Title

A Relaxed Directional Random Walk Model for Phylogenetic Trait Evolution

Journal

Systematic Biology, 66(3)

ISSN

1063-5157

Authors

Gill, Mandev S

Ho, Lam Si Tung

Baele, Guy

et al.

Publication Date

2017-05-01

DOI

10.1093/sysbio/syw093

Peer reviewed

Syst. Biol. 66(3):299–319, 2017

DOI:10.1093/sysbio/syw093
Advance Access publication October 18, 2016

A Relaxed Directional Random Walk Model for Phylogenetic Trait Evolution

MANDEV S. GILL[1], LAM SI TUNG HO[2], GUY BAELE[3], PHILIPPE LEMEY[3], AND MARC A. SUCHARD[2,4,5,*]

[1]*Department of Statistics, Columbia University, New York, NY 10027, USA;* [2]*Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, CA 90095, USA;* [3]*Department of Microbiology and Immunology, Rega Institute, KU Leuven, Minderbroedersstaat 10, 3000, Leuven, Belgium;* [4]*Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA; and* [5]*Department of Human Genetics, David Geffen School of Medicine at UCLA, Universtiy of California, Los Angeles, CA, USA*
**Correspondence to be sent to: Departments of Biostatistics, Biomathematics, and Human Genetics, University of California, Los Angeles, 695 Charles E. Young Dr., South, Los Angeles, CA 90095-7088, USA; E-mail: msuchard@ucla.edu*

Received 17 December 2015; reviews returned 30 September 2016; accepted 10 October 2016
Associate Editor: Mark Holder

Abstract.—Understanding the processes that give rise to quantitative measurements associated with molecular sequence data remains an important issue in statistical phylogenetics. Examples of such measurements include geographic coordinates in the context of phylogeography and phenotypic traits in the context of comparative studies. A popular approach is to model the evolution of continuously varying traits as a Brownian diffusion process acting on a phylogenetic tree. However, standard Brownian diffusion is quite restrictive and may not accurately characterize certain trait evolutionary processes. Here, we relax one of the major restrictions of standard Brownian diffusion by incorporating a nontrivial estimable mean into the process. We introduce a relaxed directional random walk (RDRW) model for the evolution of multivariate continuously varying traits along a phylogenetic tree. Notably, the RDRW model accommodates branch-specific variation of directional trends while preserving model identifiability. Furthermore, our development of a computationally efficient dynamic programming approach to compute the data likelihood enables scaling of our method to large data sets frequently encountered in phylogenetic comparative studies and viral evolution. We implement the RDRW model in a Bayesian inference framework to simultaneously reconstruct the evolutionary histories of molecular sequence data and associated multivariate continuous trait data, and provide tools to visualize evolutionary reconstructions. We demonstrate the performance of our model on synthetic data, and we illustrate its utility in two viral examples. First, we examine the spatiotemporal spread of HIV-1 in central Africa and show that the RDRW model uncovers a clearer, more detailed picture of the dynamics of viral dispersal than standard Brownian diffusion. Second, we study antigenic evolution in the context of HIV-1 resistance to three broadly neutralizing antibodies. Our analysis reveals evidence of a continuous drift at the HIV-1 population level towards enhanced resistance to neutralization by the VRC01 monoclonal antibody over the course of the epidemic. [Brownian Motion; Diffusion Processes; Phylodynamics; Phylogenetics; Phylogeography; Trait Evolution.]

Phylogenetic inference has emerged as an important tool for understanding patterns of molecular sequence variation over time. Along with the increasing availability of molecular sequence data, there has been a growth of associated sources of information, such as spatial and phenotypic trait data, underscoring the need for integrated models of sequence and trait evolution on phylogenies, which promise to deliver more precise insights and increase opportunities for statistical hypothesis testing.

Much of the development of trait evolution models has been motivated by phylogenetic comparative approaches focusing on phenotypic and ecological traits. Traditional comparative methods assess the correlation between traits through standard regression models that assume taxa traits are generated independently by the same distribution. This assumption is obviously violated by taxa traits due to their shared ancestry. A proper understanding of patterns of correlation between traits can be achieved only by accounting for their shared evolutionary history (Felsenstein 1985; Harvey and Pagel 1991), and comparative methods focus on relating observed phenotype information to an evolutionary history.

Trait evolution has been tackled from another angle in phylogeographic approaches focusing on geographic locations rather than phenotypic traits. Evolutionary change is better understood when accounting for its geographic context, and phylogeographic inference methods aim to connect the evolutionary and spatial history of a population (Bloomquist et al. 2010). Phylogeographic techniques have allowed researchers to better understand the origin, spread, and dynamics of emerging infectious diseases. Examples include the human influenza A virus (Lemey et al. 2009b), rabies viruses (Biek et al. 2007; Seetahal et al. 2013), dengue virus (Bennett et al. 2010; Allicock et al. 2012), and hepatitis B virus (e.g., Mello et al. 2013).

While methods for phenotypic and phylogeographic analyses are developed with different data in mind, they address similar situations and it is appropriate to speak more generally of trait evolution. Two key components required for modeling phylogenetic trait evolution are a method for incorporating phylogenetic information and a model of an evolutionary process on a phylogeny giving rise to the observed trait values. Many popular approaches first reconstruct a phylogenetic tree and condition inferences pertaining to the trait evolution process on this fixed tree. However, computational advances, particularly in Markov chain Monte Carlo (MCMC) sampling techniques, have made it possible to control for phylogenetic uncertainty

(as well as uncertainty in other important model parameters) through integrated models that jointly estimate parameters of interest (Huelsenbeck and Rannala 2003; Lemey et al. 2010).

The evolution of discrete traits has typically been modeled using continuous-time Markov chains(CTMC) (Felsenstein 1981; Pagel 1999; Lemey et al. 2009a), analogous to substitution models for molecular sequence characters. For continuously varying traits, a discrete trait analysis via partitioning of the state–space into finite parts is sometimes the best approach (Boucher and Démery 2016). In many instances, however, stochastic processes with continuous state–spaces are more suitable. A popular choice to model continuous trait evolution along the lineages of a phylogenetic tree is Brownian diffusion (Felsenstein 1985; Hansen and Martins 1996; Martins 1999; Freckleton et al. 2002). Lemey et al. (2010) and Pybus et al. (2012) have recently developed a computationally efficient Brownian diffusion model for evolution of multivariate traits in a Bayesian framework that integrates it with models for phylogenetic reconstruction and molecular evolution. Notably, their full probabilistic approach accounts for uncertainty in the phylogeny, demographic history, and evolutionary parameters. Trait evolution is modeled as a multivariate time-scaled mixture of Brownian diffusion processes with a zero-mean displacement along each branch of the possibly unknown phylogeny.

While adopting a mixture of Brownian diffusion processes is a popular and useful approach, it may not appropriately describe the evolutionary process in certain situations. Such scenarios are more realistically modeled by more sophisticated diffusion processes. There may, for example, be selection toward an optimal trait value. To address this phenomenon, there has been considerable development of mean-reverting Ornstein–Uhlenbeck process models for trait evolution, featuring a stochastic Brownian component along with a deterministic component (Hansen 1997; Butler and King 2004; Bartoszek et al. 2012).

Another trait evolutionary process inadequately modeled by standard Brownian diffusion is one characterized by directional trends. The need for relaxing the assumption of a zero-mean displacement is highlighted by a number of evolutionary scenarios in which there are apparent trends in the direction of variations, including antigenic drift in influenza (Bedford et al. 2014), the evolution of body mass in mammals (Lartillot and Poujol 2011; Pant et al. 2014) and dinosaurs (Sookias et al. 2012), the evolution of plant defense traits (Agrawal and Fishbein 2008; Agrawal et al. 2009), and dispersal patterns of viral outbreaks (Pybus et al. 2012). To address this, more general Brownian diffusion models featuring nonzero estimable displacement means have been implemented and applied in evolutionary contexts (Hunt 2006; Slater et al. 2012a). These models posit a displacement mean that is a product of elapsed evolutionary time and a constant directional trend parameter. While inclusion of a nonzero displacement mean represents a promising first step, a directional trend with a constant rate may not hold over an entire evolutionary history. Trait evolution modeling has been advanced by the development of relaxed models characterized by parameters that vary over evolutionary histories. Examples include Brownian diffusion models with displacement variance rate changes along branches of a phylogeny (O'Meara et al. 2006; Thomas et al. 2006; Lemey et al. 2010; Eastman et al. 2011; Slater et al. 2012b) as well as Ornstein–Uhlenbeck process models that feature similar variation of selection and stochastic variance parameters (Beaulieu et al. 2012) and of optimum trait parameters (Butler and King 2004; Hansen et al. 2008). In this spirit, we extend the Bayesian multivariate Brownian diffusion framework of Lemey et al. (2010) to a flexible model that permits multiple directional trend parameters on a phylogenetic tree in a computationally efficient manner. Importantly, we equip the model with machinery to infer the number of different directional trends supported by the data as well as the locations of trend changes.

Through analyses of simulated data, we demonstrate the ability of our model to accurately recover diffusion parameters, and we show the necessity of appropriate displacement mean modeling in order to do so. We also evaluate model selection techniques to compare the nested Brownian diffusion models featuring no directional trend, a constant trend parameter, and multiple trends. Finally, we apply our methodology to two viral examples of clinical importance. In the first example, we illustrate our approach in a phylogeographic setting by investigating the spatial diffusion of HIV-1 in central Africa. For the second example, we explore antigenic evolution in the context of enhanced resistance of HIV-1 to broadly neutralizing antibodies over the course of the epidemic. We demonstrate a better fit by relaxing the restrictive constant trend assumption, and an improved ability to uncover and quantify key aspects of trait evolution dynamics.

Methods

We start by assuming we have a data set of N aligned molecular sequences $\mathbf{X}=(\mathbf{X}_1,\ldots,\mathbf{X}_N)$ along with N associated values $\mathbf{Y}=(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)$ of an M-dimensional, continuously varying trait. The different coordinates of the "trait" may in fact represent several different phenotypes, but for simplicity, and without loss of generality, we consider it a single multidimensional trait. The sequence and trait data correspond to the N tips of an unknown yet estimable phylogenetic tree τ. Later we will discuss accounting for phylogenetic uncertainty, modeling the molecular evolution process giving rise to $\mathbf{X}$ and integrating it with a model for trait evolution. But first, we explore trait evolution on a fixed phylogeny via a diffusion process acting conditionally independently along its branches.

The N-tipped bifurcating phylogenetic tree τ is a graph with a set of vertices $\mathcal{V}=(\mathcal{V}_1,\ldots,\mathcal{V}_{2N-1})$ and edge weights $\mathcal{T}=(t_1,\ldots,t_{2N-2})$. The vertices correspond to nodes of the tree and, as the length of the tree τ is measured in units of time, $\mathcal{T}$ consists of times corresponding to branch lengths. Each external node (tree tip) $\mathcal{V}_i$ for $i=1,\ldots,N$ is of degree 1, with one parent node $\mathcal{V}_{pa(i)}$ from within the internal or root nodes. Each internal node $\mathcal{V}_i$ for $i=N+1,\ldots,2N-2$ is of degree 3 and the root node $\mathcal{V}_{2N-1}$ is of degree 2. An edge with weight t_i connects $\mathcal{V}_i$ to $\mathcal{V}_{pa(i)}$, and we refer to this edge as branch i. In addition to the observed trait values $\mathbf{Y}_1,\ldots,\mathbf{Y}_N$ at the external nodes, we posit for mathematical convenience unobserved trait values $\mathbf{Y}_{N+1},\ldots,\mathbf{Y}_{2N-1}$ at the internal nodes and root.

Brownian diffusion (also known as a Wiener process or unbiased random walk) is a continuous-time stochastic process originally developed to model the random motion of a physical particle (Brown 1828; Wiener 1958). For a multivariate Brownian diffusion process $\mathbf{W}(t)$, the increment $\mathbf{W}(t_2)-\mathbf{W}(t_1)$ of the process starting at time t_1 and ending at time $t_2\geq t_1$ is multivariate normally distributed with mean $\mathbf{0}$ and variance $(t_2-t_1)\mathbf{I}$, where $\mathbf{I}$ is an $M\times M$ identity matrix. The process is time-homogeneous in that the variance depends only on time differences and not on actual time points. Brownian diffusion is also characterized by independent increments: if $t_1<t_2\leq t_3<t_4$, then the displacements $\mathbf{W}(t_2)-\mathbf{W}(t_1)$ and $\mathbf{W}(t_4)-\mathbf{W}(t_3)$ are independent.

Recent phylogenetic comparative methods (Felsenstein 1988; Revell and Harmon 2008; Vrancken et al. 2015) aim to model the correlated evolution between multiple traits and, to this end, employ a correlated multivariate Brownian diffusion with displacement variance $(t_2-t_1)\mathbf{P}^{-1}$ and displacement mean $\mathbf{0}$. Here, $\mathbf{P}$ is an $M\times M$ infinitesimal precision matrix that determines the intensity and correlation of the trait diffusion after controlling for shared evolutionary history. Recall that in our development the different coordinates of an M-dimensional "trait" may in fact represent different traits, in which case the correlation between traits can be recovered from the appropriate coordinates of $(t_2-t_1)\mathbf{P}^{-1}$. The displacement mean of $\mathbf{0}$ posits that the traits do not evolve according to any systematic directional trend.

The Brownian diffusion process along a phylogeny produces the observed trait values by starting at the root node and proceeding down the branches of τ. The displacement $\mathbf{Y}_i-\mathbf{Y}_{pa(i)}$ along a branch is multivariate normally distributed, centered at $\mathbf{0}$ with variance $t_i\mathbf{P}^{-1}$ proportional to the length of the branch. Therefore, conditioning on the trait value $\mathbf{Y}_{pa(i)}$ at the parent node, we have

$$\mathbf{Y}_i|\mathbf{Y}_{pa(i)}\sim N\left(\mathbf{Y}_{pa(i)},t_i\mathbf{P}^{-1}\right). \tag{1}$$

Throughout this article, we refer to this correlated standard Brownian diffusion model for phylogenetic trait evolution as simply the random walk (RW) model. An extension that introduces branch-specific mixing parameters ν_i into the process that rescale $t_i\mapsto\nu_i t_i$ yields a mixture of Brownian processes and remains popular in phylogeography (Lemey et al. 2010).

Trends

Incorporating a nontrivial displacement mean into the diffusion process is beneficial in several ways. First, we can estimate and quantify directional trends. More importantly, it enables inference of aspects of the evolutionary process that may be poorly approximated or completely unaccounted for by standard Brownian diffusion. More accurate modeling of trait evolution dynamics opens the door to better ancestral trait reconstructions which can, for example, have important implications for elucidating the origin and spread of viral epidemics and ultimately improving disease surveillance and outbreak management (Woolhouse et al. 2015).

