The high genetic diversity of HIV is one of its most significant features, as it has consequences in global distribution, vaccine design, therapy success, disease progression, transmissibility and viral load testing. Studying HIV diversity helps to understand its origins, migration patterns, current distribution and transmission events. New advances in sequencing technologies based on the parallel acquisition of data are now used to characterize within-host and population processes in depth. Additionally, we have seen similar advances in statistical methods designed to model the past history of lineages (the phylodynamic framework) to ultimately gain better insights into the evolutionary history of HIV. We can, for example, estimate population size changes, lineage dispersion over geographic areas and epidemiological parameters solely from sequence data. In this article, we review some of the evolutionary approaches used to study transmission patterns and processes in HIV and the insights gained from such studies.

Recent advances in DNA sequencing as well as new approaches to analyzing these data allow researchers to study the impact of epidemiological factors on the evolutionary dynamics of HIV at global (worldwide), regional (single epidemics), local (transmission chains) and individual (intrahost) scales by examining genetic variation across its genome and over geographic space and time. The purpose of this review is to present an overview of HIV genetic diversity (GD) and its estimation as well as the kind of insights these new methodologies can provide and how they can improve disease control and treatment.

**GD of HIV**
GD is probably one of the most important concepts in biology. In its most simple definition, GD refers to any and every kind of genetic variation at the individual, population, interpopulation or species level. GD has a large impact on conservation biology [1] and the study of human origins [2], as well as molecular epidemiology [3], domestication [4], fitness [5] and disease [6]. In HIV-1, higher levels of GD have been associated with clinical outcomes such as immune escape of selected variants [7], emergence of drug resistance mutations and the consequent therapy failure [8], and even with disease progression [9,10]. GD has also been used to study the geographic and temporal spread of HIV-1, shedding light on global and regional population dynamics.

HIV-1 GD stems from at least three different sources: multiple introductions of HIV-1 into the human population [11–13], the low fidelity and high recombinogenic power [14,15] of its reverse transcriptase [16] and its high virus turnover [17]. HIV-1 and -2 GD has been classified within discrete groups and subtypes that largely correspond to geographic regions (Figure 1) [18]. HIV-1 includes four groups, which represent different introductions into human populations, namely group M (Main, Major), N (New, Non-major), O (Outlier) and recently P [19]. HIV-1 group M is the most commonly detected variant, which, in turn, is further divided into subtypes – that is, A–D, F–H, J and K. Also, circulating recombinant forms (CRFs) carrying genetic information from two or more subtypes have been detected in infected populations (49 to date) [20]. HIV-2 includes two groups (A and B) and it is worth noting that HIV-2 CRFs have been reported only once [21]. Genetic variation among HIV subtypes is tremendous, with within-subtype divergences reaching up to 17% and between-subtype divergences of 17–35%. For comparison, human and chimpanzee divergence can reach up to 3.9% if substitutions and insertions/deletions are considered [22].

Early on, researchers noted differences in transmissibility between HIV-1 and -2 [23]. Some transmission differences are explained on the basis of structural and evolutionary differences in *env* genes [24]. It is clear that HIV-2 exhibits lower rates of transmission, almost no vertical transmission and long incubation periods [25]. Moreover, HIV-2-infected patients have reduced immune activation, low viremia and...
rarely develop disease [26]. Within HIV-1, some researchers have proposed biological correlates to different subtypes; for example, HIV-1 subtype A-infected women are less likely to develop AIDS than non-subtype A-infected women [27]. Also, differences among subtypes have been reported in relation to chemokine coreceptor usage (tissue tropism) [28] and transmission [29]. Other researchers have suggested that the ‘subtypes’ are simply artifacts of poor or selective sampling of the overall HIV-1 GD [30].

**Measures of GD**

Given the interaction between HIV genetic and epidemiological dynamics, accurately estimating GD becomes particularly important. GD can be directly estimated from nucleotide sequence data using a variety of approaches.

