Virus dynamics is the study of the population biology of viruses and the host cells within an infected individual.\(^1,2\)

In many cases, virus dynamics studies assume the host to be well-mixed for systemic infections, or model a single well-mixed compartment for localized infections. While this is a useful abstraction when the available data are limited to a single source—typically blood samples or tissue swabs—several important dynamical processes can be missed. Most mathematical models for human-immunodeficiency virus (HIV) and hepatitis C virus (HCV), for example, rely on viral load measures from the blood and describe a single well-mixed compartment.\(^1,2\) Herpes simplex virus 2 (HSV-2) or influenza virus,\(^4\) respectively.

From the moment of transmission, the anatomy of the host is important in the progression of a viral infection. Transmission itself often constitutes a major bottleneck, where only few virus particles manage to infect the targeted cell type, after which rapid expansion establishes the viral infection. Bottlenecks at transmission have been established, for example, for HIV,\(^5\) for HCV,\(^6\) and for influenza virus.\(^7\)

In several viral infections, the virus replicates in specific compartments. Often there is a main compartment where the majority of viruses reside and most of the replication takes place. The viral dynamics in the compartments can differ, as can the rates of exchange between them. These compartments can correspond to organs, tissues or cell types, depending on the virus or dynamical process studied. HIV, for example, replicates mainly in CD4\(^+\) T cells of secondary lymphatic tissues, while the liver is considered the main site of replication of HCV.

In chronic infections, a latent compartment can be established where infected cells lie dormant not producing any new virus for
extended periods of time. Since most antiviral drugs target virus replication, these latent compartments are not cleared by therapy. The paradigmatic latency is established by herpesviruses in nerve cells, from which they can reactivate under specific conditions, such as exposure to ultra-violet light or immune suppression. HIV also establishes latency in a small minority of its target cells, which is revealed after long and effective antiretroviral treatment. Similar to latent compartments are sanctuary sites where the virus is also protected from the immune response or antiviral drugs due to strong barriers between the sanctuary site and the other compartments. However, unlike in a latent compartment, the virus can typically replicate in a sanctuary site. For example, some viruses can pass the blood-brain barrier, entering the central nervous system. This is generally considered a sanctuary site, where there is only limited exchange with the bloodstream, allowing the virus to proliferate with little influence from the host’s immune system or antiviral drugs.

A better understanding of the spread of viruses across the various anatomical compartments in the host has the potential to improve treatment and vaccination approaches. Only if we know where the infection is initiated and the path it takes to become established in the host, can we target interventions to prevent infection efficiently. Information about the early processes will aid new vaccination approaches that aim to elicit immune responses targeted to specific anatomal sites. Furthermore, identifying bottlenecks during the colonization of the host might open up promising dynamical targets for intervention, and understanding the dynamics in latent compartments will help to prevent viral rebound.

In this review, we assemble what we know about the spatial dynamics of various virus infections. We approach this topic from two angles. First, we adopt the perspective of classical virus dynamics that has been extended to account for multiple compartments. Second, we will discuss the potential of applying population genetic and phylodynamic methods to viral sequence data obtained from various compartments within the infected hosts. This application has recently been called phyloanatomy.

Phyloanatomy has its roots in phylogeography and phylodynamics that combine phylogenetics—the building of lineage trees based on sequencing data—with dynamical models of disease spread and migration. Phylodynamics and phylogeography can thus elucidate the deeper evolutionary and population history from recently sampled sequences. They are widely used in epidemiology to quantify the epidemiological spread of viruses and to understand new epidemics in real-time. This allows to locate the source of an epidemic, as well as parameters such as population size and basic reproductive number \( R_0 \) and migration rates between different locations. In phyloanatomy, these well-established methods are applied to within-host viral sequencing data to elucidate the importance of certain organs or cell types in the progress of an infection.

With this review, we aim to bridge the two methodologically separate fields of virus dynamics and virus evolution. Because of the large scope of our undertaking, we did not aim to review the literature in every relevant field comprehensively. Rather, we chose to discuss studies that shaped our thinking and that we found particularly enlightening. We restricted our efforts to the dynamics and evolution of clinically relevant viruses of humans and their respective animal models.

2 | VIRUS DYNAMICS IN MULTIPLE CELLULAR AND ANATOMICAL COMPARTMENTS

2.1 | HIV dynamics

The most paradigmatic virus dynamics studies have been performed in the context of HIV infection under treatment. These studies, relying on data collected in the blood, modeled neither the anatomical structure within the hosts nor multiple cell compartments.

The decline of the viral load observed under treatment, however, was multiphasic, which was inconsistent with the simple single-compartment models and pointed toward dynamical complexities. While this multiphasic decline is commonly interpreted as evidence for the existence of cell compartments with lower viral turnover, explanations have been put forward involving redistribution between anatomical tissue compartments. Hlavacek et al. proposed that the deceleration of viral decline under treatment could be the result of the release of virus that had been attached to follicular dendritic cells in secondary lymphatic tissues.

More recently, De Boer et al. tried to resolve the discrepancy between estimates of the lifespan of HIV in plasma, which differ about 100-fold. The analysis of the viral kinetics under treatment yielded a lifespan of plasma virion of six hours. Based on data from plasma apheresis in humans, the lifespan was estimated as 20-75 minutes. Lastly, infusion of virions into the blood of rhesus macaques resulted in an estimate of three minutes. De Boer et al. proposed that the faster estimates are measures of the migration of virus from the blood into other anatomical compartments, rather than viral clearance. They estimate that HIV is cleared in lymphoid tissue at a rate of 10-100 per day, corresponding to a lifespan of 15-120 minutes.

In the course of antiretroviral treatment, the rate of viral decline in blood decelerates to almost zero before the virus is eradicated from the host. The virus persists in a latent reservoir that constitutes a serious obstacle for curing HIV. To this day, it is therefore a focus of intense research.

There are two main questions about this compartment. First, it is not clear if the latent compartment is anatomically separate from most of the infected cells during untreated infection. Second is the question of whether there is viral turnover and evolution in the latent compartment, or if the virus in this compartment is static. These questions are very difficult to study because the viral dynamics in the latent reservoir is observable only after many months of antiretroviral treatment that prevents the faster dynamics in productively infected cell compartments. At this stage, however, the viral load is usually not detectable.
The existing literature on classical HIV dynamics models with latency has been comprehensively reviewed by Rong and Perelson. In brief, on the basis of the existing virus load data, these models cannot definitively resolve the question if there is ongoing replication and evolution in the latent reservoir. Assuming that latent cell activation is the only source of residual plasma viremia, however, quantitative analysis suggests that turnover and hence evolution in the latent compartment is unlikely.27

2.2 Spatial dynamics of other viruses

Similar to HIV, many other viruses enter their host by infecting cells in epithelial tissues, and cause a systemic infection with one or more core compartments of replication. HCV, for example, can be sampled in the blood but replicates in the liver and potentially other anatomical sites. Some viruses, however, do not enter the blood and do not become systemic, such as HSV-2 that replicates in the epithelium and establish latency in nerve cells.