To model a constant directional trend, we adopt a multivariate correlated Brownian diffusion process with a nonzero displacement mean. This extension of standard Brownian diffusion is known as a biased random walk, or Brownian diffusion with drift (in the biology literature, "trend" is often used instead of "drift" in order to avoid confusion with genetic drift). We replace the zero mean of the increment $\mathbf{W}(t_2)-\mathbf{W}(t_1)$ with the time-scaled mean vector $(t_2-t_1)\boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is referred to as the trend. The expected difference between the trait values of a descendant and its ancestor is determined by the trend vector $\boldsymbol{\mu}$ and the time elapsed between descendant and ancestor. This yields what we will call the constant directional random walk (CDRW) model:

$$\mathbf{Y}_i|\mathbf{Y}_{pa(i)}\sim N\left(\mathbf{Y}_{pa(i)}+t_i\boldsymbol{\mu},t_i\mathbf{P}^{-1}\right). \tag{2}$$

While this approach is useful for modeling a general directional tendency, it is quite restrictive in that the trend $\boldsymbol{\mu}$ is fixed over the entire phylogeny. We can relax this assumption by introducing branch-specific trend vectors $\boldsymbol{\mu}_i$:

$$\mathbf{Y}_i|\mathbf{Y}_{pa(i)}\sim N\left(\mathbf{Y}_{pa(i)}+t_i\boldsymbol{\mu}_i,t_i\mathbf{P}^{-1}\right) \tag{3}$$

for $i=1,\ldots,2N-2$. In the case of both constant and branch-specific trends, we assign the root trait value a conjugate prior

$$\mathbf{Y}_{2N-1}\sim N\left(\boldsymbol{\mu}^*,(\phi\mathbf{P})^{-1}\right), \tag{4}$$

that is relatively uninformative for small values of ϕ.

It should be noted that trait data must generally be observed at different times in order to estimate a constant directional trend. In the case of an ultrametric phylogenetic tree where the trait data are contemporaneous, a strong prior on the trait value at an internal node, particularly the root node, may still enable inference of a constant trend

(Oakley and Cunningham 2000; Slater et al. 2012a). On the other hand, contemporaneous trait data can contain sufficient information to infer directional trends that vary along the tree (even without a strong prior on an internal node trait value). As a simple example, consider an ultrametric tree with a clade whose tips correspond to relatively large univariate trait values while all tips outside of the clade correspond to small trait values.

Conditioning on the trait value $\mathbf{Y}_{2N-1}$ at the root of τ, the joint distribution of observed trait values $\mathbf{Y}_1,\ldots,\mathbf{Y}_N$ can be expressed as

$$\begin{aligned}&\text{vec}[\mathbf{Y}]\,|\,(\mathbf{Y}_{2N-1},\mathbf{P},\mathbf{V}_\tau,\boldsymbol{\mu}_\tau)\\&\quad\sim N\left(\mathbf{Y}_{\text{root}}+(\mathbf{T}\otimes\mathbf{I}_M)\boldsymbol{\mu}_\tau,\mathbf{P}^{-1}\otimes\mathbf{V}_\tau\right),\end{aligned}\tag{5}$$

building on a similar construction for standard Brownian diffusion (Felsenstein 1973; Freckleton et al. 2002). Here, vec[$\mathbf{Y}$] is the vectorization of the column vectors $\mathbf{Y}_1,\ldots,\mathbf{Y}_N$, while $\mathbf{I}_M$ is an $M\times M$ identity matrix, and $\otimes$ is the Kronecker product. $\mathbf{Y}_{\text{root}}$ is the $NM\times 1$ vector $(\mathbf{Y}^t_{2N-1},\ldots,\mathbf{Y}^t_{2N-1})^t$, and $\boldsymbol{\mu}_\tau$ is the $(2N-2)M\times 1$ trend vector $(\boldsymbol{\mu}^t_1,\ldots,\boldsymbol{\mu}^t_{2N-2})^t$. The $N\times N$ variance matrix $\mathbf{V}_\tau$ is a deterministic function of τ and represents the contribution of the phylogenetic tree to the covariance structure. Its diagonal entries V_{ii} are equal to the distance in time between the tip $\mathcal{V}_i$ and the root node $\mathcal{V}_{2N-1}$, and off-diagonal entries V_{ij} correspond to the distance in time between the root node $\mathcal{V}_{2N-1}$ and the most recent common ancestor of tips $\mathcal{V}_i$ and $\mathcal{V}_j$. Finally, the $N\times(2N-2)$ matrix $\mathbf{T}$ is defined as follows: $T_{ij}=t_j$, the length of branch j, if branch j is part of the path from the external node i to the root, and $T_{ij}=0$ otherwise.

Our development thus far clarifies some important issues. First, while it is tempting to model a unique trend on each branch, not all $\boldsymbol{\mu}_i$ are uniquely identifiable in the likelihood in Equation (5). Care must be taken to impose necessary restrictions to ensure identifiability while still permitting sufficient trend variation, and we discuss an approach to achieve this later in the article. Second, the variance matrix $\mathbf{P}^{-1}\otimes\mathbf{V}_\tau$ in Equation (5) suggests a computational order of $\mathcal{O}(N^3M^3)$ to evaluate the density. Repeated evaluation of (5) is necessary for numerical integration in Bayesian modeling, and viral data sets may encompass thousands of sequences. Fortunately, Pybus et al. (2012) demonstrate that phylogenetic standard Brownian diffusion likelihoods can be evaluated in computational order $\mathcal{O}(NM^2)$ by modeling in terms of the precision matrix $\mathbf{P}$ (as opposed to the variance) and adopting a dynamic programming approach. In the next section, we present an adaptation of their algorithm for our more general diffusion likelihood.

Multivariate Trait Peeling

To compute the distribution of the observed trait values, we extend a dynamic programming approach that is analogous to Felsenstein's pruning method (Felsenstein 1981) and has been employed for standard Brownian diffusion likelihoods (Pybus et al. 2012; Vrancken et al. 2015). The algorithm begins with the joint density of observed trait values at the tips of τ and unobserved trait values at each of the tree's root and internal nodes. The marginalized density of observed trait values can be obtained by integrating over all possible realizations of the unobserved trait values. This high-dimensional integration is accomplished through a tree traversal over the root and internal nodes. When a node is visited, we integrate out the hypothesized trait value at that node and arrive at a partial density that is used later in the traversal to integrate out the unobserved trait value at another node. The partial density yielded at a given nodal visit is the density of the trait values at the tips descendent to that node conditional on the unobserved trait value of the parent of the node. Each integration amounts to integrating the product of two such partial densities and the density of an unobserved trait value conditional on the trait value of its parent, and the structure of these three densities enables us to perform the integration by computing partial mean vectors, precision scalars, and remainder terms.

Under our Brownian diffusion process with trends, the joint distribution of all trait values is straightforwardly expressed as the product

$$\begin{aligned}P(\mathbf{Y}_1,\ldots,\mathbf{Y}_{2N-1}|\tau,\mathbf{P},\boldsymbol{\mu},\phi)=&\left(\prod_{i=1}^{2N-2}P(\mathbf{Y}_i|\mathbf{Y}_{pa(i)},\mathbf{P},t_i,\boldsymbol{\mu}_i)\right)\\&\times P(\mathbf{Y}_{2N-1}|\mathbf{P},\boldsymbol{\mu}^*,\phi),\end{aligned}\tag{6}$$

where $\boldsymbol{\mu}=(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_{2N-2},\boldsymbol{\mu}^*)$. We wish to compute the density

$$P(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)=\int\cdots\int P(\mathbf{Y}_1,\ldots,\mathbf{Y}_{2N-1})d\mathbf{Y}_{N+1}\ldots d\mathbf{Y}_{2N-1}\tag{7}$$

$$\begin{aligned}=&\int\cdots\int\left(\prod_{i=1}^{2N-2}P(\mathbf{Y}_i|\mathbf{Y}_{pa(i)})\right)\\&\times P(\mathbf{Y}_{2N-1})d\mathbf{Y}_{N+1}\ldots d\mathbf{Y}_{2N-1}.\end{aligned}\tag{8}$$

Here, we omit dependence on the parameters $\tau,\mathbf{P},\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_{2N-2},\boldsymbol{\mu}*$ and ϕ from the notation for the sake of clarity. The integration proceeds in a postorder traversal integrating out one internal node trait value at a time. Let $\{\mathbf{Y}_i\}$ denote the set of observed trait values descendant from and including the node V_i, and suppose $pa(i)=pa(j)=k$. Our traversal requires computing integrals of the form

$$P(\{\mathbf{Y}_k\}|\mathbf{Y}_{pa(k)})=\int P(\{\mathbf{Y}_i\}|\mathbf{Y}_k)P(\{\mathbf{Y}_j\}|\mathbf{Y}_k)P(\mathbf{Y}_k|\mathbf{Y}_{pa(k)})d\mathbf{Y}_k.\tag{9}$$

Because the integrand is proportional to a multivariate normal density, it suffices to keep track of partial mean vectors $\mathbf{m}_k$, partial precision scalars p_k, and normalizing constants ρ_k.

Let MVN(.;$\boldsymbol{\kappa}$,$\boldsymbol{\Lambda}$) denote a multivariate normal probability density function with mean $\boldsymbol{\kappa}$ and precision $\boldsymbol{\Lambda}$. We can rewrite conditional densities to facilitate integration with respect to the trait at the parent node:

$$P(\mathbf{Y}_i|\mathbf{Y}_k) = \text{MVN}\left(\mathbf{Y}_i; t_i\boldsymbol{\mu}_i + \mathbf{Y}_k, \frac{1}{t_i}\mathbf{P}\right) \tag{10}$$

$$= \text{MVN}\left(\mathbf{Y}_i - t_i\boldsymbol{\mu}_i; \mathbf{Y}_k, \frac{1}{t_i}\mathbf{P}\right). \tag{11}$$

For $i=1,\ldots,N$, set normalizing constant $\rho_i=1$, partial mean

$$\mathbf{m}_i = \mathbf{Y}_i - t_i\boldsymbol{\mu}_i \tag{12}$$

and partial precision

$$p_i = \frac{1}{t_i}. \tag{13}$$

Then,

$$P(\{\mathbf{Y}_i\}|\mathbf{Y}_k)P(\{\mathbf{Y}_j\}|\mathbf{Y}_k) = \rho_i\rho_j \times \text{MVN}(\mathbf{m}_i; \mathbf{Y}_k, p_i\mathbf{P}) \tag{14}$$

$$\times \text{MVN}(\mathbf{m}_j; \mathbf{Y}_k, p_j\mathbf{P})$$

$$= \rho_k \times \text{MVN}(\mathbf{Y}_k; \mathbf{m}_k^*, (p_i+p_j)\mathbf{P}) \tag{15}$$

where partial unshifted mean

$$\mathbf{m}_k^* = \frac{p_i\mathbf{m}_i + p_j\mathbf{m}_j}{p_i+p_j}, \tag{16}$$

and normalizing constant

$$\rho_k = \rho_i\rho_j\left(\frac{p_ip_j}{2\pi(p_i+p_j)}\right)^{d/2} |\mathbf{P}|^{1/2}\frac{\exp\left[-\frac{p_i}{2}\mathbf{m}_i'\mathbf{P}\mathbf{m}_i - \frac{p_j}{2}\mathbf{m}_j'\mathbf{P}\mathbf{m}_j\right]}{\exp\left[-\frac{p_i+p_j}{2}\mathbf{m}_k^{*\prime}\mathbf{P}\mathbf{m}_k^*\right]}. \tag{17}$$

Multiplying by $P(\mathbf{Y}_k|\mathbf{Y}_{pa(k)})$ and integrating with respect to $\mathbf{Y}_k$, we get

$$P(\{\mathbf{Y}_k\}|\mathbf{Y}_{pa(k)}) = \int P(\{\mathbf{Y}_i\}|\mathbf{Y}_k)P(\{\mathbf{Y}_j\}|\mathbf{Y}_k)P(\mathbf{Y}_k|\mathbf{Y}_{pa(k)})d\mathbf{Y}_k \tag{18}$$

$$= \rho_k \times \text{MVN}(\mathbf{Y}_{pa(k)}; \mathbf{m}_k, p_k\mathbf{P}), \tag{19}$$

where

$$\mathbf{m}_k = \mathbf{m}_k^* - t_k\boldsymbol{\mu}_k, \tag{20}$$

and

$$p_k = \frac{1}{t_k + \frac{1}{p_i+p_j}}. \tag{21}$$

Integrating out all internal node trait values yields

$$P(\mathbf{Y}_1,\ldots,\mathbf{Y}_N|\mathbf{Y}_{2N-1}) = \rho_{2N-1} \times \text{MVN}(\mathbf{Y}_{2N-1}; \mathbf{m}_{2N-1}^*, (p_{2N-2}+p_{2N-3})\mathbf{P}). \tag{22}$$

For the final step, we multiply by the conjugate root prior and integrate:

$$P(\mathbf{Y}_1,\ldots,\mathbf{Y}_N) = \int P(\mathbf{Y}_1,\ldots,\mathbf{Y}_N|\mathbf{Y}_{2N-1})P(\mathbf{Y}_{2N-1})d\mathbf{Y}_{2N-1} \tag{23}$$

$$= \rho_{2N-1}\text{MVN}(\mathbf{m}_{2N-1}^*; \boldsymbol{\mu}^*, p_{2N-1}\mathbf{P}), \tag{24}$$

where

$$p_{2N-1} = \frac{(p_{2N-2}+p_{2N-3})\phi}{p_{2N-2}+p_{2N-3}+\phi}. \tag{25}$$

In practice, the algorithm visits each node in the phylogeny once and computes partial unshifted means $\mathbf{m}_k^*$, partial means $\mathbf{m}_k$, partial precisions p_k, and normalizing constants ρ_k.

Identifiability and the Relaxed Directional Random Walk Model

Ideally, we would like to model a unique trend $\boldsymbol{\mu}_i$ on each branch i of the phylogenetic tree. However, such lax assumptions open the door to misleading inferences. Adopting the notation of the *Trends* section, there can exist distinct trend vectors $\boldsymbol{\mu}_\tau \neq \boldsymbol{\mu}_\tau^*$ such that

$$(\mathbf{T}\otimes\mathbf{I}_M)\boldsymbol{\mu}_\tau = (\mathbf{T}\otimes\mathbf{I}_M)\boldsymbol{\mu}_\tau^*, \tag{26}$$

yielding identical trait likelihoods in Equation (5). The lack of model identifiability presents an obstacle to uncovering the "true" values of the trends that characterize the trait evolution process.

We propose a relaxed directional random walk (RDRW) model that allows for trend variation along a phylogenetic tree while maintaining model identifiability. This is achieved by having branches inherit trends from ancestral branches by default, but allowing a random number of specific types of trend changes to occur along the tree. We describe the model here and refer readers to the Appendix for a detailed argument establishing identifiability.