**Traditional measures of GD**

The most intuitive way of measuring GD would be to simply count the pairwise differences between sequences in a sequence alignment or the number of polymorphic sites [33]. This approach, however, has some limitations, such as the existence of multiple hits in the same position, different probabilities of change in coding sequences or transition/transversion bias. In general, measures of GD based on sequence data can be classified into two basic categories: summary statistics and coalescent estimators. Some commonly used summary statistic approaches are: nucleotide diversity [32], haplotype diversity [32], allelic diversity, gene diversity and theta (θ) [33]. More recently, θ, expressed as substitution rate-scaled effective population size, has been estimated under a coalescent framework.
explicitly taking into account evolutionary history [34–37], which differentiates this model from approaches based on summary statistics. The coalescent model has been further generalized to account for varying population sizes, different time scales, structure, recombination and selection, which has probably made it the most used method to estimate GD. Detailed descriptions regarding algorithms for \( \Theta \) estimation can be found in other reviews [38–40].

**Novel approaches**

Newly developed approaches intend to estimate diversity parameters by taking advantage of the massive amounts of data that next-generation sequencing (NGS) technologies can deliver [41]. New methodologies have focused on characterizing intrahost diversity by capturing low-frequency variants [42]. Recent implementations take advantage of Bayesian inference to correct errors [43] and infer haplotypes and their frequencies (as low as 0.1% [44]) [45]. Also, the determination of HIV full genomic sequences and related measures of diversity are now feasible, which opens unexplored possibilities to comprehensively address how HIV mutates under selective pressures [46]. To date, most of the applications of NGS technologies to ‘ultra-deep sequencing’ of viral populations have focused on drug resistance characterization, fluctuations in GD through disease progression, and on certain events in HIV biology, such as tropism switch, transmission bottlenecks, immune escape [47,48], epistasis [49] and superinfections [50]. Drug resistance characterization has been performed primarily on target genes such as RT and Integrate (both within pol; Figure 2). In turn, epidemiological and subtyping studies primarily focus on the capsid proteins encoded within the env reading frame [51,52]. In the future, we expect to see more applications based on full genome data as the cost of sequencing decreases [42,53].

**Geographic & temporal spread of diversity**

PhyloDynamics [54], or the description of infectious disease behavior that arises from the blending of evolutionary and epidemiological processes, has become a hot subject in virology and epidemiology, especially following recent statistical developments (see Table 1) [55–69]. These new methods often use the coalescent under a maximum likelihood or Bayesian inference framework and have been applied in HIV to study questions related to its global, regional, local and within-host dynamics.

**Estimating phyloDynamics**

Accurate phyloDYNAMIC estimation relies on an adequate sampling strategy [70]. Sparse sampling could lead to inappropriate inferences – for example, the east Africa direct transmission of South American HIV-1 subtype C [71]. Since phyloDYNAMIC inferences rely on ‘time trees’ or dated phylogenies, it is also necessary to calibrate the molecular clock model in use. Due to the lack of fossil records in HIV, time-stamped sequence data can be used to produce those inferences. Also, incorporating independent prior knowledge about substitution rates will help any virus dating effort, generally increasing statistical power. Specialized
Table 1. Summary of software used for phylodynamic inferences.