For no other virus is our quantitative understanding of the spread across different anatomical compartments more advanced than for HIV. For other viruses some spatial models have been developed, but on a spatial scale of a few hundred cells, rather than on the roughe anatomical scale. One of the questions on this finer scale is the balance between cell-to-cell spread of virus versus longer range transmission which has also been investigated in the context of HIV (reviewed by Graw and Perelson).28

For example, for HCV infection there are spatial models for the spread of the virus between hepatocytes.29 These models were based on single-cell laser capture microdissection data from liver biopsy samples of patients chronically infected with HCV, in which the HCV RNA content within infected cells has been quantified.30 Analysis of their spatial distribution indicated that infected cells occurred in clusters, with cluster sizes ranging from 4 to 50 cells. When the virus load determined in the few hundred cells in these microdissections was scaled up to the approximate 10^11 cells of the entire liver, there was very good agreement with plasma levels of HCV in each of the four patients analyzed. The data and the analysis support the hypothesis that HCV from the blood infects a random hepatocyte, and then spreads to adjacent cells. To which extent local cell-to-cell vs long-distance spread of free virions contributed to the spread of infection still remains to be determined and cannot be inferred from these data.

Regarding the dynamics of HCV between blood and liver which is more central to the topic of this review, there are no modeling studies to-date. In principle, such anatomically resolved modeling would be possible as viral loads in the liver can be determined in ongoing HCV infections. In Talal et al.,31 for example, the viral load under treatment was determined in blood and liver. This study also contains a mathematical model of the viral dynamics in blood and liver, but does not consider the exchange between these two compartments. Such studies would be interesting to understand the recolonization of the liver after a transplant. (See below for a description of a study that analyzed sequence data before and after transplant and hypothesized nonhepatic sites of replication.)

HSV-2 represents another infection, for which spatial aspects are central. While this infection affects a single anatomical site—the genital mucosa—it causes spatially well-defined lesions in this compartment. Hence, it is characterized by a strong spatial structure.

Mathematical models have been developed to capture the spatial spread of HSV-2 in the genital mucosa. Schiffer et al.33,34 used these models to explain the high variability in the length of shedding episodes, sometimes lasting many days. This was puzzling because the virus spreads to thousands of target cells in single lesions within hours, and immune responses, mostly tissue resident CD8 T cells, clear virus from a lesion within one day. From mathematical models that linked the fast viral dynamics in multiple lesions by more infrequent spread of free virus to distant sites, the variability in the length of viral shedding episodes and the seemingly erratic, often multipeaked viral load profiles emerged as a result of the metapopulation dynamics captured in these models. The parameters characterizing viral spread within and between lesions that were most consistent with the patterns of viral shedding, differed by three orders of magnitude.

All these studies allude to anatomical aspects as one of many ways to explain discrepancies between the prediction of single-compartment models and data obtained from a single compartment. Rarely are multiple anatomical compartments sampled. Even if they are, sampling is not sufficiently frequent to yield comprehensive time series. This problem could be alleviated by phylogenetic methods that have the potential to infer processes that occurred before the time of sampling.

3 APPLICATION OF POPULATION GENETIC AND PHYLODYNAMIC METHODS TO THE WITHIN-HOST DYNAMICS OF VIRUSES

As we discussed in the previous section, standard virus dynamics models can be used to reject simple single-compartment dynamics and to propose more complex dynamics that are consistent with the data. But, because of too sparse sampling of the time courses of virus and cell population sizes in all relevant compartments, they are often insufficient to determine the full multicompartment dynamics.

There are population genetic and phylodynamic methods that have been used successfully to identify compartmentalization and the dynamics in multiple compartments. These methods require not just measures of the population size, but information on the genetic composition of the viral population. This extra information allows these methods to extrapolate contemporaneously measured genetic diversity into the past. They accomplish this by making assumptions about the underlying evolutionary dynamics.

We will first discuss studies that tested for the compartmentalization of the virus population. After that we will review studies that used genetic information to estimate the parameters of the dynamics in multiple compartments.
3.1 Identifying compartmentalization

There are several classical population genetic methods to test for compartmentalization. In principle, these methods compare the genetic diversity within and between compartments. These have been applied to various viruses.

Tests for compartmentalization can be divided into distance- and tree based. Distance-based methods require some distance measure between the viral sequences, such as the Hamming distance. A well-known example of a distance-based test is the fixation index $F_{st}$ that has been adapted for the use of sequence data using the Hamming distance instead of allele frequencies. Intuitively, these distance-based tests compare the genetic distances of viral sequences within versus between compartments.

Tree-based methods, on the other hand, require the reconstruction of a phylogenetic tree before the test can be applied. The tests rely on the particular topology of the reconstructed tree. A classical tree-based method is the Slatkin-Maddison test that infers the minimum number of migrations necessary to explain the geographical distribution of the tips on a tree. Generally, the tree-based methods consider the degree of clustering of viral sequences from a given compartment in a phylogenetic tree. (Exact definitions of the statistics in these compartmentalization tests can be found in Zárate et al.)

In HIV infection, a compartmentalization of systemic vs central nervous system infection is well established. The compartmentalization between blood, secondary lymphatic system, genital tract mucosa, and gut-associated lymphatic tissue has also been investigated, but has not been unequivocally established (see table 1 in Feder et al.).

In their methodological study on a wide variety of compartmentalization tests, Zárate et al. reanalyzed clinical sequence data, and simulated sequence data. They found that the tree-based methods (Slatkin-Maddison test, Simmonds association index, correlation coefficients) are more sensitive than distance-based methods, $F_{st}$ and analysis of molecular variance (AMOVA).

The sequence diversity of the clinical samples analyzed by Zárate et al. most likely comprises escape mutations from cytotoxic T lymphocytes (CTL) that accumulate during the course of infection. Because CTLs—the selective agent of these escapes—are distributed systemically, these escape mutations can arise independently in different compartments. The simulated sequence data, on the other hand, were generated from a neutral model, according to which such parallel evolution is very unlikely.

In Box 1, we show how parallel evolution can confound compartmentalization tests. The effect applies predominantly to the distance-based methods, $F_{st}$ and AMOVA. In comparison to the simulations of sequence evolution including selection, simulations of sequence evolution have, for example, been used to assess compartmentalization tests. However, simulations of sequence evolution are typically based on an assumption of neutrality, that is, new mutations are introduced according to a mutation rate, but potential fitness effects of a mutation are not taken into account. Since selection is an important driver in virus evolution, simulation models that include selection can be a valuable tool. Additionally, in such simulations the real genealogy is known, which can help us understand the effects of selection on the quality and shape of the reconstructed phylogenetic tree. Below we illustrate the insights one can gain from these simulations in Figures 1 and 2. The simulations were based on a model of viral sequence evolution that includes selection in line with our recent study on the mutational fitness effects distribution (MFED) in early HIV infection. Our simulation is an extension of the model of sequence evolution by. It is a discrete time model, starting with a single, randomly generated sequence and an associated fitness table, which contains the fitness effect of every possible mutation in the sequence. These fitness effects are drawn from the MFED of the virus being simulated. The fitness of a sequence is the product of the fitness effects of all mutations it carries. In every generation of the simulation, sequences generate offspring based on their fitness. Offspring sequences acquire new mutations according to a mutation rate $\mu$. A simulation of neutral evolution can be recovered by setting the MFED to be a point-mass at one.

