wild-type individuals or by reverting to wild type. They can persist in the population by acquiring compensatory mutations, which allow them to regain fitness, or by spreading into habitats where they are no longer in competition with wild-type individuals.

Nearly all organisms have many stages in their life cycles where changes in fitness can influence the longer-term evolution of the population or species as a whole. Although a fast replication rate will benefit a lineage of virus within an individual in the short term, killing the host or making the host too sick to interact with potential new hosts to pass on the infection leads to an evolutionary dead end. Replication rate within an individual can thus be in conflict with overall epidemic growth rate. The lentiviruses require close physical contact for host to host spread, and tend to have long asymptomatic periods during which an infected individual can pass the virus on to others. The term "lenti" in *Lentivirus* is Latin for slow, and these viruses were named long before the primate immunodeficiency viruses were discovered [25, 26].

One important factor in controlling HIV viral replication rate seems to be the very poor codon use in the viral genome. The genetic code is redundant, with 61 codons (plus 3 stop codons) for 20 amino acids, and different organisms have different levels of transfer RNAs matching the various synonymous codons. The lentiviral genomes have a very high frequency of codons that are read by low-level tRNAs in mammals. When synthetic DNA encoding HIV proteins but with high-level "fast" mammalian codon use is transfected into mammalian cells, protein production can be nearly 200-fold higher than transfection with DNA using the lentiviral codons.

5.3.5 Negative by Infectivity

To be passed from cell to cell within an individual and to be transmitted between host individuals, a virus must bind to cell surface receptors and maintain an ability to penetrate the host cell membrane. Although the virus evolves rapidly, host cell receptors are nearly invariant within any mammalian species. For the lentiviruses, the CD4 T-cell receptor, which is also expressed on macrophages, dendritic cells, and several other cell types, is the primary host cell receptor. The CD4 binding site on the envelope protein is thus quite highly conserved (Fig. 5.3).

5.3.6 Negative by Codon Use Bias

Codon use bias is a conundrum in HIV-1 M group and most other lentiviruses. The synonymous codons selected for use by the virus are very slow for translating the viral messenger RNA to protein in human cells. In the study of protein production levels, changing the codons from native HIV-1 sequences to codons optimized for mammalian expression, results in as much as a 200-fold increase in the level of protein produced [28, 29]. Lentiviral genomes are nearly uniformly A-rich and C-poor with the A:C ratio in any given region of the genome close to 2:1.

5.3.7 3.8 Negative by RNA Secondary Structure

In addition to encoding all of the proteins needed for viral replication and defeat of host immune defense mechanisms, the virus must also provide mechanisms of packaging viral genomes into virions and regulating gene expression levels and

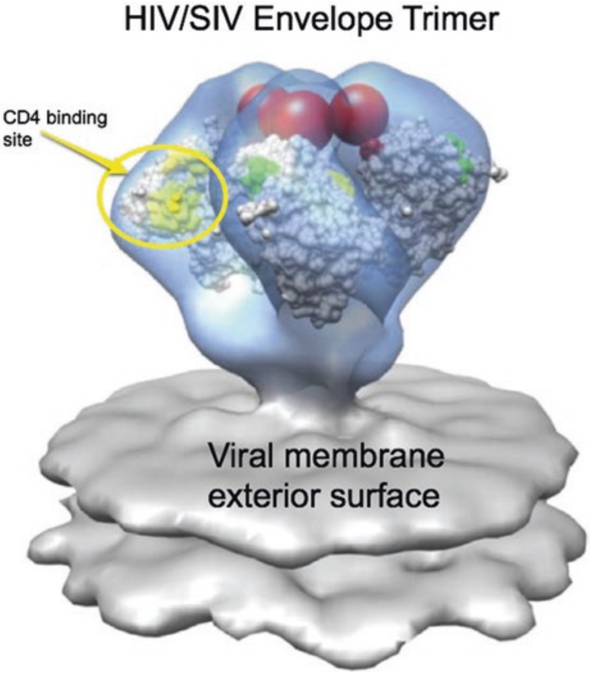

Fig. 5.3 Three-dimensional rendering of HIV/SIV envelope protein structure. Red balls illustrate the V1–V2 hypervariable loop region, green highlights the V3 loop region, and yellow highlights the relatively conserved CD4 cell surface receptor binding site (The figure is adapted from Ref. [27])

timing, and several other functions, which are known to involve self-complementarity and RNA secondary and tertiary structures [30, 31] (Fig. 5.4).

5.4 Recombination

Retroviruses package two copies of the viral RNA in each virion. During reverse transcription, the viral reverse transcriptase enzyme can switch between the two template RNA molecules and thus produce a complementary DNA that is partially derived from each of the two template genomes which were packaged in the virion. Although many or most cells are infected with only one virion such that the progeny viral genomes being packaged are identical to each other, multiple infections of a single cell are common and result in progeny virions with two different genomes packaged. Dual infection of a single individual with two different subtypes of HIV-1 can thus result in intersubtype recombination (Fig. 5.5).