<table>
<thead>
<tr>
<th>Inference</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration, spatial dispersion</td>
<td>Migrate-n [145], BEAST [146], IMa2, Lamarc [147]</td>
</tr>
<tr>
<td>Substitution rates</td>
<td>BEAST</td>
</tr>
<tr>
<td>Recombination rates and recombination breakpoints</td>
<td>Lamarc, LDhat [148], RDP3 [149]</td>
</tr>
<tr>
<td>Changes in population sizes</td>
<td>Migrate-n, BEAST</td>
</tr>
<tr>
<td>Divergence time estimation</td>
<td>BEAST, Multidivtime</td>
</tr>
<tr>
<td>Leaf ages</td>
<td>BEAST</td>
</tr>
<tr>
<td>Reproductive number</td>
<td>BEAST</td>
</tr>
<tr>
<td>Growth rate</td>
<td>BEAST, Lamarc, migrate-n</td>
</tr>
<tr>
<td>Population divergence times</td>
<td>IMa2</td>
</tr>
<tr>
<td>Haplotype reconstruction from NGS data and genetic diversity estimation</td>
<td>ShoRAH</td>
</tr>
<tr>
<td>Detection of selection</td>
<td>HyPhy [150], ADAPTSITE [151], TreeSAAP [152]</td>
</tr>
<tr>
<td>Assigning samples to populations, inferring the number of populations</td>
<td>Structurama, Structure, StructHDP</td>
</tr>
<tr>
<td>Ancestral state reconstruction</td>
<td>BEAST, MESQUITE [153]</td>
</tr>
</tbody>
</table>

NGS: Next-generation sequencing.

databases – for example, influenza [72] and HIV [201] – may help in this regard, as they can provide more clinically relevant information along with the genetic data (e.g., time of collection data).

In essence, most phylodynamic methods capitalize on analyzing distributions of trees usually obtained by sampling from the posterior distribution of a model, given the data. They work under the theoretical realization that the shape of a tree reflects dynamic processes impacting the data, such as population size changes (constant size, growth and shrinkage), selective processes (intra-host immune selection) or spatial structure (Figure 3). For example, populations with constant size are predicted to give symmetrical phylogenetic patterns, with most of the diversity happening at relative short branch lengths [73]. On the other hand, a population experiencing exponential growth will have longer branches leading to extant (sampled) sequences relative to deeper branches in the phylogeny [73]. In short, the coalescent models used in phylodynamic analyses describe a probability distribution on ancestral genealogies, given a population history. Therefore, if we can estimate the underlying phylogenetic structure of alleles from a collection of sequence data, by extension we can infer their population history and the evolutionary processes impacting that history (Figure 3).

Global spread

The best hypothesis we have regarding the origin and dispersion of HIV indicates that HIV-1 and -2 originated in Africa during the first half of the last century, and that it was the product of several cross-species transmissions between humans and non-human primates [11–13] (see [51] for further reading). Globally, approximately 33 million people were living with HIV worldwide as of 2009. In the same year, 2 million infected people died and the disease grew at a rate of 7400 new infections per day, more than 97% of which occurred in low- and middle-income countries [74]. Global distribution of groups and subtypes has remained rather constant within the last 10 years [75]. Although both HIV lineages spread exponentially at the beginning of the epidemic, HIV-2 occurred mostly in western Africa. In turn, HIV-1 is distributed worldwide, group M being the one that accounts for most infections, while groups O, N and P appear to be concentrated in central Africa. The time to the most recent common ancestors (TMRCA) of HIV lineages has been dated using different molecular-based methodologies and gene regions with the following inferences: 1905–1942 for HIV-2 group A and 1914–1945 for group HIV-2 B [76,77]. Similarly, HIV-1 TMRCA estimates were as follows: 1894–1931 for group M, 1932–1966 for group N and 1914–1925 for group O [13,76,78].

Regional spread

Founder-effect events are thought to play a major role in the spread of HIV out of Africa, although other factors cannot be ruled out completely, such as viral selective advantages, sociocultural factors and human genetic background. The Democratic Republic of Congo (DRC) is one of the places in which HIV-1 diversity is the greatest, and probably the site where cross-species transmission occurred [79–81]. Two archival samples, DRC60 and ZR59 [82], and the existence in DRC of almost all group M subtypes [75,83] support this statement (Figure 4).

Studies worldwide have attempted to infer regional spatial and temporal spread of HIV-1 in particular. In the USA, HIV-1 B seems to have emerged from a single migration out of Haiti in 1969, the place with the highest subtype B diversity. In turn, Haitian HIV-1 B emerged in 1966 from the DRC [84–86].