We extended our recent simulation model to capture viral diversification in multiple anatomical compartments, and viral migration between them. Each compartment can be assigned different simulation parameters, modeling the differences in viral dynamics in different compartments. For example, selection pressures might be the same in two populations (parallel evolution). This could be modeled by using the same fitness table in both populations. Diversifying evolution can be implemented assigning compartment-specific values to some entries in the fitness effects table. This way, some mutations that are beneficial in one population might be deleterious in another.

In Box 1 and Figure 1 we show how parallel evolution can confound compartmentalization tests. The effect applies predominantly to the distance-based methods, $F_{st}$ and AMOVA. In comparison to the simulations of sequence evolution including selection, simulations of sequence evolution have, for example, been used to assess compartmentalization tests. However, simulations of sequence evolution are typically based on an assumption of neutrality, that is, new mutations are introduced according to a mutation rate, but potential fitness effects of a mutation are not taken into account. Since selection is an important driver in virus evolution, simulation models that include selection can be a valuable tool. Additionally, in such simulations the real genealogy is known, which can help us understand the effects of selection on the quality and shape of the reconstructed phylogenetic tree. Below we illustrate the insights one can gain from these simulations in Figures 1 and 2. The simulations were based on a model of viral sequence evolution that includes selection in line with our recent study on the mutational fitness effects distribution (MFED) in early HIV infection. Our simulation is an extension of the model of sequence evolution by. It is a discrete time model, starting with a single, randomly generated sequence and an associated fitness table, which contains the fitness effect of every possible mutation in the sequence. These fitness effects are drawn from the MFED of the virus being simulated. The fitness of a sequence is the product of the fitness effects of all mutations it carries. In every generation of the simulation, sequences generate offspring based on their fitness. Offspring sequences acquire new mutations according to a mutation rate $\mu$. A simulation of neutral evolution can be recovered by setting the MFED to be a point-mass at one.

We extended our recent simulation model to capture viral diversification in multiple anatomical compartments, and viral migration between them. Each compartment can be assigned different simulation parameters, modeling the differences in viral dynamics in different compartments. For example, selection pressures might be the same in two populations (parallel evolution). This could be modeled by using the same fitness table in both populations. Diversifying evolution can be implemented assigning compartment-specific values to some entries in the fitness effects table. This way, some mutations that are beneficial in one population might be deleterious in another.
sweep once. Rather HIV and simian immunodeficiency virus (SIV) infections are characterized by sequential fixation of CTL escape mutations. Interestingly, this pattern of sequential fixation is broken by depleting CD8+ T cells in rhesus macaques without reducing the evolutionary rate or changing the degree of compartmentalization. More recently, Feder et al. studied compartmentalization using viral sequence data sampled from the blood, lymph nodes, and mucosal tissue associated with the gut or genital tract of monkeys receiving sequential monotherapy. The genetic makeup of the viral population comprised a high frequency of drug resistance mutations. The authors concluded that there is transient compartmentalization. It would be interesting to study if such transient compartmentalization can arise when standard tests, assuming neutrality, are applied to systems with strong cross-compartment selection. This could happen if identical resistance mutations appeared at different times in two compartments. The mutation would then go to fixation in the compartment it appeared in first. If the resistance mutation in the second compartment was still at low frequency, a compartmentalization test would be positive. As soon as the resistance mutation goes to fixation in the second compartment, the statistical signal for compartmentalization might disappear.

For viruses other than HIV, our insights into the compartmentalization of the dynamics are more rudimentary. While the differences in the genetic composition of HCV between liver and periphery have been investigated for many decades, formal establishment of compartmentalization is rare. For HCV, there seems to exist a similar separation between blood and central nervous system as in HIV but only in cognitively impaired patients. There is also evidence for compartmentalization between healthy and tumorous tissue based on the Mantel test in the liver.

Interestingly, the methods used to test for compartmentalization required the compartments to be known. This means that each viral sequence needed to be associated with a specific compartment from which it was sampled. There are population genetic methods that allow the identification of compartments that are a priori unknown. These methods have been extensively used to characterize the host or the vector diversity of virus infections but not for the viruses themselves, certainly not within host. One reason for the fact that these methods are not applied to viruses may be that they assume independent loci and evolutionary equilibrium, which is rarely fulfilled during any infection within the host.

Phylodynamic methods that allow insights into the number of relevant compartments from transmission trees have been conceived. The further development of such methods would be a fruitful direction for future research. This would particularly benefit the study of the within-host dynamics of viruses, in which the nature and number of relevant compartments is still unknown.

3.2 | Phyloanatomy

Phylodynamic methods are increasingly popular in macroevolutionary studies and epidemiology. In brief, these methods reconstruct one or multiple phylogenies from viral sequences, and use them to
estimate population dynamical parameters. This is accomplished by considering the phylogeny to be a realization of the assumed underlying population biological and sampling processes. Different population dynamical scenarios have been conceived, such as temporally changing population sizes ("skyline")\(^{54-57}\) and migration between multiple geographical locations ("phylogeography").\(^{14,17}\) The migration models have been used in epidemiology to pinpoint the origin of infectious diseases, and to reconstruct the spread of an infectious disease across different geographical locations.\(^{14,15}\) These methods have been implemented in dedicated software packages such as BEAST,\(^{58,59}\) or MrBayes.\(^{60,61}\)

In principle, these phylogeographic methods can be used also to infer the viral dynamics between different tissue compartments within the infected host. This application has been termed phyloanatomy\(^{9-11}\) or viral gene flow,\(^{62}\) and requires viral sequences from the compartments in question sampled at different time points throughout the infection.

There exist only few such studies to date. In a review of phylodynamic methods, Salemi and Rife\(^{10}\) highlighted the clear anatomical separation between the blood and the central nervous system in the context of HIV dynamics. Most of the studies they discuss investigate compartmentalization without providing viral population sizes and migration rates between the compartments. In one of their own studies on SIV-infected, CD8-depleted macaques, they could estimate the viral population sizes in the periphery and the brain, and the timings of migration between these two compartments.\(^{62}\) Their study was based on peripheral viral sequences sampled at several time points during infection, and viral sequences from the brain sampled after death. Viral sequence data obtained after death allows only the identification of rare migration paths, as the one between blood and brain. To infer migration between systemic compartments with potentially higher rates, earlier and more frequent sampling will be required.