We begin at the unobserved branch leading to the root, or most recent common ancestor (MRCA), of the phylogenetic tree τ and associate with it the trend $\boldsymbol{\mu}_{\text{MRCA}}$. Then the two branches emanating from the root node either both inherit the trend $\boldsymbol{\mu}_{\text{MRCA}}$, or a trend change occurs and one branch receives a new trend while the other branch assumes the trend $\boldsymbol{\mu}_{\text{MRCA}}$. Similarly, whenever a branch splits into two anywhere in τ, either both child branches assume the same trend as the parent branch, or one child branch takes on a new value while the other inherits its trend from the parent branch. Both child branches taking on different trends than the parent branch is not permitted.

Importantly, rather than fix the type of trend transfer that occurs at a given node, we estimate it from the data. The benefits of this choice are twofold. First, a trend change is not forced when the data do not suggest a need for one. Unnecessarily imposing a large number of unique trends to be inferred from limited data can

lead to high variance estimates. Second, in the event of a trend change occurring at a node, only one of the two child branches can assume a new trend. We let the data determine which of the child branches assumes the new trend. The data may support new trends on both child branches. While our model may appear too restrictive to accommodate such a scenario at first glance, we are able to infer the relative support for each child, and it is reflected in the posterior distribution in terms of the probabilities of the two types of changes. Thus, summaries of the posterior distribution can capture the true nature of trend variation in spite of the identifiability restrictions.

It is important to handle the initial trend $\boldsymbol{\mu}_{\text{MRCA}}$ with care. One option is to estimate $\boldsymbol{\mu}_{\text{MRCA}}$ from the data just as with all other trends. However, such a choice may not be ideal for data sets that exhibit relatively long periods of divergence from the MRCA to the sampling times. There is generally more information about diffusion dynamics during time periods overlapping with or close to sampling times. Likewise, the further removed the MRCA is from sampling times, the less information there is about $\boldsymbol{\mu}_{\text{MRCA}}$ and other trends near the MRCA. Because of the interconnectedness of trends associated with different branches in the RDRW model, estimates of $\boldsymbol{\mu}_{\text{MRCA}}$ and neighboring trends under such circumstances may primarily reflect information about trends on branches near sampling times. To mitigate misleading inferences of trends near the MRCA, we can adopt an initial trend $\boldsymbol{\mu}_{\text{MRCA}}=0$ and still interpret changes in trends across the tree.

To parameterize the model, we associate a ternary variable δ_k with each internal node k specifying how it passes on its trend to its child nodes. Suppose node k has left child node i and right child node j. If $\delta_k=-1$, then $\boldsymbol{\mu}_i=\boldsymbol{\mu}_k$ and node j assumes a new rate $\boldsymbol{\mu}_j=\boldsymbol{\mu}_k+\boldsymbol{\alpha}_j$. If $\delta_k=1$, then node i assumes a new rate $\boldsymbol{\mu}_i=\boldsymbol{\mu}_k+\boldsymbol{\alpha}_i$ while $\boldsymbol{\mu}_j=\boldsymbol{\mu}_k$. If $\delta_k=0$, then no trend changes occur and $\boldsymbol{\mu}_k=\boldsymbol{\mu}_i=\boldsymbol{\mu}_j$. To map the ternary δ_k variables to binary indicators $\boldsymbol{\gamma}_i$ of trend changes for child branches, we define

$$\boldsymbol{\gamma}_i=\frac{1+\delta_k}{2}|\delta_k|, \tag{27}$$

and

$$\boldsymbol{\gamma}_j=\frac{1-\delta_k}{2}|\delta_k|. \tag{28}$$

Thus,

$$\boldsymbol{\mu}_i=\boldsymbol{\mu}_{pa(i)}+\boldsymbol{\gamma}_i\boldsymbol{\alpha}_i, \tag{29}$$

for $i=1,\ldots,2N-2$. Working with the binary $\boldsymbol{\gamma}_i$ eases our understanding of the MCMC procedure to infer trend changes, discussed below. However, we parameterize the model in terms of the ternary δ_k to facilitate enforcement of the model restrictions.

Of particular interest is the random number $K\in 0,\ldots,N-1$ of trend changes that occur in τ. We can write K in terms of the δ_i,

$$K=\sum_{i=N+1}^{2N-1}|\delta_i|, \tag{30}$$

and it provides us with a natural way to think of the vector $\boldsymbol{\delta}=(\delta_{N+1},\ldots,\delta_{2N-1})$. For example, we can express our prior beliefs about $\boldsymbol{\delta}$ in terms of K. A popular prior for count data is the Poisson distribution

$$K\sim\text{Poisson}(\lambda). \tag{31}$$

Here, λ is the expected number of trend changes in τ. In our analyses, we set $\lambda=\log(2)$, which places 50% prior probability on the hypothesis of no trend changes.

In order to infer the nature of the trend transitions that occur at the nodes of the phylogenetic tree, we borrow ideas from Bayesian stochastic search variable selection (BSSVS) (George and McCulloch 1993; Kuo and Mallick 1998; Chipman et al. 2001). BSSVS is typically applied to model selection problems in a linear regression setting. In this framework, we begin with a large number P of potential predictors $\mathbf{X}_1,\ldots,\mathbf{X}_P$ and seek to determine which of them associate linearly with an N-dimensional outcome $\mathbf{Y}$. The full model with all predictors is

$$\mathbf{Y}=\mathbf{X}_1\boldsymbol{\beta}_1+\cdots+\mathbf{X}_P\boldsymbol{\beta}_P+\boldsymbol{\epsilon}, \tag{32}$$

where the $\boldsymbol{\beta}_i$ are regression coefficients and $\boldsymbol{\epsilon}$ is a vector of normally distributed error terms with mean $\mathbf{0}$. When a particular $\boldsymbol{\beta}_i$ is determined to differ significantly from 0, the corresponding $\mathbf{X}_i$ helps predict $\mathbf{Y}$. If not, $\mathbf{X}_i$ contributes little additional information and is fit to be removed from the model by forcing $\boldsymbol{\beta}_i=0$. Predictors may be highly correlated, and deterministic model selection strategies tend not to find the optimal set of predictors without exploring all possible subsets. There exists 2^P such subsets, so exploring all of them is computationally unfeasible in general and fails completely for $P>N$.

BSSVS efficiently explores the possible subsets of model predictors by augmenting the model state–space with a vector $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_P)$ of binary indicator variables that dictate which predictors to include. The indicators δ_i impose a prior on the regression coefficients $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_P)$ with mean $\mathbf{0}$ and variance proportional to a $P\times P$ diagonal matrix with its diagonal equal to $\boldsymbol{\delta}$. If $\delta_i=0$, then the prior variance on $\boldsymbol{\beta}_i$ shrinks to 0 and forces $\boldsymbol{\beta}_i=0$ in the posterior. The joint space $(\boldsymbol{\beta},\boldsymbol{\delta})$ is explored simultaneously through MCMC.

We apply BSSVS in our RDRW setting to determine the types of trend transfers that occur. We achieve this by exploring the joint space $(\boldsymbol{\alpha},\boldsymbol{\delta})$ of trend differences between parent and child branches, and ternary trend change indicators. The δ_k map to binary indicators $\boldsymbol{\gamma}_i$, as shown in Equations (27) and (28). We assume that trend differences $\boldsymbol{\alpha}_i=\boldsymbol{\mu}_i-\boldsymbol{\mu}_{pa(i)}$ are *a priori* independent and normally distributed,

$$\boldsymbol{\alpha}_i\sim N(\mathbf{0},\boldsymbol{\gamma}_i\sigma^2\mathbf{I}). \tag{33}$$

If $\gamma_i = 0$, then the prior variance on the components of $\boldsymbol{\alpha}_i$ shrinks to 0. This forces $\boldsymbol{\alpha}_i = \mathbf{0}$, and hence $\boldsymbol{\mu}_i = \boldsymbol{\mu}_{pa(i)}$, in the posterior.

We complete our directional random walk model specification by assigning the precision matrix $\mathbf{P}$ a Wishart prior with degrees of freedom v and scale matrix $\mathbf{V}$. Importantly, the Wishart distribution is conjugate to the observed trait likelihood. Indeed, invoking the notation for partial means and precisions from the trait peeling algorithm, the posterior

$$P(\mathbf{P}|\mathbf{Y}_1,\ldots,\mathbf{Y}_N) \propto P(\mathbf{Y}_1,\ldots,\mathbf{Y}_N|\mathbf{P})P(\mathbf{P}) \tag{34}$$

has a Wishart distribution with $N+v$ degrees of freedom and scale matrix

$$\left(\mathbf{V}^{-1} + p_{2N-1}(\mathbf{m}^*_{2N-1} - \boldsymbol{\mu}^*)(\mathbf{m}^*_{2N-1} - \boldsymbol{\mu}^*)' + \sum_{k=N+1}^{2N-1} [p_i \mathbf{m}_i \mathbf{m}_i' + p_j \mathbf{m}_j \mathbf{m}_j' - (p_i + p_j)\mathbf{m}^*_k \mathbf{m}^{*'}_k]\right)^{-1}. \tag{35}$$

Lemey et al. (2010) exploit a similar conjugacy to construct an efficient Gibbs sampler for $\mathbf{P}$, and our adoption of the Wishart prior conveniently allows us to extend use of the sampler to our model that now includes estimable displacement means.

Joint Modeling and Inference

A major strength of our Bayesian framework is that it jointly models sequence and trait evolution. It is possible to first reconstruct a phylogenetic tree from molecular sequence data and then condition trait evolutionary inferences on the inferred tree. However, such an approach ignores the error associated with phylogenetic uncertainty. Adoption of a tree inferred exclusively from sequence data also fails to take advantage of the additional phylogenetic information the trait data may provide, even though the sequence information will generally dominate over the trait information (Baele et al. 2016b). The extent to which traits can contribute to adequately informing the tree will depend on their phylogenetic signal. We note that such signal can be quantified using various methods including one that is implemented in the framework we focus on here (Vrancken et al. 2015). Fortunately, Bayesian inference provides the machinery to overcome these difficulties by jointly estimating the phylogeny and trait and molecular sequence evolution parameters from both trait and sequence data. Notably, Bayesian methods enable improved ancestral state reconstruction by accounting for phylogenetic and evolutionary parameter uncertainty (Ronquist 2004).

Adopting a standard phylogenetic approach, we assume the sequence data $\mathbf{X}$ arise from a CTMC model for character evolution acting along the unobserved phylogenetic tree τ. The CTMC is characterized by a vector $\mathbf{Q}$ of mutation parameters that may include, for instance, relative exchange rates among characters, an overall rate multiplier and across-site variation specifications. The trait data $\mathbf{Y}$ arise from a directional random walk diffusion process acting on τ, governed by parameters $\boldsymbol{\Lambda}$. A crucial assumption is that the processes giving rise to the observed sequences and trait values are conditionally independent given the phylogenetic tree τ:

$$P(\mathbf{X},\mathbf{Y}|\tau,\mathbf{Q},\boldsymbol{\Lambda}) = P(\mathbf{X}|\tau,\mathbf{Q})P(\mathbf{Y}|\tau,\boldsymbol{\Lambda}), \tag{36}$$

enabling us to write the joint model posterior distribution as

$$P(\tau,\mathbf{Q},\boldsymbol{\Lambda}|\mathbf{X},\mathbf{Y}) \propto P(\mathbf{X}|\tau,\mathbf{Q})P(\mathbf{Y}|\tau,\boldsymbol{\Lambda})P(\tau)P(\mathbf{Q})P(\boldsymbol{\Lambda}). \tag{37}$$

We implement the joint model by integrating our directional random walk framework for trait evolution into the open source Bayesian Evolutionary Analysis Sampling Trees (BEAST) software package (Drummond et al. 2012), available at `https://github.com/beast-dev/beast-mcmc/`. BEAST provides an array of efficient methods for Bayesian phylogenetic inference, particularly to estimate phylogenies and model molecular sequence evolution. For the phylogeny τ, we choose from flexible coalescent-based priors that do not make strong *a priori* assumptions about the population history (Gill et al. 2013, 2016). For sequence evolution, we have access to a range of classic substitution models (Kimura 1980; Felsenstein 1981; Hasegawa et al. 1985), gamma-distributed rate heterogeneity among sites (Yang 1994), and strict and relaxed molecular clock models for branch rates (Drummond et al. 2006).

Estimation of the full joint posterior in Equation (37) is achieved through MCMC sampling (Metropolis et al. 1953; Hastings 1970). We employ standard Metropolis–Hastings transition kernels available in BEAST to integrate over the parameter spaces of $\mathbf{Q}$ and τ. To sample realizations of the trait evolution precision matrix $\mathbf{P}$, we adapt a Gibbs sampler developed for standard Brownian diffusion (Lemey et al. 2010). For the RDRW model, we need transition kernels to explore the space $(\boldsymbol{\alpha}, \boldsymbol{\delta})$ of branch-specific trend differences and ternary trend change indicators. We propose new trend differences $\boldsymbol{\alpha}^*_i$ component-wise through a random walk transition kernel that adds random values within a specified window size to the current $\boldsymbol{\alpha}_i$.

For $\boldsymbol{\delta}$, we implement a trit–flip transition kernel that chooses one of the $N-1$ ternary indicators $\boldsymbol{\delta}_k$ uniformly at random and proposes a new state $\boldsymbol{\delta}^*_k$ assuming one of the two possible values not equal to $\boldsymbol{\delta}_k$ with equal probability. For example, if $\boldsymbol{\delta}_k = 0$, then

$$\boldsymbol{\delta}^*_k = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2}. \end{cases} \tag{38}$$

We have parameterized our prior on $\boldsymbol{\delta}$ in terms of the number K of trend changes, and this parameterization should be retained for the transition kernel in order to ensure the correct Metropolis–Hastings proposal ratio (Drummond and Suchard 2010). A proposed increase in trend changes occurs when we choose a $\boldsymbol{\delta}_k$ with value 0,

so

$$q(K^*=K+1|K)=\frac{N-1-K}{N-1}. \quad (39)$$

If we choose a δ_k with a nonzero value, we propose the other nonzero value with probability 0.5 (corresponding to $K^*=K$), and we propose 0 with probability 0.5 (which means a decrease in rate changes, $K^*=K-1$). Therefore,

$$q(K^*=K|K)=q(K^*=K-1|K)=\frac{1}{2}\frac{K}{N-1}. \quad (40)$$

These calculations yield the following proposal ratio for K:

$$\frac{q(K|K^*)}{q(K^*|K)}=\begin{cases}\frac{1}{2}\frac{K+1}{N-1-K} & \text{if } K^*=K+1\\ 1 & \text{if } K^*=K\\ \frac{2(N-K)}{K} & \text{if } K^*=K-1.\end{cases} \quad (41)$$

In addition to the parameters characterizing the trait and sequence evolution processes, we may wish to make inferences about the posterior distribution of trait values at the root and internal nodes, or at any arbitrary time point in the past. We equip BEAST with the ability to generate posterior trait value realizations at these nodes by implementing a preorder, tree-traversal algorithm.