South Africa has recently shown an increase in HIV infections. Almost 6 million people are infected, the majority with HIV-1 subtype C, which now accounts for 50% of all infected
The C subtype was first reported in 1990 and its TMRCA was dated to 1958. The spread of subtype C seems to have occurred eastward from South Africa to India, and also probably to China, while some founder events have also been identified from east Africa to South America and to Israel. The introduction of HIV-1 subtype C in South America goes back to the 1980s, most likely through Brazil. Nonetheless, in a more comprehensive study, South American subtype C appears to be more related to UK subtype C, with these two groups related to east African isolates, which stresses the importance of including global isolates in phylogenetic studies of HIV phylogeography.

Besides being present in the western world, HIV-1 subtype B is also present in Asia, where its introduction seems to have occurred through Thailand in 1985 and was termed subtype B’. From here, HIV-1 subtype B’ expanded into Asia, coexisting with the pandemic subtype B and others, and fueling the development of CRFs across the continent. It is worth noting that CRFs represent 20% of all HIV-1 infections, with half of these infections involving CRF02_AG and CRF01_AE. Additionally, there are some indications that certain subtypes are preferentially associated with behavioral factors, such as intravenous drug users and drug-trafficking regions, in particular within ex-USSR countries and southeast Asia (subtype A and subtype CRF01_AE, respectively).

**Figure 3. Phylodynamic patterns.** Population, selection and spatial dynamic patterns and their respective idealized trees.
HIV studies implementing phylodynamic methods have been used to address a variety of questions, including epidemic origins [84–90,93–96], correlations between epidemiological data and changes in population size or GD, viral spread over geographic regions, and evolutionary processes associated with certain risk groups, such as MSM and intravenous drug users. Some have shown that currently circulating subtype B has been introduced multiple times into MSM European (UK and Italy) populations in intervals ranging from 14 to 30 months [98–100]. Others have looked at dissemination patterns and transmission routes between risk groups in order to explain the distribution of HIV strains [101], suggesting a spread from heterosexual patients to intravenous drug users.

**Local transmissions**
Phylogenies are also useful for reconstructing transmission histories. For example, the Swedish transmission chain reconstructed transmissions in a case involving nine infected individuals, for whom the exact transmission history was known [102]. However, caution should be exercised, as convergence evolution might be a common theme in HIV-1 evolution [103]. Phylogenetic inference of transmission histories has also been accepted as evidence in court trials such as the Florida dentist case [104], a Swedish rape case [105], and healthcare-related cases in Baltimore [106] and Louisiana [107]. Bernard et al. present a detailed review on the use of HIV phylogenies in forensics [108].

HIV evolution has been thoroughly studied across transmissions because of the opportunity...
for treatment due to a reduction in GD [7,109]. Most of the work has focused on monitoring discordant couples (i.e., couples in which one partner is HIV positive and the other is not) and to test whether there is a reduction in GD down to one virus at the transmission event. Using phylogenetic methods, Edwards et al. showed that GD reductions (<1%) are no different between horizontal (homo- or heterosexual) and vertical transmissions (mother-to-child) [109]. Understanding GD at transmission events has therapeutic implications, as less diverse populations of small size are strongly influenced by genetic drift, decreasing the chance of transmission of high-fit variants. Indications that single virus variants were transmitted horizontally came from studies using Sanger sequencing and single-genome amplification coupled with phylogenetic and mathematical modeling [110]. These initial observations were additionally supported by the enhanced capabilities of ‘ultra-deep sequencing’ revealing that early HIV variants explored an extensive sequence space within epitope regions. Interestingly, as the infection proceeds, reversion to the canonical subtype sequence occurred in positions under immune selection pressure but not in positions that were not under selection pressure even in the earliest samples, suggesting that immune selection is acting earlier than previously thought [7].