Another study that Salemi and Rife\(^{10}\) cite in their review is Cybis et al.,\(^{63}\) in which viral flow between plasma, CD4\(^+\) and CD8\(^+\) T cells is investigated. This study identifies potential migration paths by Bayes factor, but does not estimate migration rates between compartments.

A recent study by Lorenzo-Redondo et al.\(^{11}\) investigated replication and migration of HIV-1 in blood and lymph nodes to address the question whether there is evolution in the latent reservoir. To this end, they use high-throughput sequencing data of HIV-1 DNA in cells from blood and lymph nodes, collected from three subjects before treatment initiation, and at three and six months after treatment. The main point of this paper was that the viral phylogenies are temporally structured during treatment, which they interpret as evidence for ongoing viral replication and evolution in the latent reservoir.

This conclusion has been heavily debated. Rosenbloom et al.\(^{64}\) pointed out that the original study may have been measuring the evolution that was ongoing in the shrinking nonlatent reservoir during the first six months of treatment. Other studies could not recapitulate that the phylogenies are temporally structured.\(^{65,66}\)

More relevant to our review are the more direct anatomical insights the study by Lorenzo-Redondo et al.\(^{11}\) provides. Using traditional measures of compartmentalization (\(F_{st}\)), they find that blood and lymph are dynamically separated. By using phylodynamic methods, they estimate migration rates between these compartments, and find higher migration rates from lymph to blood than in the reverse direction. They observe that haplotypes in the blood are always derived from haplotypes observed earlier in the lymph nodes, from which they derive that there is no source of virus to the blood other than the secondary lymphatic system. It may also indicate that the virus in the various lymph nodes of the secondary lymphatic system is not compartmentalized.

In the context of HCV infection, the case for phylodynamic analysis was made a few years ago.\(^{67}\) Yet, very few such analyses have been performed to-date. Phylodynamic methods have been applied to determine changes in the viral population size within the host,\(^{6}\) and identified two bottlenecks, one immediately after transmission, and another immunity-driven bottleneck approximately 100 days after infection. But this study did not consider any anatomical aspects. Gray et al.\(^{32}\) present an analysis of HCV sequences obtained from the blood of 10 patients who received a liver transplant. Based on sequences from before and after liver transplant, the authors could not find evidence for transplantation constituting a genetic bottleneck. This suggests the existence of a major compartment for HCV replication other than the liver, and implies high migration rates between the liver and this compartment. While this study focused on the anatomical compartmentalization of HCV dynamics into liver and periphery, it did not provide quantitative estimates of migration rate or compartment-specific replication rates.

Despite the promise of phyloanatomy, especially in terms of yielding dynamical parameter estimates, there is a scarcity of studies applying this approach. This is surprising to us, in particular, considering how popular phylodynamic methods are in epidemiology. If the virus evolves fast enough within its host and causes a sufficiently long infection, the viral population will diversify enough to enable a phylodynamic analysis. Thus, phylogenetics should be applicable to many chronic viral infections. Compared to the epidemiological context, sampling multiple compartments within the host is logistically and financially less challenging, although the ethical and clinical concerns of sampling certain compartments can admittedly be prohibitive. Additionally, in experimental systems, repeats and biological manipulations are possible which would allow the fine-tuning of the phyloanatomic inference.

### 3.3 Barcoding

Recently, "barcoding," or "genetic tagging," has become popular in many biological systems, such as viruses, bacteria, and cells. Barcodes are genetic tags introduced into the genome of the target virus, organism, or cell. This allows the tracking of individuals and their offspring throughout their life cycle.

Combined with neutral population genetics, data obtained from barcoded populations can, similar to phyloanatomy, yield estimates
of dynamic parameters. Unlike in phylodynamic approaches, diversity does not need to build up naturally. Rather one starts with a set of distinct barcodes that are typically evolutionarily static. The gain in information from barcoded population data stems from the stochastically fluctuating frequencies of tagged subpopulations that provide dynamical insight on a much finer scale than simple measurements of the total population size.

One of the main advantages of barcoding compared with natural diversity is that their effect on fitness is less pronounced. In some systems, there is even evidence for neutrality of these tags, as, for example, in Salmonella typhimurium,70,71 and in SIV.72 Barcoding can also be used in systems in which natural diversity would not build up fast enough to conduct a phylodynamic analysis, as, for example, in bacterial infections or cell biological applications. Furthermore, the distribution of tags in the inoculum can be controlled. It has been shown in the context of T-cell differentiation models that uniform tag distributions are most promising for parameter estimation.71 The downside of barcoding is that it is an experimental technique, and therefore almost exclusively applicable in animal models. The use of barcoding in a clinical setting is very restricted, although human challenge studies with influenza have been conducted.72,73 and thus the use of barcoded influenza strains in humans is conceivable.

In the context of viruses, genetic tags can be engineered into noncoding regions of the viral genome.70,73–75 This approach has been recently adopted for SIV in order to study the dynamics of the latent reservoir.74,76 A 34-nucleotide-long cassette containing a 10-random nucleotide barcode was inserted between the vpx and vpr genes of SIVmac239, yielding a theoretical upper bound of $4^{34}$ unique barcodes. These engineered viruses were used to challenge rhesus macaques that were subsequently put on antiretroviral treatment with multiple drugs. Treatment was then withdrawn after varying times, and rebound virus was analyzed.

Analyzing the structure of the rebounding viral population in terms of its barcode composition allowed Fennessey et al.74 to calculate the rate of reactivation from the latent reservoir. The reactivation rates they estimate range from 1 every 3 days to 1 every 1.5 days after 300 or more days of treatment. There was no association of reactivation rate with time of treatment initiation or length of treatment (as long as treatment duration was long enough to eliminate cell compartments in which virus replicates actively).

What is gained by using barcodes in this system? The reactivation rate would be difficult to calculate from the time course of the viral rebound alone. The first reactivation is confounded by a phase of unknown duration required for the drugs to be washed out. By using the barcodes, the time intervals between reactivations of individual, barcoded viral subpopulations can be inferred. The barcoded subpopulations thus increase the number of events that form the basis of estimating the reactivation rate, increasing the confidence of the estimate.

Genetic tagging has been applied beyond viruses. There are numerous studies, which tag bacteria with the aim to elucidate the bacterial infection dynamics. Bacterial infection dynamics are often compartmentalized. For example, in a mouse model for salmonellosis, the bacterium S. typhimurium establishes a population in the gut after oral challenge. From the gut bacteria can cross the gut wall and enter draining lymph nodes, from which they can then migrate further into other systemic sites, such as the liver, spleen, and gall bladder.

Research in this area is focused on so-called bottlenecks, which represent critical processes during the infection, such as the colonization of a new compartment. Bottlenecks are of interest because they represent points in the bacterial dynamics that, when targeted by drugs or vaccines, have the potential to cut the infection short. From the perspective of population dynamical modeling, bottlenecks are often equivalent to low migration rates between compartments in these systems.