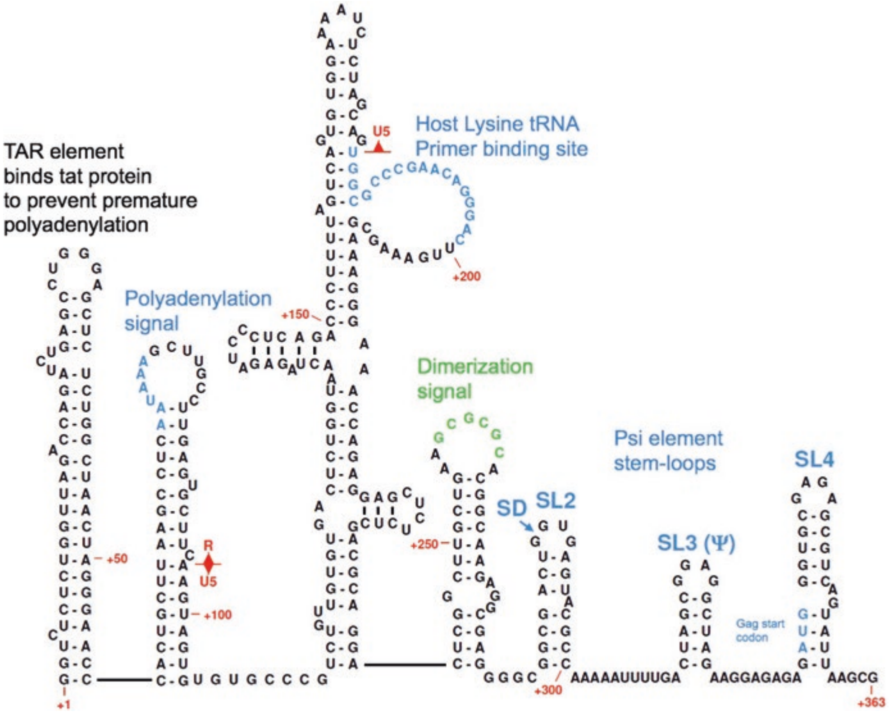

Fig. 5.4 Viral RNA secondary structure stem-loops in the 5′ long terminal repeat (LTR)

5.4.1 Recombination Within an Individual Virus Lineage

The vast majority of HIV-1 infections are derived from a single virus. This is known as the transmission bottleneck. It remains unproven whether the transmission bottleneck is the result of a single virion crossing the mucosal surface at the time of transmission or alternatively whether several viruses typically are transmitted but one of them rapidly outgrows the others in the initial days of infection. Multiple infections with more than one strain of virus are somewhat rare, but in populations of people at high risk of infection, such as commercial sex workers and IV drug users, it is not unusual to find people infected with more than one strain. If a person is infected with a second virus before seroconversion to the first virus has taken place, it is termed a dual infection. Infection with a second strain after seroconversion is known as superinfection.

Within an individual that was infected with only one virus, recombination happens, but the recombinants are derived from two nearly identical template genomes such that detecting the recombination events is often impossible. However, after many years of infection, the viruses within an individual have acquired some diversity, and it then becomes possible to detect the recombinant genomes [33, 34]. Although intrapatient recombination in a singly infected individual does not have