A popular choice to summarize the results of a Bayesian phylogenetic analysis is a maximum clade credibility (MCC) tree. To form an MCC tree, the posterior sample of trees is examined to determine posterior clade probabilities, and the tree with the maximum product of posterior clade probabilities is the MCC tree. The branches and nodes in MCC tree clades can be annotated with inferred drift rates and trait values, along with other quantities of interest, from matching clades in the posterior sample. Annotated MCC trees can be summarized using TreeAnnotator, available as part of the BEAST distribution, and visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Model Selection

We can formally compare the CDRW and RDRW models through Bayes factors. A Bayes factor (Jeffreys 1935, 1961) compares the fit of two models, M_1 and M_0, to observed data $(\mathbf{X},\mathbf{Y})$ by taking the ratio of marginal likelihoods:

$$BF_{10}=\frac{P(\mathbf{X},\mathbf{Y}|M_1)}{P(\mathbf{X},\mathbf{Y}|M_0)}=\frac{P(M_1|\mathbf{X},\mathbf{Y})}{P(M_0|\mathbf{X},\mathbf{Y})}\Big/\frac{P(M_1)}{P(M_0)}. \quad (42)$$

BF_{10} quantifies the evidence in favor of model M_1 over M_0. Kass and Raftery (1995) provide guidelines for assessing the strength of the evidence against M_0: Bayes factors between 1 and 3 are not worth more than a bare mention, while values between 3 and 20 are considered positive evidence against M_0. Bayes factors in the ranges 20–150 and >150 are considered to be strong and a very strong evidence against M_0, respectively.

Evaluation of Bayes factors has become a popular approach to model selection in Bayesian phylogenetics (Sinsheimer et al. 1996; Suchard et al. 2005). Marginal likelihood estimation can be quite difficult in a phylogenetic context, and (generalized) stepping-stone sampling estimators have been implemented to address this (Baele et al. 2016a). Following the approach of Drummond and Suchard (2010), however, we are able to straightforwardly compute the Bayes factor BF_C supporting the CDRW model M_C over the RDRW model M_R. The model M_C is nested within the more general model M_R and occurs when the number of trend changes $K=0$. This enables us to write

$$BF_C=\frac{P(\mathbf{X},\mathbf{Y}|M_C)}{P(\mathbf{X},\mathbf{Y}|M_R)}=\frac{P(M_C|\mathbf{X},\mathbf{Y})}{P(M_R|\mathbf{X},\mathbf{Y})}\Big/\frac{P(M_C)}{P(M_R)} \quad (43)$$

$$=\frac{P(K=0|\mathbf{X},\mathbf{Y},M_R)}{1-P(K=0|\mathbf{X},\mathbf{Y},M_R)}\Big/\frac{P(K=0|M_R)}{1-P(K=0|M_R)}, \quad (44)$$

requiring only our prior probability of no trend changes under the RDRW model, and the posterior probability of zero trend changes.

Model Performance

To evaluate the performance of the CDRW and RDRW models, we simulate univariate trait values on the tips of the MCC tree estimated from a data set of HIV-1 sequences and traits that measure neutralization by several broadly neutralizing antibodies (Euler et al. 2011). (We analyze this data set in detail with our trait evolution framework later in the article.) Trait values are simulated on the tree tips according to a directional random walk process with one or more trends. We consider three different simulation scenarios: one with a constant trend, and two with multiple trends. We set the diffusion displacement variance rates to 0.13 and 0.06 for simulated data sets with a constant trend and with multiple trends, respectively. These are the estimated posterior mean displacement variance rates for neutralization by antibody VRC01 from the Euler et al. (2011) data under the CDRW and RDRW models, respectively. For each simulation scenario, we generate 100 simulated data sets and analyze the data sets with the model used to generate the data (the CDRW model or RDRW model) as well as with the model nested immediately within it (the RW model or CDRW model, respectively). We assess the performance of our models by monitoring frequentist coverage of true parameter values by 95% Bayesian credibility intervals (BCIs). Ideally, estimated coverage should approach the nominal coverage of 95%. Also, to compare displacement variance rate estimates yielded by the different models, we compute the percentage error, which we define as

$$\text{Percentage error}=100\times\frac{|\text{Estimated posterior mean rate - True rate}|}{|\text{True rate}|}. \quad (45)$$

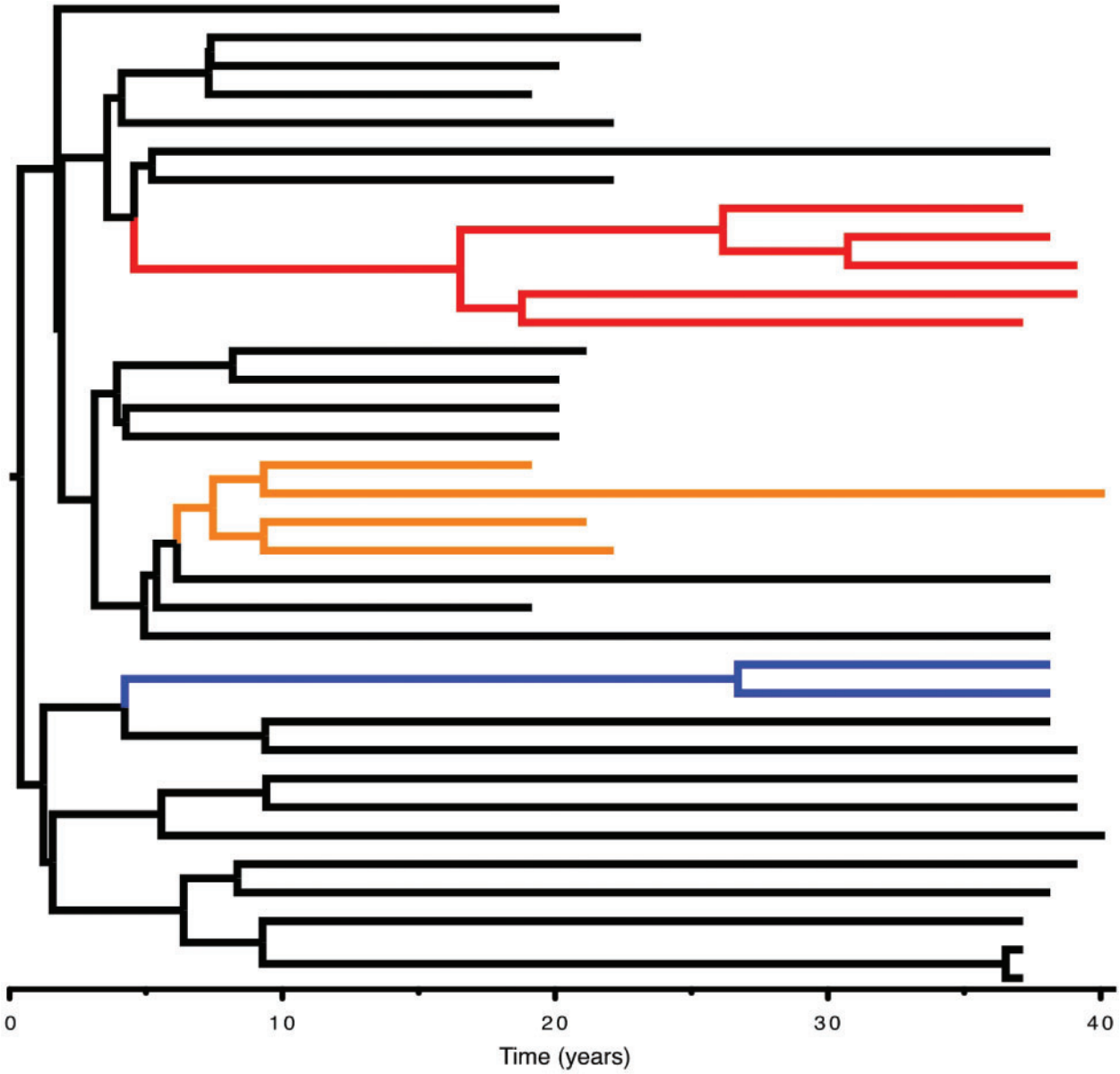

FIGURE 1. Phylogenetic tree with multiple simulated trends. Branches are colored according to different trend values. Simulation scenario 1 features a trend of 0.15 on all branches. In simulation scenario 2, red, orange, blue, and black branches have trends of 0.3, −0.5, 0.075, and 0, respectively. Simulation scenario 3 is identical to scenario 2, except that blue branches have trends of 0.35. Trends are reported in units of trait value units per year.

Finally, we evaluate our model selection procedure by recording the proportional of simulants for which the Bayes factor supports the true model in favor of the nested model according to different thresholds.

The first simulation scenario features a constant trend of 0.15. For the second and third simulation scenarios, we divide the branches of the phylogenetic tree into four different groups (Fig. 1). In the second simulation scenario, we assign red, orange, blue and black branches trends of 0.3, −0.5, 0.075, and 0, respectively. The third simulation scenario is exactly like the second, except we replace the relatively modest trend of 0.075 on blue branches with 0.35. In this synthetic example, the trait is abstract and is not measured in concrete units. However, trends are generally reported in units of whatever the trait value units are, per year.

The results (Table 1) demonstrate that the directional random walk models do adequately well at recovering the true model and its parameters when applied to data simulated according to the same model. Bayes factors express support in favor of true models over the simpler nested models at high rates. Furthermore, coverage of true trends and displacement variance rates is generally close to nominal for the CDRW model in scenario 1 and for the RDRW model in scenarios 2 and 3. The notable exception is the poor coverage of the trends on

TABLE 1. Performance of trait evolution models on data simulated under three different simulation scenarios. We report coverage of true trends and displacement variance rates and, for simulation scenarios with multiple trends, the posterior mean number of trend changes. We also report Bayes factor (BF) support for the true model.

	Scenario 1	Scenario 2	Scenario 3
Constant Trend Coverage (CDRW)	92	—	—
Red Trend Coverage (RDRW)	—	94	93
Orange Trend Coverage (RDRW)	—	94	92
Blue Trend Coverage (RDRW)	—	50	93
Black Trend Coverage (RDRW)	—	94	95
Displacement Variance Rate Coverage (RW)	78	—	—
Displacement Variance Rate Coverage (CDRW)	93	0	0
Displacement Variance Rate Coverage (RDRW)	—	93	93
Percentage Error for D.V. Rate Estimate (RW)	39	—	—
Percentage Error for D.V. Rate Esimate (CDRW)	22	54	65
Percentage Error for D.V. Rate Esimate (RDRW)	—	26	22
Percentage of BF > 3	93	100	100
Percentage of BF > 20	62	100	100
Percentage of BF > 150	30	100	100
Mean Trend Changes (RDRW)	—	2.21	3.11

blue branches in scenario 2. In the second simulation scenario, the mean number of trend changes, averaged over 100 simulated data sets, is 2.21. Although the RDRW model detects the two trend changes that correspond to the red and orange lineages, it does not identify the trend change that corresponds to the blue lineage. This may be due to the relatively small deviation of the 0.075 blue lineage trend from the trivial trend of 0 along black branches. Indeed, when the blue branches are assigned a greater trend in scenario 3, we infer a mean number of 3.11 trend changes under the RDRW model. Finally, in each of the three simulation scenarios, the models under which the data were generated greatly outperform the more restrictive nested models in estimation of displacement variance rates. This demonstrates that a failure to adequately model trends in the diffusion process can lead to inaccurate displacement variance rate estimates. We discuss this issue in more detail in the HIV-1 phylogeography example.

THE SPREAD OF HIV-1 IN CENTRAL AFRICA

Faria et al. (2014) explore the early spatial expansion and epidemic dynamics of HIV-1 in central Africa by analyzing sequence data sampled from countries in the Congo River basin. The authors employ a discrete phylogeographic inference framework (Lemey et al. 2009a) and show that the pandemic likely originated in Kinshasa (in what is now the Democratic Republic of Congo) in the 1920s. Furthermore, viral spread to other population centers in sub-Saharan Africa was aided by a combination of factors, including strong railway networks, urban growth, and changes in sexual behavior.

We follow up the analysis of Faria et al. (2014) by applying our continuous directional random walk approach to one of the data sets analyzed in their study. The data set consists of HIV-1 sequences sampled between 1985 and 2004 from the Democratic Republic of Congo and the Republic of Congo and includes 96 sequences from Kinshasa (Vidal et al. 2000, 2005; Kalish et al. 2004; Yang et al. 2005), 96 sequences from Mbuji-Mayi (Vidal et al. 2000, 2005), 96 from Brazzaville (Bikandou et al. 2004; Niama et al. 2006), 76 from Lubumbashi (Vidal et al. 2005), 33 from Bwamanda (Vidal et al. 2000), 24 from Likasi (Kita et al. 2004), 23 from Kisangani (Vidal et al. 2005), and 22 sequences from Pointe-Noire (Bikandou et al. 2000). Sampling bias is a common obstacle in phylogeographic analyses, and to mitigate sampling effects, Faria et al. (2014) down-sample from the most densely sampled location (Kinshasa, for which 422 sequences were available). We reconstruct the spatial dynamics under standard correlated Brownian (RW), CDRW, and RDRW diffusion models on a MCC tree estimated from the sequences and their locations of sampling. The traits in this instance are bivariate longitude and latitude coordinates, with observed trait values corresponding to sampling locations. Table 2 reports posterior estimates of trends.

Under the CDRW model, we infer a significant longitudinal trend with posterior mean 0.30 degrees per year and a 95% BCI (0.26, 0.33), as well as a significant latitudinal trend with posterior mean of −0.09 degrees per year and BCI (−0.11, −0.06). Furthermore, for each coordinate the Bayes factor in favor of a CDRW model over the RW model is greater than 1000, indicating substantially a better fit for the CDRW model. These results imply general eastward and southward trends in the spread of HIV from the Kinshasa–Brazzaville–Pointe-Noire area to other population centers. They also reflect the composition of sampling locations: in terms of longitude, a majority are far to the east of the believed origin while the rest are relatively close to it. Similarly, nearly 90% of the sequences come from locations south of Kinshasa or from neighboring locations of similar latitude. On the other hand, the existence of samples from cities north of Kinshasa,

TABLE 2. Spatiotemporal dynamics of HIV-1 in central Africa. Model comparison of RW, CDRW, and RDRW diffusion models. We report posterior mean estimates along with 95% BCIs for trends and displacement variance rates. Trends for latitude and longitude coordinates are reported in units of degrees per year.