Within-host dynamics

Different evolutionary forces are in action when we look at viral evolution within the host. The viral population is targeted by both cellular and humoral immune responses, resulting in relatively strong diversifying selection. As a result, when reconstructing within-host evolutionary histories, we observe a ladder-like pattern (continual immune selection in Figure 3 and [111]), as opposed to the strong spatial structure we observed in HIV population phylogenies (spatial dynamics in Figure 3). However, whether within-host HIV-1 evolution is governed by natural selection or genetic drift (deterministic or stochastic models, respectively) has been the subject of considerable debate. This dispute stems from disparate estimates of effective population sizes [112]: the smaller the population size, the more susceptible the population is to genetic drift. However, signatures of natural selection have been demonstrated using different estimators [113]. Recombination also plays an important role in shaping intra-host viral diversity. Conventionally, it can purge ‘bad mutations’ out of the gene pool and ‘put together’ novel combinations of genome regions, increasing allele diversity. Although it is not entirely clear whether recombination is associated with increased fitness, simulations have shown that under strong drug pressures, recombination will favor the appearance of drug resistance variants [114].

Within-host variation has been extensively studied because of the development of drug resistance, viral reservoirs and clinical conditions related to some tissues such as dementia and lymphoma. For example, phylogenetic analyses of post mortem brain tissues have revealed that HIV is evolving at different rates in different brain compartments. Apparently, this is not due to selective pressure, but rather to inherent drift associated with macrophage-tropic viral expansion after immune failure [115]. Within-host variation has been also addressed under a phylodynamic framework for coreceptor usage [116]. For example, coreceptor dynamics in tissues and peripheral blood mononuclear cells showed temporal structure between CCR5-tropic (R5) virus and the appearance of CXCR4 (X4) variants, suggesting that the majority of X4 virus found in thymus tissue seemed to come from peripheral blood mononuclear cell viruses [116]. Similarly, a high degree of compartmentalization and differences in population size have been described in HIV viruses from macrophages found in tumor and nontumor tissues and an intermixing of HIV strains obtained from axillary lymph nodes [117].

Population structure

Population structure refers to the degree of subdivision or differentiation a population exhibits. It has evolutionary consequences, as subdivided populations can evolve somewhat independently. Examples of structured HIV populations can be found in within-host samples from different tissues, which usually exhibit a large degree of compartmentalization [117], as well as in among-host samples, for instance founder effects and the global HIV-1 distribution (Figure 4). Traditional methods for estimating population structure include summary statistics that attempt to measure the diversity of randomly chosen sequence markers within the same subpopulation, relative to what is found in the entire population. These include the fixation index ($F_{st}$), as well as all its relatives ($G_{st}$, $R_{st}$, $D_{st}$). Summary statistical methods have been used to assess compartmentalization in lymphocyte reservoir populations [118,119].
New methodologies can also take advantage of the flexibility of the Bayesian framework to accommodate uncertainties in parameter estimation and to incorporate population history using the coalescent (see [120]). Bayesian approaches to infer population structure have been developed, for example, to estimate the posterior probability of viruses to belong to different populations (K) [121–123]. These methods could also be used to study HIV tissue differentiation, viral reservoirs, and their link to plasma GD.

**Implications of the GD of HIV**
Estimating the GD of HIV and mapping its distribution across space and time has clinical implications in terms of disease progression, drug resistance and vaccine development.

**Disease progression & drug resistance**
Disease progression is seemingly related to GD. Several studies have found a correlation between them when estimating relative and absolute substitution rates from time-stamped sequence data [10,124,125]. In general, there are three phases that can be identified on the basis of diversity: diversity increases linearly in association with initial features of infection and the appearance of X5 HIV populations; diversity levels off or even decreases, which could be correlated with the appearance of X4 HIV populations; and decline in CD4+ T cells and failure of T-cell homeostasis with a general reduction in GD. However, Carvajal-Rodriguez *et al.* have reported no relationship between disease progression and substitution rates when analyzed separately as adaptive and neutral categories of variation [9].