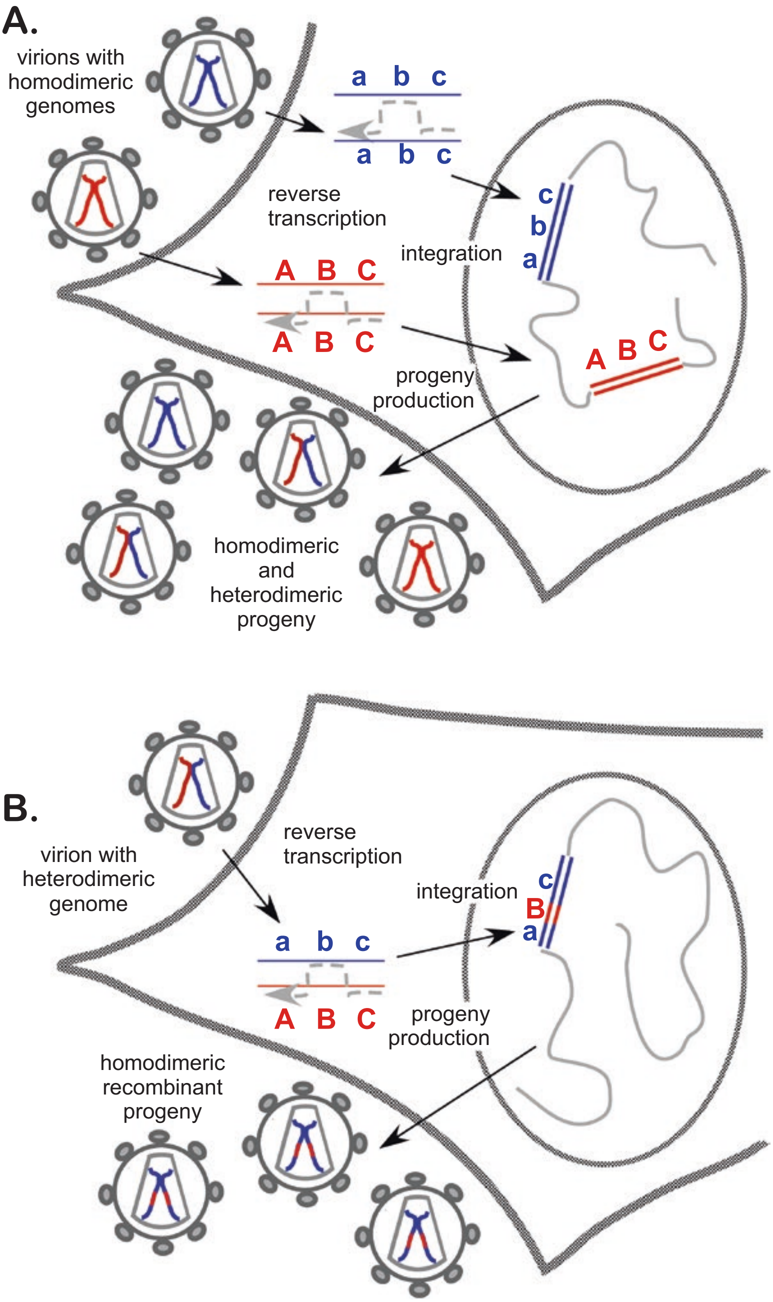

Fig. 5.5 Generation of recombinant genomes. (**a**) Coinfection of a single cell with two genetically distinct viruses can, if both proviruses are activated to production at the same time, lead to production of some virions containing one genome from each, depicted as red and blue here. (**b**) The recombinant genomes are generated during reverse transcription in a cell infected with a virion containing two different genome templates. The recombinant genomes are generated by template switching during the reverse transcription and do not require any nucleic acid strand breakage or repair (Artwork kindly provided by Alice Telesnitsky [32])

any significant impact on virus evolution for vaccine design or phylogenetic analyses of the overall epidemic, it does have large impacts on the ability of the virus to escape numerous selection pressures. For example, recombination between a virion with a drug resistance mutation in the pol gene and another virion with a mutation allowing it to escape immune selection in the env gene can produce progeny viruses that are resistant to both of these selection pressures.

5.4.2 Intrasubtype Recombination

Virus recombination is typically only noted when the two participating viruses are of different genetic subtypes. Intrasubtype recombination is of course far more frequent, but it is usually difficult to detect and so commonplace that it is usually not of interest to report on it. Dual infections and superinfections with the same subtype of virus but from a different donor are more common than infection with multiple subtypes, because in most communities in the world, only one subtype of the virus is present [34, 35]. Intrasubtype recombination has little or no impact on epidemiology, vaccine design, and many other aspects of HIV biology, but it can be a driving influence for recombining different regions of the viral genome carrying different selective advantages within an individual such as antibody escape mutations in the envelope gene and CTL escape mutations in the *gag* gene [33].

5.4.3 Intersubtype Recombination

Multiple infections with different subtypes soon result in intersubtype recombinant viruses. Although in theory recombination could happen often enough within such an individual to soon generate scrambled genomes with only very short regions derived from each of the parental viruses, in practice the observed intersubtype recombinants are not scrambled and usually have fewer than ten recombination breakpoints [36, 37]. It is possible that genetic distance between the subtypes has resulted in genomes that are not fully compatible with each other, and thus not all recombinants are equally fit [38]. This would prevent many intersubtype recombinant virions from thriving in an individual and prevent the scrambling of genomes over time. It is also possible that recombination just does not happen as often as theory would predict [39].

5.5 Phylogenetic Reconstruction

Phylogenetics is the study of the evolutionary relationships between organisms. Although the evolutionary histories of many plants and animals can be accurately inferred by analyzing phenotypic traits such as leaf structures or wing and beak shapes, the use of DNA sequences is much more common today and more accurate [3, 40]. Reconstructing the evolutionary tree or ancestral history of viruses has many uses, and there are many data sets available in which the biology is very well known such that the theory and practice can be tested [16, 41, 42].

5.5.1 Data Set Choices

Several factors influence the types of information that can be gained by phylogenetic analysis of virus genetic sequences. When analyzing the evolution of the viruses within a single patient, the most common type of sample is a blood sample, but the viruses present in the blood at any one moment in time may not be an accurate representation of the viruses present in lymph nodes, central nervous system, or other sites in the body. Several studies have attempted to assess the degree to which different sites in the body tend to host their own sublineages of virus, which is known as compartmentalization [43–46]. Within a population of people, it is similarly difficult to obtain a truly random sample of infected people. In addition to the choice of biological samples to use, there are choices to be made about which region(s) of the genome to sequence and how many sequences will be needed to obtain the desired statistical power.