	RW		CDRW		RDRW	
Trend (Lat.)	—	—	—0.09	(−0.11, −0.06)	—0.03	(—0.05, —0.01)
Trend Changes (Lat.)	—	—	—	—	28.13	(27.0, 29.0)
Variance (Lat.)	0.25	(0.23, 0.27)	0.23	(0.21, 0.25)	0.13	(0.12, 0.14)
Trend (Long.)	—	—	0.30	(0.26, 0.33)	0.12	(0.08, 0.16)
Trend Changes (Long.)	—	—	—	—	1.48	(1.00, 3.00)
Variance (Long.)	0.59	(0.55, 0.64)	0.37	(0.34, 0.40)	0.43	(0.39, 0.45)
Correlation	−0.47	(−0.54, −0.40)	−0.40	(−0.47, −0.32)	−0.84	(−0.87, −0.82)

Bwamanda, and Kisangani, suggests that diffusion with a northward trend may more accurately characterize part of the spatial history. We explore this possibility with the RDRW model.

Under the RDRW model, there is significant evidence of multiple longitudinal trends. We estimate a posterior mean of 1.48 trend changes with BCI (1, 3), and the Bayes factor in favor of RDRW over CDRW is greater than 1000. The posterior mean longitudinal trend across all branches is 0.12 with BCI (0.08, 0.16). The MCC tree (Fig. 2) consists of a green-colored clade with inferred eastward trends of 0.18 degrees per year and a brown-colored clade with inferred trends close to or equal to 0. The observed and inferred longitudes provide better understanding of the difference in trends between the two clades. During the period 1960–2004, there is generally greater eastward movement along the lineages of the green clade. Although the lineages of the brown clade show greater eastward spread during the first half of the evolutionary history, the trends in the two clades are driven by the trends of the second half of the evolutionary history. The second half accounts for a much greater proportion of tree branches and, because it overlaps with all sampling times, contains more information about the spatial diffusion process.

For the latitudinal trend, we estimate 28.13 trend changes with BCI (27, 29). Furthermore, RDRW is supported over CDRW with a Bayes factor greater than 1000. The overall posterior mean trend across all branches of the MCC tree (Fig. 3) is −0.03 with BCI (−0.05, −0.01). Notably, external branches with positive inferred trends lead to samples from locations north of the origin (Bwamanda and Kisangani), whereas external branches with nonpositive trends lead to samples from locations with latitudes south of or similar to the origin.

Importantly, adopting a zero-mean displacement distribution when the diffusion process is more accurately described with a nontrivial mean can result in inflated displacement variance rate estimates. (Recall that the displacement variance along a branch of length t_i is $t_i\mathbf{P}^{-1}$, so by "displacement variance rates," we mean the diagonal elements of the variance rate matrix $\mathbf{P}^{-1}$.) By incorporating trends into the model, we are able to disentangle the displacement mean from the displacement variance and uncover a clearer picture of the movement. The variance rate estimates in Table 2 illustrate this point. Including a constant trend reduces the variance rate of the displacement of the longitude coordinate from 0.59 to 0.37. For the latitude, on the other hand, the variance rate decreases modestly from 0.25 to 0.23 and the BCIs of (0.23, 0.27) and (0.21, 0.25) overlap. The lack of an appreciable reduction in variance rate may be explained by the fact that the apparent northward trend to Bwamanda and Kisangani remains unaccounted for by inclusion of a constant southward trend. Indeed, by accommodating trend changes under the RDRW model, the variance rate of the latitudinal displacement drops to 0.13 with BCI (0.12, 0.14).

HIV-1 Resistance to Broadly Neutralizing Antibodies

It is widely believed that a successful HIV-1 vaccine will require the elicitation of neutralizing antibodies (Johnston and Fauci 2007; Barouch 2008; Walker and Burton 2008). Most neutralizing antibodies are strain-specific and therefore not so attractive for vaccine design (Weiss et al. 1985; Mascola and Montefiori 2010). It is important to identify and characterize antibody specificities that are effective against a large number of currently circulating HIV-1 variants (Burton 2002, 2004). Several broadly neutralizing monoclonal antibodies have been recently isolated, including PG9 and PG16 (Walker et al. 2009), and VRC01 (Zhou et al. 2010).

Studies comparing viruses isolated from individuals who seroconverted early in the HIV-1 epidemic to viruses from individuals who seroconverted in recent years have shown that HIV-1 has become increasingly resistant to antibody neutralization over the course of the epidemic (Bunnik et al. 2010; Euler et al. 2011; Bouvin-Pley et al. 2013). Bunnik et al. (2010) demonstrate a decreased sensitivity to polyclonal antibodies and to monoclonal antibody b12. Euler et al. (2011) extend those findings by investigating whether HIV-1 has adapted to the neutralization activity of PG9, PG16, and VRC01. Their results show that HIV-1 has become significantly more resistant to neutralization by VRC01 and also provide some support for increased resistance to neutralization by PG16.

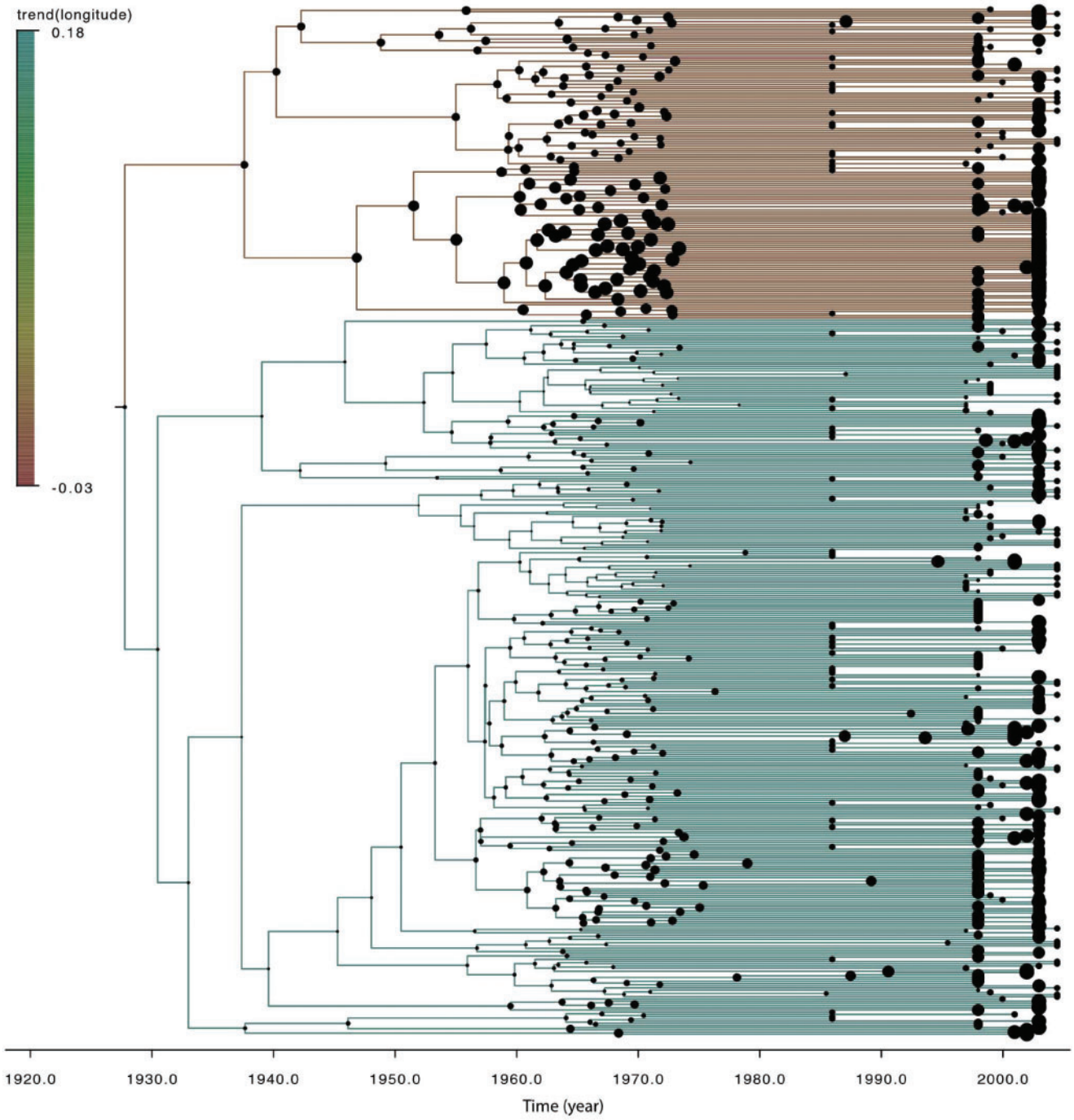

FIGURE 2. Maximum clade credibility tree for spread of HIV-1 in central Africa. The posterior mean longitudinal trend, in units of degrees per year, is depicted using a color gradient along the branches. Green indicates an eastward trend while brown signifies trends close to or equal to zero. Tree nodes are depicted as circles of different sizes. The size of each circle is determined by the longitude of the observed or inferred location corresponding to the node. Larger circles represent more eastward locations.

These studies typically do not account for phylogenetic dependence among the sampled viruses. Vrancken et al. (2015) examine the data set of Euler et al. (2011) with a Brownian diffusion trait evolution model that simultaneously infers phylogenetic signal, the degree to which resemblance in traits reflects phylogenetic relatedness. They find moderate phylogenetic signal and, through ancestral trait value reconstruction, more evidence of decreased sensitivity of HIV-1 to VRC01 and PG16 neutralization at the population level.

We follow up on the analysis of Vrancken et al. (2015) by incorporating trends into the Brownian trait evolution. The data set comprises clonal HIV-1 variants

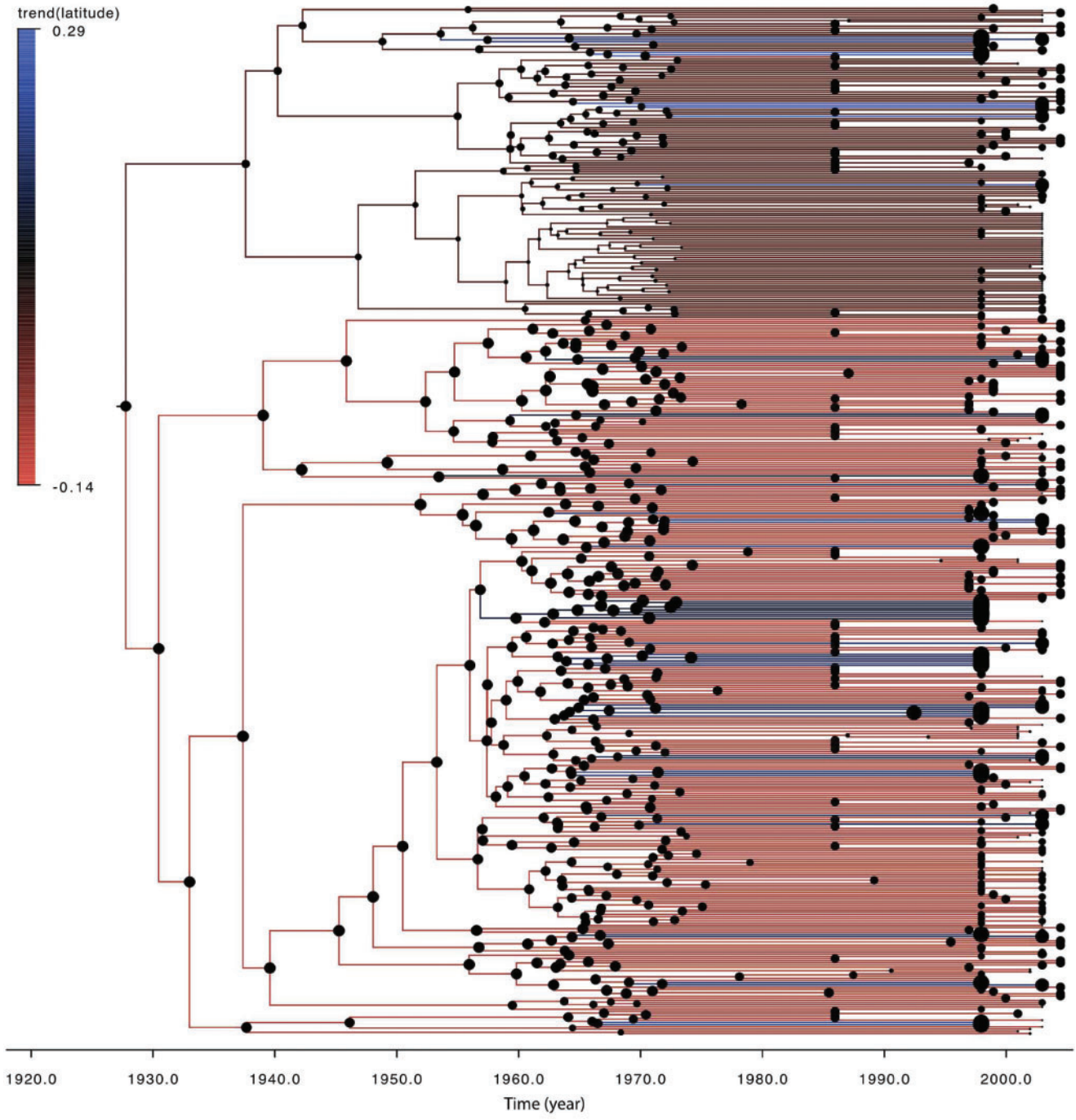

FIGURE 3. Maximum clade credibility tree for spread of HIV-1 in central Africa. The posterior mean latitudinal trend, in units of degrees per year, is depicted using a color gradient along the branches. The colors range between red and blue, with the former indicating a southward trend and the latter a northward trend. Tree nodes are depicted as circles of different sizes. The size of each circle is determined by the longitude of the observed or inferred location corresponding to the node. Larger circles represent more northward locations.

from "historic" and "contemporary" seroconverters with an acute or early subtype B HIV-1 infection. The 14 historic seroconverters have a known seroconversion date between 1985 and 1989, and the 21 contemporary seroconverters have a seroconversion date between 2003 and 2006. The percentage neutralization is determined by calculating the reduction in p24 production in the presence of the neutralizing agent compared to the p24 levels in the cultures with virus only. The trait values of interest are 50% inhibitory concentration (IC_{50}) assay values that summarize the percentage neutralization by antibodies PG9, PG16, and VRC01, measured in units of $\mu g/ml$. We take the log-transform of IC_{50} values in order to ensure that concentration values are strictly

TABLE 3. HIV-1 resistance to broadly neutralizing antibodies. Mean trends under CDRW, and RDRW models for log(IC_{50}) measurements corresponding to monoclonal neutralizing antibodies PG9, PG16, and VRC01. Higher log(IC_{50}) values represent lower sensitivity to antibody neutralization, and positive trends indicate a drift over time toward greater resistance. We report posterior means along with 95% BCIs. Trends are reported in units of log-μg/ml per year.

Antibody	CDRW		RDRW	
PG9	0.08	(−0.05, 0.19)	0.07	(−0.04, 0.18)
PG16	0.10	(−0.02, 0.20)	0.11	(−0.01, 0.24)
VRC01	0.15	(0.06, 0.24)	0.15	(0.09, 0.21)

positive under the diffusion process. Higher log(IC_{50}) values correspond to greater resistance to antibody neutralization. For viruses with log(IC_{50}) values that fall outside the tested antibody concentration range, we integrate out the concentration over a plausible IC_{50} interval.