Disease progression is also related to transmitted drug resistance (TDR) and the clinical impact of low-frequency variants. TDRs are thought to exhibit diminished fitness compared with ‘wild-types’, although additional compensatory mutations might restore fitness levels [126]. Low-frequency variants, typically characterized using NGS, have clinical significance, especially when the drug-resistant mutation has a low genetic barrier – that is, few mutations instead of several mutations that confer resistance. Non-nucleoside reverse transcriptase inhibitors and protease inhibitors are examples of low and high genetic barriers, respectively [127,128]. When using NGS, TDRs seem fairly common (~30% for B and non-B subtypes) and subjects with multiple protease TDRs are infrequent [129]. However, Sanger sequencing reveals values of TDR prevalence of only 10% [130]. Knowing what TDR estimate is accurate remains an open question.

**Vaccine strategies**
Designing an HIV vaccine is an extraordinarily difficult challenge. A successful vaccine is typically thought to be capable of eliciting broadly cross-reactive neutralizing antibodies in order to cope with the extreme GD of HIV [121]. New vaccine designs have considered accounting for HIV geographical distribution and HLA allele frequencies in different countries or regions [120]. This makes sense in the light of the geographical structure of HIV (see above). Thus, in southwest China a vaccine must be capable of neutralizing mainly HIV-1 M C/B’ recombinants, whereas a South African vaccine must consider, among others, HIV-1 M subtype C variants (Figure 4) [131,132]. Although this strategy seems sound, several considerations have led to a search for a global vaccine. Even in countries where the HIV epidemic is dominated by one subtype, for example, India, there are other subtypes that might increase its frequency if the vaccine targets just one variant. Also, and apart from cost concerns, nations with high HIV prevalence have complex epidemics, such as central and southern Africa, which will be less impacted by single subtype vaccines (pie charts in Figure 4). Alternative strategies have tried to account for HIV GD by looking for potential T-cell epitopes within the env 3D structure [133].

Novel strategies for vaccine design include, but are not limited to, poly-epitope vaccines, the use of conserved regions on the proteome, central vaccines (ancestral state, consensus and center-of-the-tree) and polyvalent mosaic vaccines. For details, the reader is referred to Korber *et al.* [134]. Central vaccines are appealing from a phylogenetics standpoint and have also proven useful [135]. The idea is to maximize GD coverage by inferring the ancestral sequence of a particular clade or tree region, under the assumption that such a sequence, once translated, will elicit antibodies against any and every member of the clade or group of interest. In effect, central vaccines are designed to minimize the genetic distance to circulating strains. Other strategies include immunization against cell coreceptors such as the CCR5 chemokine receptor [136–138].

Recent research has focused on characterizing broadly neutralizing antibodies that occur naturally. Independently, and using different approaches, Wu *et al.* [139] and Scheid *et al.* [140] converged in identifying CD4 binding sites in antibodies derived from multiple unrelated HIV-infected individuals. These promising results
have obvious vaccine design implications as broadly neutralizing antibodies can protect new cells from infection.

**Conclusion**
Understanding the geographical and temporal spread of HIV GD is crucial for planning effective intervention strategies to combat infection, whether through educational programs, drug treatments or vaccine strategies. Is there a sequence type that is preferentially transmitted? If yes, is this sequence type similar among different transmission routes? What is the GD that needs to be covered by a vaccine? What proportion of TDR variants is clinically relevant? New sequencing technologies have the potential to improve our understanding of the patterns and processes that lead to the temporal and spatial spread of HIV and to answer the abovementioned questions. Methods such as ultra-deep sequencing have already impacted our understanding of HIV in terms of within-host variation: emergence and transmission of drug resistance, identification of super-infections, characterization of early dynamics and immune escape. At the population level, NGS has provided data to test for coevolving sites and added to our knowledge regarding the clinical impact of low-frequent variants. Then, phylogenetic and population genetic approaches have provided an appropriate probabilistic framework to analyze NGS data and understand better the impact of genetic diversity.