In all cases, some compromise must be reached between the theoretically ideal data set and the data that can actually be obtained given biological, ethical, funding, and other constraints. In many cases it is possible to enhance the statistical power of a given study by supplementing the new data from a given study with data obtained by other studies and available in the genetic databases such as GenBank and the HIV Databases at the Los Alamos National Laboratory [47, 48].

5.5.2 Method and Model Choices

The simplest model of evolution assumes that all DNA base changes are equally likely and that there is no selection pressure or other influence on the rate of evolution of different sites in each gene. The Kimura two-parameter model adds just one factor, stating that transitions and transversions have different rates [49]. More complex models can evaluate a different rate for each base change and also allow each site in a gene (column in a multiple sequence alignment) to have a different rate of evolution. There are programs such as ModelTest and PartitionFinder to assist in the rational decision of which model of evolution to use on a given data set [50, 51]. As the size of the data set grows, the computational resources needed to perform the most complex analyses increase factorially with the number of sequences and linearly with the length of the sequences. Although it is impossible to compute the absolutely correct or best tree from large data sets, there are heuristics employed to greatly reduce the number of computations needed to arrive at a very reasonable result. Poor choice of samples, sequencing errors, and other problems with the input data sets are far more often the cause of serious problems than suboptimal choices of computational methods.

5.5.3 Recombination Detection

Phylogenetic reconstruction of evolutionary history in general assumes that the sequences being analyzed are not recombinant and that each sequence has one history. Although in practice HIV does undergo recombination, the scope of the recombination is limited. The viruses in one infected individual are not recombining with viruses in any other infected individual. Recombination can confound or invalidate phylogenetic analyses, but it is not always a problem, and there are many methods available for detecting recombination [34, 40, 52].

5.5.4 Alignment

Almost any analysis of multiple sequences from the same organism requires that all of the sequences be aligned to one another in a multiple sequence alignment. Pairwise alignment of any two sequences, or of each sequence in a set to one reference sequence, is usually simple using the Smith-Waterman algorithm [53]. Aligning many sequences to each other, when each of the sequences has insertions and deletions relative to other sequences, becomes a much more difficult problem, but many programs have been written to automate the task [54]. The HIV Databases at Los Alamos National Laboratory have developed tools specifically designed for aligning HIV sequences, which take into account the multiple overlapping reading frames used by the virus [47, 55].

Obtaining a very good multiple sequence alignment often involves iterations of producing a multiple sequence alignment, analyzing the alignment by methods such as building a phylogenetic tree and using Simplot to look for uniform diversity between sequences, and then adjusting the alignment if the analyses indicated any region of the alignment or sequences in the alignment were aberrant.

5.6 Rates and Dates

5.6.1 Molecular Clock Tests

The gene sequences of viruses and other organisms change over time due to mutations and selection pressures. The consistency of the rate of change over time is known as the molecular clock hypothesis. Individual mutations happen in a stochastic manner with little predictability, but the sum of changes over large regions of the genome and longer timescales tends to be more uniform. Many factors including population sizes, selection pressures, generation times, and fidelity of replication influence the clock rate; thus most data sets do not show evidence of a strict molecular clock. However, given large data sets, the average behavior is clocklike enough to allow many inferences about the past history of populations such as effective population sizes and dates of divergence from a common ancestor [3, 40, 56].

5.6.2 Examples

Very early in the study of the HIV/AIDS pandemic, it was noted that there was a great diversity between HIV isolates in comparison to the diversity observed in most other viruses [57, 58]. The first estimates to estimate the rates of evolution and to use the rate to date the origin of the pandemic were hampered by small sample sizes and by missing information about the natural history of the primate lentiviruses. Very good estimates have now been made by many groups, using independent methods and sample collections, with very high levels of agreement between them [3, 9, 59–62].

Within the HIV-1 M group, many studies have analyzed the growth of subsets of the AIDS pandemic using sequences from viruses collected over time in various parts of the world. Several studies, for example, have attempted to pinpoint the time and location of the beginning of the HIV-1 subtype B epidemic in the USA [63–65]. Molecular clock analyses of HIV-1 subtype B in the USA and Europe agree that the date of the common origin of subtype B was between 1960 and 1970, and many papers speculate that HIV was incubating in the USA for nearly 20 years before being detected in 1981. However, it is also possible that the subtype B viruses were evolving in other parts of the world, and then multiple introductions of HIV-1 subtype B entered the USA in the late 1970s and early 1980s [66]. Analyses of virus sequences can provide accurate information on the date of the common ancestor of the viruses, but this does not provide information on the geographic location of the ancestor.