First, we analyze the data with the CDRW model (see Table 3). The results are essentially consistent with the findings of Euler et al. (2011). For VRC01, we estimate a posterior mean trend of 0.15 with 95% BCI (0.06, 0.24), signaling a significant trend toward higher resistance to VRC01 neutralization. Furthermore, a Bayes factor of 32.33 lends strong support to the CDRW model over the standard RW. There is not as much evidence of a trend for PG16. On the one hand, the posterior probability that the trend is positive is 0.953, providing some corroboration for a decreased sensitivity to PG16 neutralization. However, we infer a mean trend of 0.10 with a 95% BCI (−0.02, 0.20) that contains zero. Furthermore, the Bayes factor in favor of CDRW over RW is just 1.32, showing little support for inclusion of a trend vector. We do not detect a significant trend in the case of PG9: the posterior mean is 0.08 with 95% BCI (−0.05, 0.19) and the Bayes factor is 1.17.

To take a closer look, we fit the data to the RDRW model. Along with branch-specific trends, we examine the posterior mean trends over the entire evolutionary history (see Table 3). First, we consider the results for PG9 and PG16. The posterior mean trends are similar to the inferred trends under the CDRW model, and their 95% BCIs contain 0. There is little support for any trend changes occurring along the phylogeny. The mean estimated number of trend changes are 0.17 and 0.2 for PG9 and PG16, respectively, and the Bayes factors in favor of RDRW over CDRW are 0.19 and 0.20. Hence there is not much evidence of localized directional trends that differ from the overall directional trends.

We illustrate the evolutionary pattern for resistance against VRC01 under the RDRW model in Figure 4. We obtain a posterior mean estimate of 1.13 trend changes, with a posterior probability of 0.88 for exactly one trend change and probability greater than 0.99 for at least one trend change. Furthermore, the Bayes factor in favor of RDRW over CDRW is 359.04, providing very strong support for RDRW. As shown in Table 4, the displacement variance rate decreases as we move from the RW model with no trend to the CDRW model, and then to the RDRW model. Figure 4 shows a trend change occurring at the common ancestral node of samples from subjects P001 and P002. Apart from the two branches leading to tips P001 and P002 (which we refer to as "branch P001" and "branch P002," respectively), the branches have essentially identical mean trends of about 0.15 with 95% BCI (0.09, 0.21). For the blue-colored branch P001, we have an estimated trend of 2.13 with 95% BCI (0.06, 4.89), and for the red-colored branch P002, the estimated trend is −1.18 with 95% BCI (−4.46, 0.22). Both estimated trends are drastically different from the parent branch trend, and their BCIs are also much wider. If either subject P001 or P002 is deleted from the data set, we infer a constant, significant trend of 0.15 over the entire evolutionary history. Notably, we do not infer any trend changes after deleting either P001 or P002.

Under the RDRW model, situations in which both child branches have trends that differ from the trend of the parent branch must be handled with care. In any given tree in the posterior sample, one of the two branches must inherit its trend from the parent branch. An examination of the posterior distribution of the trend change indicator at the parent node of tips P001 and P002 reveals that branch P001 inherits the parent branch trend with posterior probability of about 40% and branch P002 inherits the parent trend with the other 60% probability mass. Thus, the posterior trend estimate for each of the branches P001 and P002 averages over the cases where it inherits the parent trend and the cases where it differs from the parent trend. While the potentially large departures from the parent trend are still apparent in the "averaged" posterior estimates, it is of interest to find out how representative they are of the true trends. It is conceivable that the "inherited" portion of the posterior may bias the estimate, shifting the mean and widening the BCI. In the case of branch P002, for example, the posterior mass near the parent trend extends the BCI into the positive axis so that it includes zero. It is also conceivable that the "inherited" part of the posterior represents a part of the distribution that would show up even without the restrictions of the RDRW model.

To elucidate the true nature of the trend change that occurs at the parent of tips P001 and P002, we conduct a follow-up analysis. We introduce a new parameterization of our directional random walk diffusion model that posits three unique trends: one each corresponding to branches P001 and P002, and another for all remaining branches in the phylogenetic tree. The results, presented in Table 5, are similar to the findings from the RDRW model. Notably, the 95% BCIs for the trend on branches P001 and P002 still contain the range of credible values for the parent trend. There are also some key differences between the two analyses. The distributions for the trend on branches P001 and P002 are bell-shaped rather than bimodal, and the 95% BCIs are wider than under the RDRW model. Unlike the RDRW model estimate, the 95% BCI for the trend on branch P001 contains 0. However, it has a 0.95 posterior probability of being positive, so there is still some support that the trend on branch P001 is statistically significant.

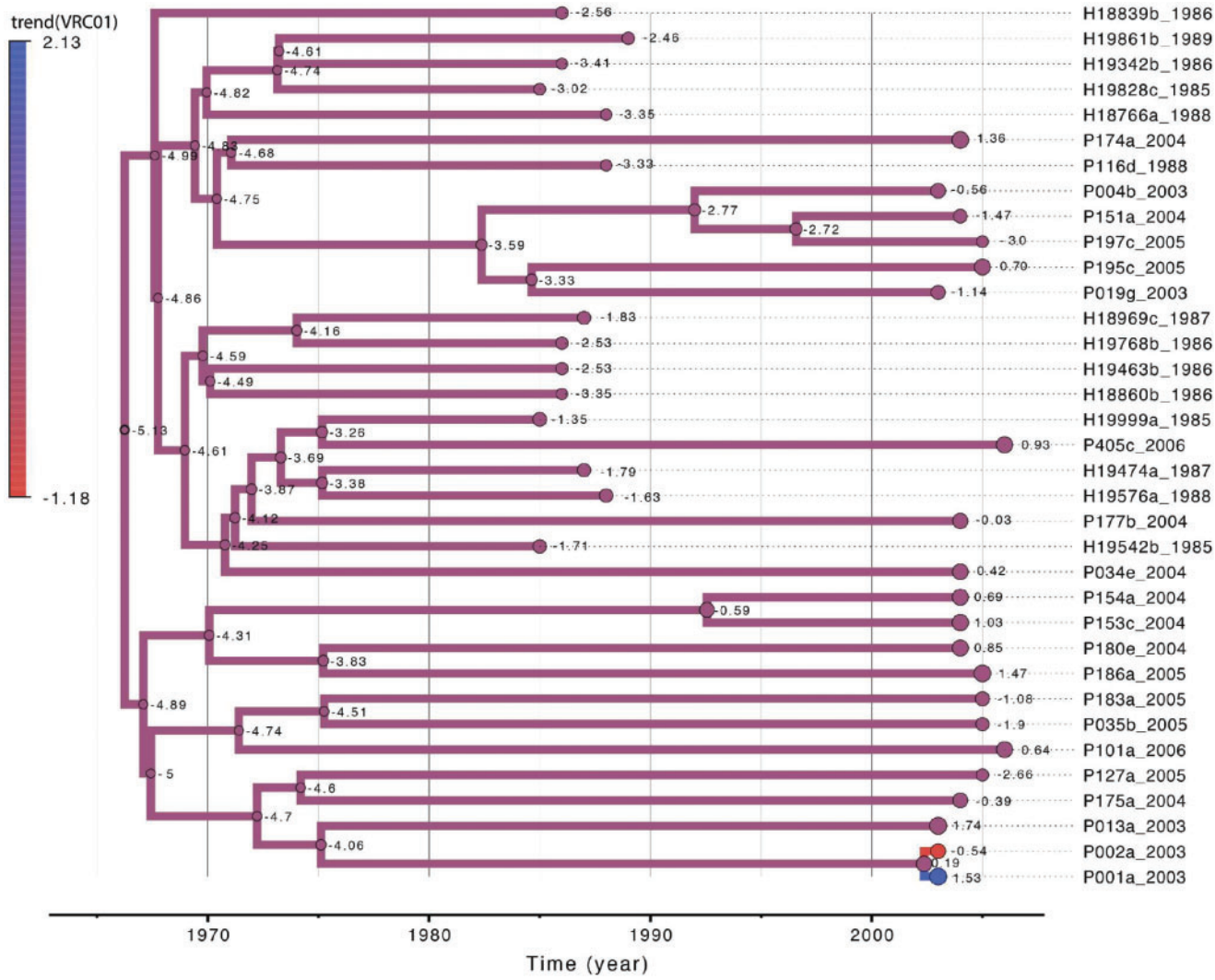

FIGURE 4. Maximum clade credibility tree depicting evolutionary pattern of HIV-1 resistance to neutralization by antibody VRC01. Subject identifiers corresponding to tree tips are listed to the right of the tree. The posterior mean trend is depicted using a color gradient along the branches. Positive trends correspond to a trend toward greater resistance to neutralization. The estimated posterior mean trend along the purple colored branches that make up most of the tree is 0.15. A trend change is inferred at the common ancestor of tips P001 and P002. The red-colored branch leading to the tip P002 has an estimated posterior mean trend of -1.18, and the blue-colored branch leading tip P001 has an estimated posterior mean trend of 2.13. Tree tips and internal nodes are annotated with observed and inferred log(IC_{50}) values, measured in units of log-μg/ml. Trends are in units of log-μg/ml per year.

TABLE 4. HIV-1 resistance to broadly neutralizing antibodies. Displacement variance rate under RW, CDRW and RDRW models for log(IC_{50}) measurements corresponding to monoclonal neutralizing antibodies PG9, PG16, and VRC01. We report posterior means along with 95% BCIs.

Antibody	Displacement Variance RW		CDRW		RDRW	
PG9	0.28	(0.18, 0.57)	0.26	(0.16, 0.53)	0.26	(0.17, 0.55)
PG16	0.26	(0.17, 0.51)	0.23	(0.15, 0.49)	0.23	(0.15, 0.51)
VRC01	0.19	(0.11, 0.43)	0.13	(0.08, 0.27)	0.06	(0.04, 0.14)

An examination of the MCC tree annotated with observed log(IC_{50}) values and inferred ancestral trait value realizations (Fig. 4) clarifies why we infer a trend change. Consider node triples consisting of two nodes and their common parent node. The triple of tips P001, P002, and their common ancestor features a relatively large difference in trait values at the child nodes as well as relatively short branch lengths connecting nodes P001 and P002, to their parent. While there are other triples with child nodes possessing a comparable difference in log(IC_{50}) values, they have much longer branches leading from the parent node to the children.

TABLE 5. Rate changes in HIV-1 drift toward resistance to antibody VRC01. We report posterior means, 95% BCIs, and posterior probabilities that the drift > 0.

Branch	RDRW Mean	RDRW 95% BCI	RDRW P(Drift > 0)	Fixed-Changes Mean	Fixed-Changes 95% BCI	Fixed-Changes P(Drift > 0)
P001	2.13	(0.06, 4.89)	> 0.99	2.27	(−0.36, 5.10)	0.95
P002	−1.18	(−4.46, 0.22)	0.60	−1.81	(−4.61, 1.03)	0.10
Other	0.15	(0.09, 0.21)	> 0.99	0.14	(0.08, 0.20)	> 0.99

Similarly, while other triples feature relatively short branches connecting the parent to the children, the trait values at the child nodes do not differ as much. The unique combination of short branches coupled with a large difference between $\log(IC_{50})$ values at the child nodes explains why the trend present on most of the tree may be incompatible with the triple of P001, P002 and their parent.

Although there is strong evidence of a trend change at the ancestral node of samples P001 and P002, the wide BCIs for the branch P001 and branch P002 trend estimates suggest that they are poorly informed by the data. The type of trend change that occurs at the ancestral node of tips P001 and P002 also remains unclear. Subjects P001 and P002 are a transmission couple and it appears that one person mounted a very different antibody response than the other to a highly similar virus. Further research may clarify the situation. Nevertheless, we infer a clear, significant trend toward increased resistance to neutralization by VRC01, and it is robust to deletion of either subject P001 or P002. At the population level, the phylogenetic structure of HIV is "starlike," featuring multiple co-circulating lineages, the dynamics of which generally reflects neutral epidemiological processes (Grenfell et al. 2004). It is therefore notable to find evidence of population level evolution toward increased resistance.

DISCUSSION

Standard Brownian diffusion is a popular and, in many ways, natural starting point for modeling continuous trait evolution in a phylogenetic context. On the other hand, it is very restrictive and may not adequately describe the dynamics of the underlying evolutionary process. Development of non-Brownian models, such as mean-reverting Ornstein–Uhlenbeck processes, represents a promising avenue. However, substantial gains can also be made through building upon standard Brownian diffusion approaches. For example, the displacement along a branch is typically assumed to have variance equal to the product of the branch length and a diffusion variance rate matrix $\mathbf{P}^{-1}$, where $\mathbf{P}^{-1}$ does not vary along the phylogenetic tree. O'Meara et al. (2006) and Lemey et al. (2010) demonstrate improvements by relaxing this homogeneity assumption via branch-specific diffusion rate scalars that yield a mixture of Brownian processes. On the other hand, Hunt (2006) and Slater et al. (2012a) show that progress toward a more realistic trait evolution process can also be made by relaxing the assumption of a zero-mean displacement to incorporate a constant directional trend. Here, we take the development of the displacement mean a step further by introducing a framework that accommodates multiple directional trends. Such development is foundational to highly general trait evolution processes. Notably, the Ornstein–Uhlenbeck process is nested within the diffusion process defined by

$$\mathbf{Y}_i|\mathbf{Y}_{pa(i)} \sim N\left(\boldsymbol{\beta}_1(t_i)\mathbf{Y}_{pa(i)} + \boldsymbol{\beta}_2(t_i)\boldsymbol{\mu}_i, \boldsymbol{\Sigma}(t_i)\right). \quad (46)$$

Consider the special case where $\boldsymbol{\mu}_i = \boldsymbol{\mu}$ for every branch, and

$$\boldsymbol{\beta}_1(t_i) = e^{-\alpha t_i}, \ \boldsymbol{\beta}_2(t_i) = 1 - e^{-\alpha t_i}, \text{ and}$$
$$\boldsymbol{\Sigma}(t_i) = \frac{\sigma^2}{2\alpha}\left[1 - e^{-2\alpha t_i}\right]. \quad (47)$$

This is equivalent to an Ornstein–Uhlenbeck process on the phylogenetic tree defined by the stochastic differential equation

$$d\mathbf{Y}_t = \alpha(\boldsymbol{\mu} - \mathbf{Y}_t)dt + \sigma d\mathbf{W}_t, \quad (48)$$

where $\mathbf{W}_t$ is a standard Brownian diffusion process (Hansen 1997). Here, $\boldsymbol{\mu}$ can be thought of as an optimal trait value, α represents the strength of selection toward $\boldsymbol{\mu}$, and σ^2 is the variance of the Brownian diffusion component. Such generality enables formal testing between a wide class of different Gaussian process models.