Using tagged strains in a mouse model of Salmonella infection, we have estimated the migration rate between gut and mesenteric lymph nodes to be 300 bacteria per day.69 This migration rate is modulated by inflammation and innate immunity,69 and vaccine-induced immunity.79

Bacterial replication rates in different compartments can also be estimated. This has led to the insight that the replication and drug-induced clearance of bacteria in the mesenteric lymph node decreases during the course of antibiotic treatment, resulting in a static, persisting subpopulation.69 The so-called persister cells are not cleared by the antibiotic and can lead to relapse when antibiotic treatment is stopped. Dynamically, bacterial persistence is equivalent to the phenomenon of latency in HIV infection during highly active antiretroviral treatment, and bacterial relapse is analogous to viral rebound. As for HIV latency, one of the key questions is if the persister compartment turns over. Continuous evolution in the persister compartment is usually not considered to be relevant because of lower mutation rates in bacterial systems, although the occurrence of single mutations conferring antibiotic resistance is conceivable.

Genetic tagging is also being adopted to elucidate the dynamics of blood and immune cells. T cells, for example, have been tagged by various methods and the resulting data have been paired with population dynamical analyses to estimate dynamical parameters characterizing replication and differentiation.83,84

While the anatomical aspects in these systems do not take a central position in this research, the process of differentiation is formally similar to migration: instead of investigating at which rate a subpopulation of cells migrates from one compartment to the next, these studies estimate the rate at which T cells differentiate to become a member of another T-cell subpopulation. The analogy to viral latency or bacterial persistence in T cells is that memory T cells have a lower turnover rate, and seed the secondary response upon reinfection.

The main result of the studies by Buchholz et al.84 is that the T-cell response in these systems unfolds according to a linear differentiation pathway, going from naive to central memory to effector memory to effector cells. This pathway has challenged the consensus that existed in this field for a long time that memory T cells are

Genetic tagging has been applied beyond viruses. There are numerous studies, which tag bacteria with the aim to elucidate the bacterial infection dynamics. Bacterial infection dynamics are often compartmentalized. For example, in a mouse model for salmonellosis, the bacterium S. typhimurium establishes a population in the gut after oral challenge. From the gut bacteria can cross the gut wall and enter draining lymph nodes, from which they can then migrate further into other systemic sites, such as the liver, spleen, and gall bladder.

Research in this area is focused on so-called bottlenecks, which represent critical processes during the infection, such as the colonization of a new compartment. Bottlenecks are of interest because they represent points in the bacterial dynamics that, when targeted by drugs or vaccines, have the potential to cut the infection short. From the perspective of population dynamical modeling, bottlenecks are often equivalent to low migration rates between compartments in these systems.

Using tagged strains in a mouse model of Salmonella infection, we have estimated the migration rate between gut and mesenteric lymph nodes to be 300 bacteria per day.69 This migration rate is modulated by inflammation and innate immunity,69 and vaccine-induced immunity.79

Bacterial replication rates in different compartments can also be estimated. This has led to the insight that the replication and drug-induced clearance of bacteria in the mesenteric lymph node decreases during the course of antibiotic treatment, resulting in a static, persisting subpopulation.69 These so-called persister cells are not cleared by the antibiotic and can lead to relapse when antibiotic treatment is stopped. Dynamically, bacterial persistence is equivalent to the phenomenon of latency in HIV infection during highly active antiretroviral treatment, and bacterial relapse is analogous to viral rebound. As for HIV latency, one of the key questions is if the persister compartment turns over. Continuous evolution in the persister compartment is usually not considered to be relevant because of lower mutation rates in bacterial systems, although the occurrence of single mutations conferring antibiotic resistance is conceivable.

Genetic tagging is also being adopted to elucidate the dynamics of blood and immune cells. T cells, for example, have been tagged by various methods and the resulting data have been paired with population dynamical analyses to estimate dynamical parameters characterizing replication and differentiation.83,84

While the anatomical aspects in these systems do not take a central position in this research, the process of differentiation is formally similar to migration: instead of investigating at which rate a subpopulation of cells migrates from one compartment to the next, these studies estimate the rate at which T cells differentiate to become a member of another T-cell subpopulation. The analogy to viral latency or bacterial persistence in T cells is that memory T cells have a lower turnover rate, and seed the secondary response upon reinfection.

The main result of the studies by Buchholz et al.84 is that the T-cell response in these systems unfolds according to a linear differentiation pathway, going from naive to central memory to effector memory to effector cells. This pathway has challenged the consensus that existed in this field for a long time that memory T cells are
produced from effector T cells. The newly determined pathway is consistent with the concept of “stemness”: after the first infection is cleared, a small population of memory T cells remain that can seed a response against secondary challenge, which then unfolds faster because the first differentiation step is already taken. By much finer barcoding, Gerlach et al. could investigate the individual differentiation rates of a large number naive T cells. They found large heterogeneity in the fates of individual cells, adding support for the fact that stochasticity is an essential feature of the differentiation dynamics.

Comparing the analysis of experimental data involving barcoding between the viral and nonviral systems, there are a few differences. First, the analysis of bacterial and cellular systems was firmly rooted in stochastic process theory. This allowed not just the identification of a single process rate, as for SIV infection, but opened the door to estimating multiple rate constants. Especially when several processes influence the dynamics similarly, such as migration into a compartment and replication therein, estimating the corresponding rate constants reliably is challenging. With barcoding, the resulting stochasticity of the dynamics can often be exploited to disentangle the contribution of competing processes.

Second, in the case of T-cell differentiation, the tagging paired with stochastic models can be used to identify the model structure. Thus, rather than just estimating the rate constants of potentially competing processes, the rich data from studies using barcodes can be used to compare models with very different structures. Buchholz et al., for example, considered hundreds of conceivable differentiation models.

In summary, due to the fact that barcoding is more controllable and potentially neutral, it can be a very powerful tool to elucidate the within host dynamics of pathogens and host cells. In the case of viruses, a fully stochastic modeling and inference framework might allow one to gain further insights that go deeper than those gained to-date. Especially in the context of anatomical spread, barcoding will allow one to identify migration pathways between multiple compartments.

### 4 | SELECTION AS A POTENTIAL CONFOUNDER IN PHYLOANATOMICAL ANALYSIS

Phylogenetic and -dynamic analysis works best if evolution is neutral. Interestingly, one of the main motivations for phylodynamics came from immune-mediated selection that could affect the shape of viral phylogenies. In one of the earliest phylodynamics studies, for example, the shape of influenza virus phylogenies was found to be consistent with an epidemiological model that includes short-lived, strain-transcending immunity.

Despite the central role of selection, both as a driver and a potential confounder of phylogenetic analysis, it is not usually captured quantitatively in phylodynamic inference schemes. Furthermore, selection is not considered explicitly when reconstructing genealogies or transmission trees, which provide the empirical input into phylodynamic analysis. Commonly variation in evolutionary rates is used to capture the effect of selection in phylogenetic reconstruction. However, from the perspective of population genetics, the relationship between the evolutionary rate and selection is too indirect. In our view, selection can best be captured in terms of fitness values of individual viral variants.