5.6.3 Saturation Effects

The DNA bases thymine and cytosine are pyrimidines with one ring, while guanine and adenine are purines with two rings. Because of the size difference and other factors, substitutions of one base for another do not all happen equally. Transitions far outnumber transversions, with the lentiviruses being especially prone to G to A transitions. Rather than simply counting all point differences between sequences equally, models of evolution can calculate different rates for different types of mutations and attempt to correct for multiple mutations at a given site. One method of

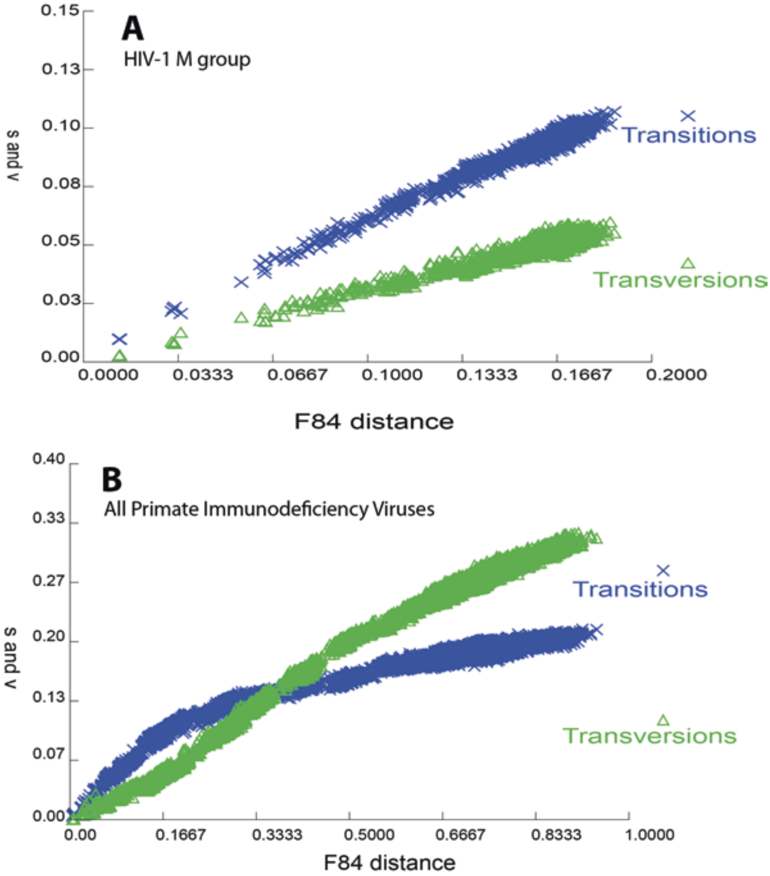

Fig. 5.6 Saturation of sequences, multiple hits per site. Transitions (G <-> A and C <-> T) outnumber transversions (A <-> C, A <-> T, G <-> C, and G <-> T) when distances are relatively small, but as mutations accumulate such that more variable sites have mutated more than once, saturation is reached, and computation of the phylogenetic or molecular clock time distance from the observed distance becomes difficult to impossible. (**a**) complete genomes of HIV-1 M group viruses were analyzed. (**b**) the complete genomes of all primate immunodeficiency viruses were analyzed

testing for saturation of mutable sites is to calculate the transition to transversion ratios of all pairs of sequences in the data set. The DAMBE phylogenetic analysis package [40] provides a tool for producing a graphical plot of transitions and transversions versus pairwise distances. Figure 5.6 shows the results of analyzing the data sets used to make Figs. 5.1 and 5.2.

Most phylogenetic tree building programs also calculate a matrix of substitutions observed in the data and have an option for outputting the full matrix. However, this matrix is an average over all comparisons and will not show whether or not saturation is observed in the data.

Conflict of interest The authors report no conflicts of interest.

References

1. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF et al (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature [Internet] 397(6718):436–441. [cited 2016 Jun 7]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9989410
2. D'arc M, Ayouba A, Esteban A, Learn GH, Boué V, Liegeois F et al (2015) Origin of the HIV-1 group O epidemic in western lowland gorillas. Proc Natl Acad Sci U S A [Internet] 112(11):E1343–E1352. [cited 2016 Mar 16]. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4371950&tool=pmcentrez&rendertype=abstract
3. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ et al (2014) HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science [Internet] 346(6205):56–61. [cited 2016 Jun 24]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25278604
4. Katzourakis A, Gifford RJ, Tristem M, Gilbert MTP, Pybus OG (2009) Macroevolution of complex retroviruses. Science [Internet] 325(5947):1512. [cited 2012 Dec 19]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19762636
5. Hron T, Farkašová H, Padhi A, Pačes J, Elleder D (2016) Life History of the oldest lentivirus: characterization of ELVgv integrations in the dermopteran genome. Mol Biol Evol [Internet] 33(10):2659–2669. [cited 2016 Sep 20]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27507840
6. Keckesova Z, LMJ Y, Towers GJ, Gifford RJ, Katzourakis A (2009) Identification of a RELIK orthologue in the European hare (Lepus europaeus) reveals a minimum age of 12 million years for the lagomorph lentiviruses. Virology 384:7–11
7. Puvvada MP, Patel SS (2013) Role of trim5α in the suppression of cross-species transmission and its defence against human immunodeficiency virus. Curr HIV Res [Internet] 11(8):601–609. [cited 2016 Jun 7]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24606328
8. Etienne L, Bibollet-Ruche F, Sudmant PH, Wu LI, Hahn BH, Emerman M (2015) The role of the antiviral APOBEC3 gene family in protecting chimpanzees against lentiviruses from monkeys. PLoS Pathog [Internet] 11(9):e1005149. [cited 2016 Jun 24]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26394054
9. Hogan CA, Iles J, Frost EH, Giroux G, Cassar O, Gessain A et al (2016) Epidemic history and iatrogenic transmission of blood-borne viruses in mid-20th century Kinshasa. J Infect Dis [Internet] 214(3):353–360. [cited 2016 Jun 24]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26768251
10. Pépin J, Labbé A-C, Mamadou-Yaya F, Mbélesso P, Mbadingaï S, Deslandes S et al (2010) Iatrogenic transmission of human T cell lymphotropic virus type 1 and hepatitis C virus