We introduce a flexible new Bayesian framework for phylogenetic trait evolution, modeling the evolutionary process as Brownian diffusion with a nontrivial drift. By allowing an estimable mean vector in the displacement distribution, we can account for and quantify a directional trend. However, imposing a constant rate directional trend can make for an unrealistic approximation of the underlying process. We overcome this limitation through the RDRW model. The RDRW model permits trend variation along a phylogenetic tree while maintaining model identifiability. Trends are generally passed on from parent branches to child branches, and variation is achieved by allowing at most one branch of any given pair of child branches to assume a different trend than their common parent branch.

The utility of incorporating a trend into the diffusion is corroborated by our analyses of two viral examples. We apply our methodology to both geographic traits in

a phylogeographic setting as well as phenotypic traits. Our phylogeographic analysis of the spread of HIV-1 in central Africa confirms the findings obtained by discrete phylogeographic inference (Faria et al. 2014). Directional random walk models fit the data better than an unbiased random walk (standard Brownian diffusion), and we uncover directional trends in the dispersal of the virus from its origin to sampling locations. We also see that trend variation characterizes real spatiotemporal diffusion processes. The absence of a directional trend in the diffusion model can lead to conflation of the latent trend with other parameters, particularly the displacement variance. Our analysis of the spread of HIV-1 illustrates how inferred displacement variance rates can decrease with appropriate directional trend modeling, revealing a clearer, more detailed picture of dispersal dynamics. Notably, this enables improved ancestral location reconstructions, which are widely used in in epidemiological analyses (Pybus et al. 2015).

While it is tempting to assume trend variation and seek out the additional insight it may provide, the data may not support multiple trends, even when there is a significant constant trend. Parameterizing the model to enforce the maximal number of unique trends can result in numerous small trend changes that are a consequence of the modeling choice and are not necessarily driven by the data. Our RDRW framework overcomes this issue by inferring the locations and types of trend changes directly from the data as opposed to making *a priori* assumptions about the number of unique trends and their appropriate assignments. Bayesian stochastic search variable selection enables efficient exploration of all possible trend configurations.

Although our focus has been on directional trends, a major strength of our approach is its implementation in the larger Bayesian phylogenetic framework of BEAST. Through BEAST, we have access to a plethora of different models for molecular character substitution, demographic history, and molecular clocks. Bayesian inference provides a natural framework for controlling for different sources of uncertainty in evolutionary models, including the phylogenetic tree and trait and sequence evolution parameters, and testing evolutionary hypotheses.

The gains from introducing directional trends into our real data examples are encouraging, and we anticipate that our RDRW approach will be useful in other scenarios not examined here. For example, antigenic drift is the process by which influenza viruses evolve to evade the immune system, and understanding its dynamics is essential to public health efforts. Bedford et al. (2014) have recently developed an integrated approach to mapping antigenic phenotypes that combines it with genetic information. It may be fruitful to model the diffusion of the antigenic phenotype in their framework with the RDRW model. Also, while we have focused on viral examples in this article, comparative studies of morphology have largely motivated continuous trait evolutionary modeling. An important example can by found in the work of Hunt (2007), who examines the fit of different evolutionary models to a sample of over 250 cases of phenotypic traits evolving along fossil lineages. The study finds that Brownian diffusion with a constant directional trend was more strongly supported than standard Brownian diffusion and stasis models in only 5% of the cases. Hunt (2007) suggests that this may be due to directional evolution typically occurring only for relatively short durations. The RDRW model may provide the flexibility necessary to characterize such evolutionary scenarios.

While the RDRW framework has proven to be flexible and useful, there is a need for continued development of more realistic trait evolution models. The restrictive parameterization of the RDRW model may render it inappropriate for some evolutionary scenarios. One of its limiting aspects is that trend changes can only occur at phylogeny nodes. While permitting changes at nodes has proven to be a popular technique for evolutionary parameter variation in phylogenetic inference models, it may be beneficial to relax this assumption to allow trends to change between nodes. For example, Revell et al. (2012) propose a standard Brownian diffusion model with only two distinct displacement variance rates in a tree and allow the single transition between rates to occur at any position along a branch. However, allowing trends to change more than once in such a way results in identifiability issues in the data likelihood. Overcoming these issues would require strong prior information or considerable restrictions on the number of unique trends. Another major restriction of the RDRW parameterization is the way it forces trends to be inherited. For example, once a trend appears anywhere in the phylogenetic tree, the restrictions mandate that it must be passed on and "survive" until it reaches an external branch. From this perspective, sequences and traits sampled throughout evolutionary history, as can be obtained for rapidly evolving pathogens, represent particularly interesting applications. We note that lack of support for the survival of a specific trend may not preclude its inclusion under the RDRW model. In the analysis of HIV-1 resistance to neutralization, for example, the parent branch of the two branches leading to P001 and P002 must pass on its trend to one of these two child branches in each tree in the posterior sample. Yet, the branch which is forced to inherit the parent trend alternates between P001 and P002 in the posterior sample, reflecting the uncertainty and resulting in posterior trend distributions for both child branches that differ from that of their parent trend. However, this unnatural mechanism of deflecting an unsupported trend may contribute to misleading trend estimates. It would be preferable to sidestep such problems by developing alternative models that accommodate trend variation while retaining identifiability.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.v43rd.

FUNDING

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement [no. 278433-PREDEMICS] and ERC Grant Agreement [no. 260864] and the National Institutes of Health [R01 AI107034, R01 HG006139, R01 LM011827, and T32 AI007370] and the National Science Foundation [DMS 1264153].

ACKNOWLEDGMENTS

We would like to thank Luke Harmon, Mark Holder, Frank Anderson, and two anonymous reviewers for constructive comments that helped improve this manuscript.

APPENDIX

Identifiability

To ensure that our results are meaningful, it is important to understand the conditions under which our model is identifiable. For convenience, and without loss of generality, we assume here that traits are one-dimensional. Following the development in the "Methods" section, the observed trait values $\mathbf{Y}=(\mathbf{Y}_1,\ldots,\mathbf{Y}_N)^t$ at the tips of the phylogeny τ are multivariate-normal distributed:

$$P\left(\mathbf{Y}|\mathbf{Y}_{2N-1},\mathbf{P},\mathbf{V}_\tau,\boldsymbol{\mu}_\tau\right) = \mathrm{MVN}\left(\mathbf{Y};\mathbf{Y}_{\mathrm{root}}+\mathbf{T}\boldsymbol{\mu}_\tau,\mathbf{P}^{-1}\otimes\mathbf{V}_\tau\right). \quad (A.1)$$

Here, $\mathbf{Y}_{\mathrm{root}}$ is an $N\times 1$ vector with the root trait value $\mathbf{Y}_{2N-1}$ repeated N times. The vector $\boldsymbol{\mu}_\tau=(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_{2N-2})^t$ consists of branch-specific trends $\boldsymbol{\mu}_i$. The $N\times N$ variance matrix $\mathbf{V}_\tau$ is a deterministic function of τ and represents the contribution of the phylogenetic tree to the covariance structure. Its diagonal entries V_{ii} are equal to the distance in time between the tip $\mathcal{V}_i$ and the root node $\mathcal{V}_{2N-1}$, and off-diagonal entries V_{ij} correspond to the distance in time between the root node $\mathcal{V}_{2N-1}$ and the most recent common ancestor of tips $\mathcal{V}_i$ and $\mathcal{V}_j$. Finally, the $N\times(2N-2)$ matrix $\mathbf{T}$ is defined as follows: $T_{ij}=t_j$, the length of branch j, if branch j is part of the path from the external node i to the root, and $T_{ij}=0$ otherwise. In other words, the ith row of $\mathbf{T}$ specifies the path of branches in τ connecting external node i to the root.

Let $\mathbf{V}(\boldsymbol{\mu}_\tau)$ denote the vector space of permissible values of $\boldsymbol{\mu}_\tau$ for our model. With respect to the trend, the model is identifiable if the equality

$$P\left(\mathbf{Y}|\mathbf{Y}_{2N-1},\mathbf{P},\mathbf{V}_\tau,\boldsymbol{\mu}_\tau\right)=P\left(\mathbf{Y}|\mathbf{Y}_{2N-1},\mathbf{P},\mathbf{V}_\tau,\boldsymbol{\mu}_\tau^*\right) \quad (A.2)$$

implies that $\boldsymbol{\mu}_\tau=\boldsymbol{\mu}_\tau^*$. Because the trend vector appears only in the mean of the distribution, we have an identifiability problem if the same mean $\mathrm{E}(\mathbf{Y})$ can be realized from different values of $\boldsymbol{\mu}_\tau$. In other words, if there exist $\boldsymbol{\mu}_\tau\neq\boldsymbol{\mu}_\tau^*$ such that $\mathbf{T}\boldsymbol{\mu}_\tau=\mathbf{T}\boldsymbol{\mu}_\tau^*$. This can happen if and only if the linear transformation $\mathbf{T}$ has a nontrivial kernel. We know that

$$\dim\mathbf{V}(\boldsymbol{\mu}_\tau)=\dim\ker(\mathbf{T})+\dim\mathrm{range}(\mathbf{T}), \quad (A.3)$$

and we also know that for any phylogeny τ, $\mathbf{T}$ is of full rank because its rows are linearly independent. It follows that for the kernel of $\mathbf{T}$ to be trivial, we must have

$$\dim\mathbf{V}(\boldsymbol{\mu}_\tau)\leq N. \quad (A.4)$$

If we allow a unique trend on each branch of τ, we have $\mathbf{V}(\boldsymbol{\mu}_\tau)=\mathbb{R}^{2N-2}$. Therefore, we must take a different approach.

It is illuminating to look at identifiability from the perspective of linear equations. For a given $\boldsymbol{\mu}_\tau$, $\mathbf{T}$ maps $\boldsymbol{\mu}_\tau$ to an $N\times 1$ vector $\boldsymbol{\gamma}$:

$$\mathbf{T}\boldsymbol{\mu}_\tau=\boldsymbol{\gamma}. \quad (A.5)$$

Identifiability is then equivalent to the system (A.5) of N linear equations having a unique solution.

To achieve identifiability, we introduce the RDRW model. Starting with a trend on the unobserved branch leading to the root node and moving down the tree toward the external nodes, every time a branch splits into two branches, one of two things happens. Either both of the child branches inherit the trend of the parent branch, or exactly one of the child branches inherits the trend from the parent branch while the other gets a new trend. Both child branches taking on different trends than the parent branch is not permitted. To avoid confusion, we continue to denote the $2N-2$ branch-specific trends as $\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_{2N-2}$, with the understanding that they are not all unique. We let $\boldsymbol{\mu}_1^*,\ldots,\boldsymbol{\mu}_K^*$ denote the unique trends, where $K\leq N$.

Definition: We say that a row $\mathbf{T}_i=(T_{i1},T_{i2},\ldots,T_{i,2N-2})$ in $\mathbf{T}$ is $\boldsymbol{\mu}_k^*-$*dominated* if its associated path from the root of τ to a tip ends with a branch with trend $\boldsymbol{\mu}_k^*$. We also refer to the sum $\mathbf{T}_i\boldsymbol{\mu}_\tau=\sum_{j=1}^{2N-2}T_{ij}\boldsymbol{\mu}_j$ and the path associated with $\mathbf{T}_i$ as $\boldsymbol{\mu}_k^*-$*dominated*.

Note that each unique trend dominates at least one path.

Definition: An *initial branch* of the trend $\boldsymbol{\mu}_k^*$ is a branch whose parent branch has a different trend. The unobserved branch leading to the root node is also defined to be an initial branch.

Observe that every branch with trend $\boldsymbol{\mu}_k^*$ is an initial branch of $\boldsymbol{\mu}_k^*$ or a descendant of an initial branch of $\boldsymbol{\mu}_k^*$. A trend may have more than one initial branch. In order to quantify how deep into the tree τ a trend extends (starting from the tips and going toward the root), we make the following definition.

Definition: By a *descendant path* of a branch b, we mean a series of connected branches, starting with a child branch of b and ending with a branch leading to a tip. The *depth* of a branch b is equal to the maximal number of branches

in descendant paths of b. The depth of a trend $\boldsymbol{\mu}_k^*$ is equal to the maximal depth of its initial branches.

For example, if $\boldsymbol{\mu}_k^*$ has one initial branch leading to a tip, then $\boldsymbol{\mu}_k^*$ has depth 0. If the number of unknowns K is less than the number of equations N in the system $\mathbf{T}\boldsymbol{\mu}_\tau = \boldsymbol{\gamma}$, a unique solution can be established by working with a reduced system. We form the reduced system by choosing K of the N rows in $\mathbf{T}$, say $\mathbf{T}_{i_1}, \ldots, \mathbf{T}_{i_K}$, such that each is dominated by a different trend. If a trend dominates more than one path, we choose a path containing a maximal depth initial branch of the trend for the reduced system.

Claim: The RDRW model is identifiable.

Proof: The reduced linear system

$$\sum_{j=1}^{2N-2} T_{i_1 j}\boldsymbol{\mu}_j = \boldsymbol{\gamma}_{i_1} \tag{A.6}$$

$$\sum_{j=1}^{2N-2} T_{i_2 j}\boldsymbol{\mu}_j = \boldsymbol{\gamma}_{i_2} \tag{A.7}$$

$$\ldots \tag{A.8}$$

$$\sum_{j=1}^{2N-2} T_{i_K j}\boldsymbol{\mu}_j = \boldsymbol{\gamma}_{i_K}, \tag{A.9}$$

consists of K equations and K variables. Therefore to show that the solution is unique, it suffices to show that the linear system is independent. To establish independence, it suffices to show that if

$$a_1 \sum_{j=1}^{2N-2} T_{i_1 j}\boldsymbol{\mu}_j + a_2 \sum_{j=1}^{2N-2} T_{i_2 j}\boldsymbol{\mu}_j \cdots + a_N \sum_{j=1}^{2N-2} T_{i_K j}\boldsymbol{\mu}_j = 0, \tag{A.10}$$

for some constants $a_1, \ldots, a_K$, then we must have

$$a_1 = a_2 = \cdots = a_K = 0. \tag{A.11}$$

Suppose Equation (A.10) holds. The idea behind the proof is as follows: we consider all trends of depth 0, conclude that each sum in (A.10) dominated by a trend of depth 0 must have its corresponding coefficient $a_i = 0$, then consider all trends of depth 1, conclude that each sum in (A.10) dominated by a trend of depth 1 must have corresponding coefficient $a_i = 0$, and so on until we have gone through all possible depth values of trends in τ.

Suppose $\boldsymbol{\mu}_k^*$ has depth 0. Then $\boldsymbol{\mu}_k^*$ only appears in the single $\boldsymbol{\mu}_k^*$-dominated sum and cannot be canceled out by a linear combination of the other sums. This forces the coefficient a_i of the $\boldsymbol{\mu}_k^*$-dominated sum in (A.10) to be equal to zero. Having shown that any sum dominated by a trend of depth 0 must have a zero coefficient in (A.10), we can move on to the case of depth 1. Rather than handle the case of depth 1 separately, we present a general argument.