In the context of compartmentalization analysis, selection can affect the reconstructed phylogeny in various ways. If the selection pressures are the same across locations, parallel evolution can occur, where the same mutation arises independently in different locations. This might confound measures of evolutionary distance, which are often used in phylogenetics/dynamics, and measures to determine compartmentalization. The occurrence of independent, parallel mutations in multiple compartments can increase the estimated migration rate, because, assuming a neutral model, it is more likely that the mutation had arisen once and migrated rather than occurred twice independently. As shown in Figure 1, some commonly used methods fail to identify compartmentalization in a parallel evolution scenario even when the two compartments are not connected.

Conversely, selection pressures may differ between anatomical compartments, leading to diversifying evolution. In this scenario, mutations that are adaptive in one compartment can be detrimental or neutral in another. Migrants then have a lower chance of survival in the new compartment due to the high probability that they are maladapted, leading to lower effective migration rates and stronger signals for compartmentalization. Migration rates might be underestimated in this case, as only well-adapted migrants will survive.

Evolution under selection can have the advantage to generate diversity fast, thus providing viral variants diverse enough for phylodynamic analysis. This is because under neutral evolution, mutations can only rise in frequency due to drift, which is a slow process in large populations. The faster diversification will come at the cost of having to put up with the biases in the estimates of the population dynamic parameters discussed above. In particular parameter regimes, these biases might be buffered; in very large populations with high mutation rates, for example, beneficial mutations will arise independently in different compartments, but likely on different genetic backgrounds.

This will allow one to distinguish the independent mutations, thus alleviating the bias introduced by parallel evolution.

These effects are illustrated in Figure 2, where sequence evolution was simulated under neutral, parallel and diversifying conditions, leading to different tree topologies.

There is accumulating evidence for strong selection on infectious agents. In experimental model systems, involving bacteriophages as the pathogen and Escherichia coli or Salmonella as hosts, fitness is observed to increase dramatically over the course of evolution, and identical mutations emerge and fix across independent evolutionary lines. Fitness increases and parallel evolution was also observed in long-term HIV evolution experiments in vitro (van Opijnen et al. and our own unpublished experiments).
In HIV-infected cohorts, there is evidence for parallel evolution in early infection across individuals before strong immune responses are mounted. The degree of shared mutations cannot be explained by neutral evolutionary models and is indicative of strong evolutionary pressures. The observed patterns of shared mutations are most consistent with a fitness effects distribution involving beneficial, deleterious, and lethal mutations. In particular, while the majority of mutations is deleterious, approximately 5% of mutations is expected to have a beneficial effect across patients.

Also, for influenza virus, parallel evolution was observed in immunocompromised individuals, which indicates similar selection pressures in these hosts.

There is conflicting evidence to what extent parallel evolution impacts our ability to correctly infer the evolutionary history. While parallel evolution has been found to impair the reconstruction of the evolutionary history in the bacteriophage \( \phi X 174 \) experimental genealogies of the bacteriophage T7 could be correctly inferred by phylogenetic methods. Also, HIV transmission histories could be correctly inferred by phylogenetic methods. This remains an area of active investigation.

Selection pressures have been found to differ between anatomical compartments. The most extreme form of differing selection pressures arise when specific mutations are required to invade a compartment. While we are not aware of evidence for such an extreme form of diversifying selection, there is evidence for specific mutations that confer an advantage to invade certain compartments. In HIV infection, for example, the virus can evolve to switch the coreceptor it uses to enter target cells. Although there are still uncertainties about the population genetic or dynamic driver behind the switch from using CCR5 to CXCR4, it is known that the switch expands the range of cells that can be infected, and genetic signatures for the switch have been identified. In the context of bacterial infections, \( Haemophilus influenzae \) has been found to evolve higher rates of invasion from the nasal tissues into the blood. In poliovirus infection in mice, it has been observed that colonization of the central nervous system is associated with the diversity of the systemic strain. In this system, it appears to be the composition and the cooperation between viral strains, rather than specific mutations, that facilitate invasion.

Selection pressures within an infected host are likely to be larger than those on the epidemiological scale. For HIV and HCV infections, it is well established that the evolutionary rates within hosts are faster than those between hosts.

One of the major within-host selection pressures is exerted by CTL. Selection exerted by CTL responses has been estimated to have selection coefficients close to 100% in animal models, but can also be large in humans. The strength of this selection pressure is often determined by the rate at which CTL escape mutants outcompete the wildtype virus. Due to potential clonal interference between escape mutants, the selection pressure exerted by CTL might be underestimated. Also nonlytic mechanisms of CTL action that have been considered to lead to unspecific selection can, in theory, lead to escape if spatial aspects of the interaction are taken into consideration. Selection pressures of a strength

![Figure 2](image-url)
similar to those exerted by CTL have been estimated for antibody responses.\textsuperscript{122} Immune-mediated selection strength within hosts therefore dwarfs the typical selection coefficients in other systems, in which beneficial mutations confer a 10% increase in fitness.\textsuperscript{123} For example, one of the most advantageous alleles in recent human history confers lactase persistence and has a selection coefficient of 10%.\textsuperscript{124}

Moreover, immune-mediated selection pressures within the host are likely to be more homogeneous than on the epidemiological scale. On the one hand, this is due to the mobility of immune effectors between different anatomical compartments within the host.\textsuperscript{125,126} On the other hand, host heterogeneity—mediated by, for example, diversity on human leukocyte antigen—may lead to diversifying selection on the epidemiological scale, rather than to parallel evolution. We hypothesize that parallel evolution should therefore be more prevalent at the within-host scale. This does not apply to immune-privileged sites, which may explain the accumulating evidence for compartmentalization between the central nervous system and the periphery in the context of various viral infections.

### 5 | CONCLUSION

In this review, we have assembled approaches from different fields to elucidate the spread of viruses between different compartments within their hosts. We started by briefly reviewing mathematical models describing the dynamics of viruses between various compartments. Across many viruses, these multicompartment models have been proposed to explain discrepancies between observed viral load measurements and prediction from single-compartment models. We argued that, while the multicompartment models have the potential to explain viral load patterns, their validity can rarely be confirmed.

We then discussed the various population genetic methods to infer compartmentalization that have been applied to virus sequences obtained from multiple compartments within an infected host. These methods test against a null model, in which virus populations are well mixed, and do not provide quantitative estimates for migration or replication rates. Moreover, we showed that their sensitivity and specificity strongly depend on the nature and extent of selection (Figures 1 and 2).

Last, we considered approaches that yield dynamical parameter estimates: phyloanatomy and barcoding. In our view, phyloanatomy has great potential but can be confounded by selection. The combination of barcoding and neutral population genetics, on the other hand, is not necessarily confounded by selection, but is applicable only in experimental systems.