through parenteral treatment and chemoprophylaxis of sleeping sickness in colonial Equatorial Africa. Clin Infect Dis [Internet] 51(7):777–784. [cited 2016 Oct 18]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20735238

11. Deuchert E, Brody S (2006) The role of health care in the spread of HIV/AIDS in Africa: evidence from Kenya. Int J STD AIDS [Internet] 17(11):749–752. [cited 2016 Nov 8]. Available from: http://std.sagepub.com/lookup/doi/10.1258/095646206778691167
12. Ganusov VV, Goonetilleke N, Liu MKP, Ferrari G, Shaw GM, McMichael AJ et al (2011) Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. J Virol [Internet] 85(20):10518–10528. [cited 2016 Oct 20]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21835793
13. Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T et al (2010) Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PLoS One [Internet] 5(8):e12303. [cited 2016 Oct 20]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20808830
14. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics [Internet] 26(19):2455–2457. [cited 2016 Nov 8]. Available from: http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btq429
15. Leenen FAD, Muller CP, Turner JD, Garcia-Carpizo V, Ruiz-Llorente L, Fraga M et al (2016) DNA methylation: conducting the orchestra from exposure to phenotype? Clin Epigenetics [Internet] 8(1):92. [cited 2016 Sep 22]. Available from: http://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-016-0256-8
16. Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J (1996) Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci U S A [Internet] 93(20):10864–10869. [cited 2016 Nov 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8855273
17. Mooers, Holmes (2000) The evolution of base composition and phylogenetic inference. Trends Ecol Evol [Internet] 15(9):365–369. [cited 2016 Nov 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10931668
18. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B et al (2013) Great ape genetic diversity and population history. Nature [Internet] 499(7459):471–475. [cited 2014 Mar 20]. Available from: http://dx.doi.org/10.1038/nature12228
19. Krishnan NM (2004) Ancestral sequence reconstruction in primate mitochondrial DNA: compositional bias and effect on functional inference. Mol Biol Evol [Internet] 21(10):1871–1883. [cited 2016 Nov 8]. Available from: http://mbe.oupjournals.org/cgi/doi/10.1093/molbev/msh198
20. Malim MH (2009) APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos Trans R Soc Lond B Biol Sci [Internet] 364(1517):675–688
21. Sharp PM, Bailes E, Stevenson M, Emerman M, Hahn BH (1996) Gene acquisition in HIV and SIV. Nature [Internet] 383(6601):586–587. [cited 2016 Oct 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8857532
22. Tristem M, Marshall C, Karpas A, Hill F (1992) Evolution of the primate lentiviruses: evidence from vpx and vpr. EMBO J [Internet] 11(9):3405–3412. [cited 2016 Oct 3]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1324171
23. Losos JB, Arnold SJ, Bejerano G, Iii EDB, Hibbett D, Moritz C et al (2013) Evolutionary biology for the 21st century. PLoS Biol [Internet] 11(1):e1001466. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3539946&tool=pmcentrez&rendertype=abstract
24. Dacheux L, Moreau A, Ataman-Onal Y, Biron F, Verrier B, Barin F (2004) Evolutionary dynamics of the Glycan shield of the human immunodeficiency virus envelope during natural infection and implications for exposure of the 2G12 epitope. J Virol [Internet] 78(22):12625–12637. [cited 2016 Nov 8]. Available from: http://jvi.asm.org/cgi/doi/10.1128/JVI.78.22.12625-12637.2004