**Future perspective**
NGS data have already demonstrated their utility in HIV biology; nevertheless, there are sequencing applications and aspects of phylodynamics that are presently beyond the reach of current sequencing technologies, paving the way for additional innovation. It is imperative for the continuous progress of phylodynamics, then, to capitalize and incorporate in a ‘biologist-accessible’ manner the power of this technology and of the just-arriving ‘third-generation sequencing technologies’, namely single-molecule sequencing (PacBio, Helicos) and semiconductor sequencing technology (Ion Torrent)\(^\text{[14]}\). Additionally, an increase in the number of epidemiological models, coupled with population genetic and phylogenetic models, will provide researchers with more flexibility when applying these tools. Examples of this are models borrowed from epidemiology\(^\text{[67]}\) and other recent advances in estimating epidemiological parameters directly from sequence data\(^\text{[68,69]}\). For instance, models for cophylodynamics among coinfecting viruses that are transmitted similarly, notably HIV and HCV, will be necessary to explore the still controversial consequences of this coinfection\(^\text{[3,142–144]}\). Over the next few years, this race to generate large amounts of data should be coupled with a similar effort to expand our models of HIV evolution to incorporate these data into the analysis of HIV dynamics and disease progression.

**Executive summary**

- The high genetic diversity of HIV stems from at least three different sources, namely high mutation and recombination rates, high replication error and multiple introductions into human populations.
- Traditional measures of genetic diversity (summary statistics and coalescent estimators) have helped to understand HIV diversity and distribution. Next-generation sequencing is providing a new dimension to explore within-host and population diversity in depth.
- The integration of epidemiological and evolutionary processes through the use of sequence data under a phylogenetic framework (phyldynamic) has allowed us to investigate changes in population size over time, estimate relevant parameters as the reproductive number and substitution rates, and most notably, the dispersion of lineages over a geographic region.
- Through the use of phylogenetics, we now know that HIV-1 and -2 originated in central and west Africa, respectively. HIV-1, sub-classified in groups and subtypes, is present worldwide and exhibits a large geographic distribution, probably because of several founder-effect events.
- Evolutionary forces within and among hosts differ. Within-host HIV-1 evolution seems to be governed by continual immune selection, whereas among-hosts evolution seems to be driven by founder effects, as reflected in the strong geographic structure observed.
- Increased HIV genetic diversity has implications in almost every aspect of its biology, including vaccine design, drug treatment, disease progression, viral reservoirs, transmissibility and viral load testing. It also allows us to apply ambitious statistic models to interrogate these very same aspects. On the other hand, its high genetic diversity has prompted HIV as a model entity to test new developments in phylogenetics and evolutionary theory.
- Over the next few years, phylodynamic studies will couple with next-generation sequencing and third-generation sequencing to integrate the large amount of data into a robust and informative framework for testing hypotheses of HIV evolution, epidemiology and diversity. In turn, HIV biologists will need to better understand the strengths and weaknesses of these evolutionary and population genetic approaches to gain insights as applied to HIV data.
Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:
• of interest
• of considerable interest


8. Characterization of early HIV dynamics by ultra-deep sequencing, showing the reduction of genetic diversity at the point of transmission and the early diversification of transmitted viruses.


Excellent review covering next-generation sequencing technologies, their pros and cons and applications.


Original description of an algorithm that can take next-generation sequencing data and return genetic diversity measures and haplotype number and frequencies using a Bayesian framework.


Drummond A, Suchard M. Bayesian random local clocks, or one rate to rule them all. BMC Biol. 8, 114 (2010).


Excellent article that deals with phylogenetic patterns and their associated processes. It also describes algorithms to integrate epidemiological and phylogenetic models.


89. Relevant study showing the loss of GD across vertical and horizontal transmissions. It also illustrates the use of phylogenetic approaches to answer clinically important questions.
Genetic diversity & molecular epidemiology of HIV transmission

Review


Excellent review covering novel approaches to vaccine design and their known performances. It summarizes novel attempts to cope with HIV GD in the design of vaccines.


Website
201. HIV Databases. www.hiv.lanl.gov