The effect of selection on phylodynamics is not yet understood, and little research has been done on the subject. While selection pressures exist on the epidemiological and ecological scales where phylodynamics is commonly used, we consider its effect much stronger on the within-host scale. Evolutionary rates are known to be higher on the within-host scale than on the between-host level, and there is strong evidence for parallel evolution in viruses on the within-host scale, both in experimental settings and in clinical data. We have illustrated that selection can be advantageous for phylodynamic inferences, because it can create sufficient diversity on shorter time scales. We also showed that the resulting migration rates between compartments might be over- or underestimated due to the presence of parallel or divergent evolution (see Box 1 and Figure 2). These stronger selection pressures indicate that a clarification of the role of selection is more important for phyloanatomy than for phylodynamic applications in the epidemiological setting.

We believe that the most promising path toward the reliable application of phyloanatomical methods to clinical data will require insights from both theoretical and experimental work. On the theoretical side, simulation of sequence evolution under specific selection regimes can help us understand how these specific selection pressures bias inferences. Experiments can help us assess exactly how much selection biases the estimates obtained from current methods, by working in a system where the dynamical parameters are known or can be independently determined. In vitro, a multicompartment model system could be set up, in which separate cell cultures represent compartments, and virus infection and migration can be controlled. In animal models, infection experiments with barcoded populations can be used to independently determine dynamical parameters, which can then be compared phylodynamic inferences from untagged sequences in the same system. The combination of theoretical and experimental work could lead to either new phylodynamic inference schemes specific to systems with strong selection, or to formulae that correct for biases arising from the application of classical phylodynamic methods in these systems.

### ACKNOWLEDGEMENTS

We thank Sebastian Bonhoeffer, Veronika Boskova, Judith Bouman, and Frederik Graw for valuable feedback on our manuscript. We gratefully acknowledge the financial support from the Swiss National Science Foundation (grant number 31003A_149769 to RRR).

### CONFLICT OF INTEREST

The authors declare that no conflict of interest exists.

### REFERENCES

1. Perelson AS. Modelling viral and immune system dynamics. Nat Rev Immunol. 2002;2:28-36.
2. Nowak MA, May RM. Virus Dynamics: Mathematical Principles of Immunology and Virology. Oxford: Oxford University Press; 2000.
3. Schiffer JT, Swan D, Prlic M, Lund J. Herpes simplex virus-2 dynamics as a probe to measure the extremely rapid and spatially localized tissue-resident T cell response. *Immunol Rev*. 2018;285:113-133.
4. Sobel Leonard A, McClain MT, Smith GJD, et al. Deep sequencing of influenza A virus from a human challenge study reveals a
selective bottleneck and only limited intrahost genetic diversification. J Virol. 2016;90:11247-11258.
5. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci USA. 2008;105:7552-7557.
6. Bull RA, Luciani F, McElnay K, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. PLoS Pathog. 2011;7:e1002243.
7. Sobel Leonard A, Weissman DB, Greenbaum B, Ghedin E, Koelle K. Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. J Virol. 2017;91:e00171-17.
8. Belyakov IM, Ahlers JD. What role does the route of immunization play in the generation of protective immunity against mucosal pathogens? J Immunol. 2009;183:6883-6892.
9. Shirreff G, Garcia V, Vanderford T, Silvestri G, Regoes R. The phyloanatomy of early SIV infection. Mathematical and Computational Evolutionary Biology. Saint Martin de Londres, France; 2013.
10. Salemi M, Rife B. Phylogenetics and phyloanatomy of HIV/SIV intra-host compartments and reservoirs: the key role of the central nervous system. Curr HIV Res. 2016;14:110-120.
11. Lorenzo-Redondo R, Fryer HR, Bedford T, et al. Persistent HIV-1 replication maintains the tissue reservoir during therapy. Nature. 2016;530:51-56.
12. Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303:327-332.
13. Volz EM, Koelle K, Bedford T. Viral phylodynamics. PLoS Comput Biol. 2013;9:e1002947.
14. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogenography finds its roots. PLoS Comput Biol. 2009;5:e1000520.
15. Faria NR, Rambaut A, Suchard MA, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014;346:56-61.
16. Stadler T, Kouyos R, von Wyl V, et al. Estimating the basic reproductive number from viral sequence data. Mol Biol Evol. 2012;29:347-357.
17. Vaughan TG, Kühnert D, Popinga A, Welch D, Drummond AJ. Efficient Bayesian inference under the structured coalescent. Bioinformatics. 2014;30:2272-2279.
18. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature. 1995;373:123-126.
19. Wei X, Gosh SK, Taylor ME, et al. Viral dynamics in human immunodeficiency virus type 1 infection. Nature. 1995;373:117-122.
20. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science. 1996;271:1582-1586.
21. Perelson AS, Ribeiro RM. Modeling the within-host dynamics of HIV infection. BMC Biol. 2013;11:96.
22. Hlavacek WS, Stilianakis NI, Notermans DW, Danner SA, Perelson AS. Influence of follicular dendritic cells on decay of HIV during antiretroviral therapy. Proc Natl Acad Sci USA. 2000;97:10966-10971.
23. De Boer RJ, Ribeiro RM, Perelson AS. Current estimates for HIV-1 production imply rapid viral clearance in lymphoid tissues. PLoS Comput Biol. 2010;6:e1000906.
24. Ramratnam B, Bonhoeffer S, Binley J, et al. Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. Lancet. 1999;354:1782-1785.
25. Zhang L, Dailley PJ, He T, et al. Rapid clearance of simian immunodeficiency virus particles from plasma of rhesus macaques. J Virol. 1999;73:855-860.
48. Goonetilleke N, Liu MKP, Salazar-Gonzalez JF, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J Exp Med. 2009;206:1253-1272.

49. Rife Magalis B, Nolan DJ, Autissier P, Burdo TH, Williams KC, Salemi M. Insights into the impact of CDB⁺ immune modulation on human immunodeficiency virus evolutionary dynamics in distinct anatomical compartments by using simian immunodeficiency virus-infected macaque models of AIDS progression. J Virol. 2017;91:e01162-17.

50. Maggi F, Fornai C, Vatteroni ML, et al. Differences in hepatitis C virus quasispecies composition between liver, peripheral blood mononuclear cells and plasma. J Gen Virol. 1997;78:1521-1525.

51. Tully DC, Hjerrild S, Leutscher PD, et al. Deep sequencing of hepatitis C virus reveals genetic compartmentalization in cerebrospinal fluid from cognitively impaired patients. Liver Int. 2016;36:1418-1424.

52. Harouaka D, Engle RE, Wollenberg K, et al. Diminished viral replication and compartmentalization of hepatitis C virus in hepatocellular carcinoma tissue. Proc Natl Acad Sci USA. 2016;113:1375-1380.

53. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945-959.

54. Strimmer K, Pybus OG. Exploring the demographic history of DNA sequences using the generalized skyline plot. Mol Biol Evol. 2001;18:2298-2305.

55. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22:1185-1192.

56. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969-1973.

57. Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014;10:e1003537.