25. Thormar H (2013) The origin of lentivirus research: maedi-visna virus. Curr HIV Res [Internet] 11(1):2–9. [cited 2016 Nov 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23278353
26. Stowring L, Haase AT, Charman HP (1979) Serological definition of the lentivirus group of retroviruses. J Virol [Internet] 29(2):523–528. [cited 2016 Nov 8]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/85722
27. White TA, Bartesaghi A, Borgnia MJ, de la Cruz MJV, Nandwani R, Hoxie JA et al (2011) Three-dimensional structures of soluble CD4-bound states of trimeric simian immunodeficiency virus envelope glycoproteins determined by using cryo-electron tomography. J Virol [Internet] 85(23):12114–12123. [cited 2016 Nov 9]. Available from: http://jvi.asm.org/cgi/doi/10.1128/JVI.05297-11
28. Honarmand Ebrahimi K, West GM, Flefil R (2014) Mass spectrometry approach and ELISA reveal the effect of codon optimization on N-linked glycosylation of HIV-1 gp120. J Proteome Res [Internet] 13(12):5801–5811. [cited 2016 Nov 9]. Available from: http://pubs.acs.org/doi/10.1021/pr500740n
29. Gao F, Li Y, Decker JM, Peyerl FW, Bibollet-Ruche F, Rodenburg CM et al (2003) Codon usage optimization of HIV type 1 subtype C *gag*, *pol*, *env*, and *nef* genes: *in vitro* expression and immune responses in DNA-vaccinated mice. AIDS Res Hum Retroviruses [Internet] 19(9):817–823. [cited 2016 Nov 10]. Available from: http://www.liebertonline.com/doi/abs/10.1089/088922203769232610
30. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM et al (2013) Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. PLoS Pathog [Internet] 9(4):e1003294. [cited 2016 Nov 10]. Available from: http://dx.plos.org/10.1371/journal.ppat.1003294
31. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R et al (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. Nature [Internet] 460(7256):711–716. [cited 2016 Nov 10]. Available from: http://www.nature.com/doifinder/10.1038/nature08237
32. Onafuwa-Nuga A, Telesnitsky A (2009) The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. Microbiol Mol Biol Rev [Internet] 73(3):451–480. [cited 2016 Oct 18]. Table of Contents. Available from: http://mmbr.asm.org/cgi/content/long/73/3/451
33. Philpott S, Burger H, Tsoukas C, Foley B, Anastos K, Kitchen C et al (2005) Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intrapatient recombination. J Virol [Internet] 79(1):353–363. [cited 2016 Oct 18]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15596829
34. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG et al (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci [Internet] 105(21):7552–7557. [cited 2016 Nov 8]. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0802203105
35. Kiwelu IE, Novitsky V, Margolin L, Baca J, Manongi R, Sam N et al (2013) Frequent intra-subtype recombination among HIV-1 circulating in Tanzania. PLoS One [Internet] 8(8):e71131. [cited 2016 Nov 8]. Available from: http://dx.plos.org/10.1371/journal.pone.0071131
36. Rousseau CM, Learn GH, Bhattacharya T, Nickle DC, Heckerman D, Chetty S et al (2007) Extensive intrasubtype recombination in South African human immunodeficiency virus type 1 subtype C infections. J Virol [Internet] 81(9):4492–4500. [cited 2016 Nov 8]. Available from: http://jvi.asm.org/cgi/doi/10.1128/JVI.02050-06
37. Cromer D, Grimm AJ, Schlub TE, Mak J, Davenport MP (2016) Estimating the in-vivo HIV template switching and recombination rate. AIDS [Internet] 30(2):185–192. [cited 2016 Nov 8]. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00002030-201601140-00004
38. Golden M, Muhire BM, Semegni Y, Martin DP (2014) Patterns of recombination in HIV-1M are influenced by selection disfavouring the survival of recombinants with disrupted genomic RNA and protein structures. PLoS One [Internet] 9(6):e100400. [cited 2016 Nov 15]. Available from: http://dx.plos.org/10.1371/journal.pone.0100400

39. Chin MPS, Rhodes TD, Chen J, Fu W, Hu W-S (2005) Identification of a major restriction in HIV-1 intersubtype recombination. Proc Natl Acad Sci [Internet] 102(25):9002–9007. [cited 2016 Nov 8]. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0502522102
40. Salemi M, Vandamme A-M, Lemey P (2009) The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, Cambridge, p 723
41. Lemey P, Derdelinckx I, Rambaut A, Van Laethem K, Dumont S, Vermeulen S et al (2005) Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain. J Virol [Internet] 79(18):11981–11989. [cited 2016 Nov 10]. Available from: http://jvi.asm.org/cgi/doi/10.1128/JVI.79.18.11981-11989.2005
42. Paraskevis D, Magiorkinis E, Magiorkinis G, Kiosses VG, Lemey P, Vandamme A-M et al (2004) Phylogenetic reconstruction of a known HIV-1 CRF04_cpx transmission network using maximum likelihood and Bayesian methods. J Mol Evol [Internet] 59(5):709–717. [cited 2016 Nov 10]. Available from: http://link.springer.com/10.1007/s00239-004-2651-6
43. Chaillon A, Gianella S, Wertheim JO, Richman DD, Mehta SR, Smith DM (2014) HIV migration between blood and cerebrospinal fluid or semen over time. J Infect Dis [Internet] 209(10):1642–1652. [cited 2016 Nov 10]. Available from: http://jid.oxfordjournals.org/lookup/doi/10.1093/infdis/jit678
44. Smith DM, Zárate S, Shao H, Pillai SK, Letendre SL, Wong JK et al (2009) Pleocytosis is associated with disruption of HIV compartmentalization between blood and cerebral spinal fluid viral populations. Virology [Internet] 385(1):204–208. [cited 2016 Nov 10]. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0042682208007514
45. Kemal KS, Ramirez CM, Burger H, Foley B, Mayers D, Klimkait T et al (2012) Recombination between variants from genital tract and plasma: evolution of multidrug-resistant HIV type 1. AIDS Res Hum Retroviruses [Internet] 28(12):1766–1774. [cited 2016 Nov 10]. Available from: http://online.liebertpub.com/doi/abs/10.1089/aid.2011.0383
46. Kemal KS, Foley B, Burger H, Anastos K, Minkoff H, Kitchen C et al (2003) HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation. Proc Natl Acad Sci [Internet] 100(22):12972–12977. [cited 2016 Nov 10]. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.2134064100
47. Kuiken C, Korber B, Shafer RW (2016) HIV sequence databases. AIDS Rev [Internet] 5(1):52–61. [cited 2016 Nov 10]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12875108
48. NCBI Resource Coordinators NR (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res [Internet] 41(Database issue):D8–20. [Internet]. Oxford University Press; [cited 2016 Nov 10]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23193264
49. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol [Internet] 16(2):111–120. [cited 2016 Nov 10]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7463489
50. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods [Internet] 9(8):772–772. [cited 2016 Nov 10]. Available from: http://www.nature.com/doifinder/10.1038/nmeth.2109
51. Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol [Internet] 29(6):1695–1701. [cited 2016 Nov 10]. Available from: http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/mss020
52. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol [Internet] 1(1):vev003. [cited 2016 Nov 15]. Available from: http://ve.oxfordjournals.org/cgi/doi/10.1093/ve/vev003
53. Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol [Internet] 147(1):195–197. [cited 2016 Nov 15]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7265238

54. Neuwald AF, Altschul SF (2016) Bayesian top-down protein sequence alignment with inferred position-specific gap penalties. PLoS Comput Biol [Internet] 12(5):e1004936. [cited 2016 Nov 15]. Available from: http://dx.plos.org/10.1371/journal.pcbi.1004936
55. Gaschen B, Kuiken C, Korber B, Foley B (2001) Retrieval and on-the-fly alignment of sequence fragments from the HIV database. Bioinformatics [Internet] 17(5):415–418. [cited 2016 Nov 15]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11331235
56. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme A-M (2003) Tracing the origin and history of the HIV-2 epidemic. Proc Natl Acad Sci U S A [Internet] 100(11):6588–6592. [cited 2016 Jun 7]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12743376
57. Smith TF, Srinivasan A, Schochetman G, Marcus M, Myers G (1988) The phylogenetic history of immunodeficiency viruses. Nature [Internet] 333(6173):573–575. [cited 2013 Apr 1]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3131682
58. Myers G, Korber B, Berzofksy JA, Smith RF, Pavlakis GN (1992) Human retroviruses and AIDS 1992: a compilation and analysis of nucleic acid and amino acid sequences [Internet]. Los Alamos National Laboratory, Los Alamos. Available from: http://www.hiv.lanl.gov/
59. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ, Emerman M (2009) Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. PLoS Pathog [Internet] 5(9):e1000590. [cited 2016 Nov 15]. Available from: http://dx.plos.org/10.1371/journal.ppat.1000590
60. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A et al (2000) Timing the ancestor of the HIV-1 pandemic strains. Science [Internet] 288(5472):1789–1796. [cited 2016 Nov 15]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10846155
61. Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, Brown TM et al (2003) U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. J Virol [Internet] 77(11):6359–6366. [cited 2016 Nov 15]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12743293
62. Pépin J, Labbé A-C (2008) Noble goals, unforeseen consequences: control of tropical diseases in colonial Central Africa and the iatrogenic transmission of blood-borne viruses. Trop Med Int Health [Internet] 13(6):744–753. [cited 2016 Oct 18]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18397182
63. Worobey M, Watts TD, McKay RA, Suchard MA, Granade T, Teuwen DE et al (2016) 1970s and "Patient 0" HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature [Internet] 539(7627):98–101. [cited 2016 Dec 5]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27783600
64. Pagán I, Holguín Á (2013) Reconstructing the timing and dispersion routes of HIV-1 subtype B epidemics in the Caribbean and Central America: a phylogenetic story. PLoS One [Internet] 8(7):e69218. [cited 2016 Dec 5]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23874917
65. Cabello M, Junqueira DM, Bello G (2015) Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. AIDS [Internet] 29(4):483–492. [cited 2016 Dec 5]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25630042
66. Foley B, Pan H, Buchbinder S, Delwart EL (2000) Apparent founder effect during the early years of the San Francisco HIV type 1 epidemic (1978–1979). AIDS Res Hum Retroviruses [Internet] 16(15):1463–1469. [cited 2016 Dec 5]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11054259