58. Huelsenbeck JP, Ronquist F. MrBayes: Bayesian inference of phylogenic trees. Bioinformatics. 2001;17:754-755.

59. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19:1572-1574.

60. Strickland SL, Rife BD, Lamers SL, et al. Spatiotemporal dynamics of simian immunodeficiency virus brain infection in CDB⁺ lymphocyte-depleted rhesus macaques with neuroAIDS. J Gen Virol. 2014;95:2784-2795.

61. Cybis GB, Sinhelsma JS, Lemey P, Suchard MA. Graph hierarchies for phylogeography. Philos Trans R Soc Lond B Biol Sci. 2013;368:20120206.

62. Rosenbloom DIS, Hill AL, Laskey SB, Siliciano RF. Re-evaluating evolution in the HIV reservoir. Nature. 2017;551:E6-9.

63. Kearney MF, Wiegand A, Shao W, et al. Ongoing HIV replication during ART reconsidered. Open Forum Infect Dis. 2017;4:ofx173.

64. Van Zyl GU, Katsulisme MG, Wiegand A, et al. No evidence of HIV replication in children on antiretroviral therapy. J Clin Invest. 2017;127:3827-3834.

65. Gray RR, Salemi M, Klennerman P, Pybus OG. A new evolutionary model for hepatitis C virus chronic infection. PLoS Pathog. 2012;8:e1002656.

66. Grant AJ, Restif O, McKinley TJ, Sheppard M, Maskell DJ, Mastroeni P. Modelling within-host spatiotemporal dynamics of invasive bacterial disease. PLoS Biol. 2008;6:e74.

67. Van Zyl GU, Katusiime MG, Wiegand A, et al. No evidence of HIV-1 in the host-cell environment. PLoS ONE. 2007;2:e271.
93. Wood N, Bhattacharya T, Keele BF, et al. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. PLoS Pathog. 2009;5:e1000414.
94. Bertels F, Metzner KJ, Regoes RR. Convergent evolution as an indicator for selection during acute HIV-1 infection. bioRxiv. 2017;168620.
95. Bons E, Bertels FR, Regoes RR. Estimating the mutational fitness effects distribution during early HIV infection. bioRxiv. 2017;185678.
96. Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ. Experimental phylogenetics: generation of a known phylogeny. Science. 1992;255:589-592.
97. Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci USA. 1996;93:10864-10869.
98. Romero-Severson E, Skar H, Albert J, Leitner T. Timing and order of transmission events is not directly reflected in a pathogen phylogeny. Mol Biol Evol. 2014;31:2472-2482.
99. Volz EM, Romero-Severson E, Leitner T. Phylogenetic inference across epidemic scales. Mol Biol Evol. 2017;34:1276-1288.
100. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol. 1999;17:657-700.
101. Romano KJ, Regoes RR, Bonhoeffer S. The HIV coreceptor switch: a population dynamical perspective. Trends Microbiol. 2005;13:269-277.
102. Ribeiro RM, Hazenberg MD, Perelson AS, Davenport MP. Nave and memory cell turnover as drivers of CCR5-to-CXCR4 tropism switch in human immunodeficiency virus type 1: implications for therapy. J Virol. 2006;80:802-809.
103. Jensen MA, Li FS, van ’t Wout AB, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol. 2003;77:13376-13388.
104. Margolis E, Levin BR. Within-host evolution for the invasiveness of commensal bacteria: an experimental study of bacteremias resulting from Haemophilus influenzae nasal carriage. J Infect Dis. 2007;196:1048-1075.
105. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature. 2006;439:344-348.
106. Xiao Y, Dolan PT, Goldstein EF, et al. Poliovirus intrahost evolution is required to overcome tissue-specific innate immune responses. Nat Commun. 2017;8:375.
107. Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci USA. 1996;93:10864-10869.
108. Xue KS, Stevens-Ayers T, Campbell AP, et al. Parallel evolution of influenza across multiple spatiotemporal scales. Elife. 2017;6:e26875.
109. Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ. Experimental phylogenetics: generation of a known phylogeny. Science. 1992;255:589-592.
110. Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci USA. 1996;93:10864-10869.
111. Xue KS, Stevens-Ayers T, Campbell AP, et al. Parallel evolution of influenza across multiple spatiotemporal scales. Elife. 2017;6:e26875.
112. Fernandez CS, Stratov I, De Rose R, et al. Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytoprotective T-lymphocyte epitope exacts a dramatic fitness cost. J Virol. 2005;79:5721-5731.
113. Yates A, Shaw F, Barber DL, Ahmed R, Regoes RR, Antia R. Revisiting estimates of CTL killing rates in vivo. PLoS ONE. 2007;2:e1301.
114. Regoes RR, Yates A, Antia R. Mathematical models of cytotoxic T-lymphocyte killing. Immunol Cell Biol. 2007;85:274-279.
115. Asquith B, McLean AR. In vivo CDB8+ T cell control of immunodeficiency virus infection in humans and macaques. Proc Natl Acad Sci USA. 2010;107:6365-6370.
116. Mandl JN, Regoes RR, Garber DA, Feinberg MB. Estimating the effectiveness of simian immunodeficiency virus-specific CDB8+ T cells from the dynamics of viral immune escape. J Virol. 2007;81:11982-11991.
117. Asquith B, Edwards CTT, Lipsitch M, McLean AR. Inefficient cytotoxic T lymphocyte-mediated killing of HIV-1-infected cells in vivo. PLoS Biol. 2006;4:e90.
118. Ganusov VV, Goonetilleke N, Liu MKP, et al. 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. 2011;85:10518-10528.
119. Pandit A, de Boer RJ. Reliable reconstruction of HIV-1 whole genome haplotypes reveals clonal interference and genetic hitchhiking among immune escape variants. Retrovirology. 2014;11:56.
120. Garcia V, Regoes RR. The effect of interference on the CDB8+ T cell escape rates in HIV. Front Immunol. 2014;5:661.
121. Seich Al Basatena NK, Chatzimichalis K, Graw F, Frost SD, Regoes RR, Asquith B. Can non-lytic CD8+ T cells drive HIV-1 escape? PLoS Pathog. 2013;9:e1003656.
122. Bar KJ, Cy Tsao, Iyer SS, et al. Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. PLoS Pathog. 2012;8:e1002721.
123. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8:610-618.
124. Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG. The origins of lactase persistene in Europe. PLoS Comput Biol. 2009;5:e1000491.
125. Westermann J, Ehlers EM, Exton MS, Kaiser M, Bode U. Migration and phyloanatomy: Merging population dynamic and phylogenetic approaches. Immunol Rev. 2018;285:134–146.
126. Textor J, Henrickson SE, Mandl JN, et al. Random migration and T-lymphocyte killing. PLoS ONE. 2014;10:e1003752.
127. Lee HY, Giorgi EE, Keele BF, et al. Modeling sequence evolution in acute HIV-1 infection. J Theor Biol. 2009;261:341-360.