Suppose the coefficients of all sums in (A.10) that are dominated by trends of depth less than m have been shown to be zero. Consider trends of depth m. If $\boldsymbol{\mu}_k^*$ has depth m, then it appears in the $\boldsymbol{\mu}_k^*$-dominated path, and it may appear in paths dominated by other rates. Suppose $\boldsymbol{\mu}_k^*$ appears in a path P_i dominated by a different trend, say $\boldsymbol{\mu}_i^*$. By construction of the reduced system, P_i contains an initial branch $\mathbf{b}_i$ of $\boldsymbol{\mu}_i^*$ of maximal depth. This means the depth of $\mathbf{b}_i$ is equal to the depth of $\boldsymbol{\mu}_i^*$. Because $\mathbf{b}_i$ is a descendant of a branch with trend $\boldsymbol{\mu}_k^*$, the depth of $\boldsymbol{\mu}_i^*$ must be less than the depth of $\boldsymbol{\mu}_k^*$. But sums dominated by trends of depth less than m have already been shown to have zero coefficients in (A.10). Thus the sum associated with P_i has coefficient zero. Because $\boldsymbol{\mu}_k^*$ appears in only one sum which is not already known to have a zero coefficient, the $\boldsymbol{\mu}_k^*$-dominated sum, it follows that in order for (A.10) to hold, the $\boldsymbol{\mu}_k^*$-dominated sum must also have a zero coefficient. Therefore sums dominated by trends of depth m must have zero coefficients in Equation (A.10).

Invoking this argument until we have gone through all possible values of trend depth, it follows that $a_1 = a_2 = \cdots = a_K = 0$. ■

References

Agrawal A., Fishbein M. 2008. Phylogenetic escalation and decline of plant defense strategies. Proc. Natl. Acad. Sci. 105:10057–10060.

Agrawal A., Fishbein M., Halitschke R., Hastings A., Rabosky D., Rasmann S. 2009. Evidence for adaptive radiation from a phylogenetic study of plant defenses. Proc. Natl. Acad. Sci. 106:18067–18072.

Allicock O.M., Lemey P., Tatem A.J., Pybus O.G., Bennett S.N., Mueller B.A., Suchard M.A., Foster J.E., Rambaut A., Carrington C.V. 2012. Phylogeography and population dynamics of dengue viruses in the americas. Mol. Biol. Evol. 29:1533–1543.

Baele G., Lemey P., Suchard M.A. 2016a. Genealogical working distributions for bayesian model testing with phylogenetic uncertainty. Sys. Biol. 65:250–264.

Baele G., Suchard M.A., Rambaut A., Lemey, P. 2016b. Emerging concepts of data integration in pathogen phylodynamics. Systematic Biology, in press.

Barouch D. 2008. Challenges in the development of an HIV-1 vaccine. Nature 455:613–619.

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen, T. 2012. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biol. 314:204–215.

Beaulieu J., Jhwueng D., Boettiger C., O'Meara, B. 2012. Modeling stabilizing selection: Expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution 66: 2369–2383.

Bedford T., Suchard M., Lemey P., Dudas G., Gregory V., Hay A., McCauley J., Russell C., Smith D., Rambaut, A. 2014. Integrating influenza antigenic dynamics with molecular evolution. Elife 3:e01914.

Bennett S., Drummond A., Kapan D., Suchard M., Munoz-Jordan J., Pybus O., Holmes E., Gubler D. 2010. Epidemic dynamics revealed in dengue evolution. Mol. Biol. Evol. 27:811–818.

Biek R., Henderson J., Waller L., Rupprecht C., Real L. 2007. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. PNAS 104:7993–7998.

Bikandou B., Ndoundou-Nkodia M., Niama F., Ekwalanga M., Obengui O., Taty-Taty R., Parra H., Saragosti S. 2004. Genetic subtyping of gag and env regions of HIV type 1 isolates in the Republic of Congo. AIDS Res. Hum. Retroviruses 20.

Bikandou B., Takehisa J., Mboudjeka I., Ido E., Kuwata T., Miyazaki Y., Moriyama H., Harada Y., Taniguchi Y., Ichimura H., Ikeda M., Ndolo

P., Nzoukoudi M., M'Vouenze R., M'Pandi M., Parra H., M'Pele P., Hayami M. 2000. Genetic subtypes of HIV type 1 in Republic of Congo. AIDS Res. Hum. Retroviruses 16:613–619.
Bloomquist E.W., Lemey P., Suchard M.A. 2010. Three roads diverged? routes to phylogeographic inference. Trends Ecol. & Evol. 25: 626–632.
Boucher F. Démery V. 2016. Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. Syst. Biol. 65: 651–661.
Bouvin-Pley M., Morgand M., Moreau A., Jestin P., Simonnet C., Tran L., Goujard C., Meyer L., Barin F., Braiband M. 2013. Evidence for a continuous drift of the HIV-1 species towards higher resistance to neutralizing antibodies over the course of the epidemic. PLoS Pathog. 9:e1003477.
Brown R. 1828. A brief account of microscopial observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Philos. Mag. 4:161–173.
Bunnik E., Euler Z., Welkers M., Boeser-Nunnink B., Grijsen M., Prins J., Schuitemaker H. 2010. Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level. Nat. Med. 16: 995–997.
Burton D. 2002. Antibodies, viruses and vaccines. Nat. Rev. Immunol. 2:706–713.
Burton D. 2004. HIV vaccine design and the neutralizing antibody problem. Nat. Immunol. 5:233–236.
Butler M. King A. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164.
Chipman, H., George E., McCulloch R. 2001. The practical implementation of Bayesian model selection. IMS Lecture Notes - Monograph Series 38:67–134.
Drummond A., Ho S., Phillips M., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.
Drummond A. Suchard M. 2010. Bayesian random local clocks, or one rate to rule them all. BMC Biol. 8:114.
Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.
Eastman J., Alfaro M., Joyce P., Hipp A., Harmon L. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–89.
Euler Z., Bunnik E., Burger J., Boeser-Nunnink B., Grijsen M., Prins J., Schuitemaker H. 2011. Activity of broadly neutralizing antibodies, including PG9, PG16, and VRC01, against recently transmitted subtype B HIV-1 variants from early and late in the epidemic. J. Virol. 85:7236–7245.
Faria N., Rambaut A., Suchard M., Baele G., Bedford T., Ward M., Tatem A., Sousa J., Arinaminpathy N., Pepin J., Posada D., Peeters M., Pybus O., Lemey P. 2014. The early spead and epidemic ignition of HIV-1 in human populations. Science 346: 56–61.
Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25: 471–492.
Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 13:93–104.
Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.
Felsenstein J. 1988. Phylogenies and quantitative characters. Ann. Rev. Ecol. Syst. 19:445–471.
Freckleton R., Harvey P., Pagel M. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. Am. Nat. 160: 712–726.
George E., McCulloch R. 1993. Variable selection via Gibbs sampling. JASA 88:881–889.
Gill M., Lemey P., Bennett S., Biek R., Suchard M. 2016. Understanding past population dynamics: Bayesian coalescent-based modeling with covariates. Syst. Biol. 65:1041–1056.
Gill M., Lemey P., Faria N., Rambaut A., Shapiro B., Suchard M. 2013. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. 30: 713–724.
Grenfell B., Pybus O., Gog J., Wood J., Daly J., Mumford J., Holmes E. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327–332.
Hansen T. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 52:1341–1351.
Hansen T., Martins E. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.
Hansen T., Pienaar J., Orzack S. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–1967.
Harvey P., Pagel M. 1991. The comparative method in evolutionary biology. Oxford, UK: Oxford University Press.
Hasegawa M., Kishino H., Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.
Hastings W. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.
Huelsenbeck J., Rannala B. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. Evolution 57:1237–1247.
Hunt G. 2006. Fitting and comparing models of phyletic evolution: Random walks and beyond. Paleobiology 32:578–601.
Hunt G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proc. Nat. Acad. Sci. USA 104:18404–18408.
Jeffreys H. 1935. Some tests of significance, treated by the theory of probability. Math. Proc. Cambridge Philos. Soc. 31:203–222.
Jeffreys H. 1961. Theory of Probability. Oxford, UK: Oxford University Press.
Johnston, M., Fauci A. 2007. An HIV vaccine: evolving concepts. N. Eng. J. Med. 356:2073–2080.
Kalish M., Robbins K., Pieniazek D., Schaefer A., Nzilambi N., Quinn T., Louis M.S., Youngpairoj A., Phillips J., Jaffe H., Folks T. 2004. Recombinant viruses and early global HIV-1 epidemic. Emerg. Infect. Diseases 10:1227–1234.
Kass R. Raftery A. 1995. Bayes factors. JASA 90:773–795.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111–120.
Kita K., Ndembi N., Ekwalanga M., Ido E., Kazadi R., Bikandou B., Takehisa J., Takemura T., Kageyama S., Tanaka J., Parra H., Hayami M., Ichimura H. 2004. Genetic diversity of HIV type 1 in Likasi, southeast of the Democratic Republic of Congo. AIDS Res. Hum. Retroviruses 20:1352–1357.
Kuo L., Mallick B. 1998. Variable selection for regression models. Sankhya B 60:65–81.
Lartillot N., Poujol R. 2011. A phylogenetic model for investigating correlated evolution of subsitution rates and continuous phenotypic characters. Mol. Biol. Evol. 28:729–744.
Lemey P., Rambaut A., Drummond A., Suchard M. 2009a. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5:e1000520.
Lemey P., Rambaut A., Welch J., Suchard M. 2010. Phylogeography takes a relaxed random walk in continuous space and time. Mole. Biol. Evol. 27:1877–1885.
Lemey P., Suchard M., Rambaut A. 2009b. Reconstructing the initial global spread of a human influenza pandemic: A Bayesian spatial-temporal model for the global spread of H1N1pdm. PLoS Curr. 1:RRN1031.
Martins E. 1999. Estimation of ancestral states of continuous characters: a computer simulation study. Syst. Biol. 48:642–650.
Mascola J., Montefiori D. 2010. The role of antibodies in HIV vaccines. Ann. Rev. Immun. 28:413–444.
Mello F., Araujo O., Lago B., Motta-Castro A., Moraes M., Gomes S., Bello G., Araujo N. 2013. Phylogeography and evolutionary history of hepatitis B virus genotype F in Brazil. Virol. J. 10:236.
Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E. 1953. Equation of state calculation by fast computing machines. J. Chem. Phys. 21:1087–1092.
Niama F., Toure-Kane C., Vidal N., Obengui P., Bikandou B., Ndoundou-Nkodia M., Montavon C., Diop-Ndiaye H., Mombouli J., Mokondzimobe E., Diallo A., Delaporte E., Parra H., Peeters M., Mboup S. 2006. HIV-1 subtypes and recombinants in the Republic of Congo. Infect., Genet. Evol. 6:337–343.

Oakley T., Cunningham C. 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. Evolution 54:397–405.

O'Meara B., Ane C., Sanderson M., Wainwright P. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Pagel M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol. 48:612–622.

Pant S., Goswami A., Finarelli J. 2014. Complex body size trends in the evolution of sloths (Xenarthra: Pilosa). BMC Evol. Biol. 14: 184.

Pybus O., Suchard M., Lemey P., Bernardin F., Rambaut A., Crawford F., Gray R., Arinaminpathy N., Stramer S., Busch M., Delwart E. 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proc. Nat. Acad. Sci. USA 109: 15066–15071.

Pybus O., Tatem A., Lemey P. 2015. Virus evolution and transmission in an ever more connected world. Proc. Roy. Soc. B 282:20142878.

Revell L., Harmon L. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10:311–331.

Revell L., Mahler D., Peres-Neto P., Redelings B. 2012. A new phylogenetic method for identifying exceptional phenotypic diversification. Evolution 66:135–146.

Ronquist F. 2004. Bayesian inference of character evolution. Trends Ecol. Evol. 19:475–481.

Seetahal J., Velasco-Billa A., Allicock O., Adesiyun A., Bissessar J., Amour K., Phillip-Hosein A., Marston D., McElhinney L., Shi M., Wharwood C., Fooks A., Carrington C. 2013. Evolutionary history and phylogeography of rabies viruses associated with outbreaks in Trinidad. PLoS Negl. Trop. Dis. 7:e2365.

Sinsheimer J., Lake J., Little R. 1996. Bayesian hypothesis testing of four-taxon topologies using molecular sequence data. Biometrics 52:193–210.

Slater G., Harmon L., Alfaro M. 2012a. Integrating fossils with molecular phylogenies improves inference of trait evolution. Evolution 66:3931–3944.

Slater G., Harmon L., Wegmann D., Joyce P., Revell L., Alfaro M. 2012b. Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate Bayesian computation. Evolution 66:752–762.

Sookias R., Butler R., Benson R. 2012. Rise of dinosaurs reveals major body-size transitions are driven by passive processes of trait evolution. Proc. Roy. Soc. B 279:2180–2187.

Suchard M., Weiss R., Sinsheimer J. 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. Biometrics 61:665–673.

Thomas G., Freckleton R., Szekely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc. Roy. Soc. B 273:1619–1624.

Vidal N., Mulanga C., Bazepeo S., Mwamba J., Tshimpaka J., Kashi M., Mama N., Laurent C., Lepira F., Delaporte E., Peeters M. 2005. Distribution of HIV-1 variants in the Democratic Republic of Congo suggests increase of subtype C in Kinshasa between 1997 and 2002. J. Acquir. Immune Defic. Syndr. 40:456–462.

Vidal N., Peeters M., Mulanga-Kabeya C., Nzilambi N., Robertson D., Ilunga W., Sema H., Tshimanga K., Bongo B., Delaporte E. 2000. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J. Virol. 74:10498–10507.

Vrancken B., Lemey P., Rambaut A., Bedford T., Longdon B., Günthard H.F., Suchard M.A. 2015. Simultaneously estimating evolutionary history and repeated traits phylogenetic signal: applications to viral and host phenotypic evolution. Methods Ecol. Evol. 6:67–82.

Walker B., Burton D. 2008. Toward an AIDS vaccine. Science 320: 760–764.

Walker L., Phogat S., Chan-Hui P., Wagner D., Phung P., Goss J., Wrin T., Simek M., Fling S., Mitcham J., Lehrman J., Priddy F., Olsen O., Frey S., Hammond P., Investigators P.G.P., Kaminsky S., Zamb T., Moyle M., Koff W., Poignard P., Burton D. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science 326:285–289.

Weiss R., Clapham P., Cheingsong-Popov R., Dalgleish A., Carne C., Weller I., Tedder R. 1985. Neutralization of human T-lymphocyte virus type III by sera of AIDS and AIDS-risk patients. Nature 316: 69–72.

Wiener N. 1958. Nonlinear problems in random theory. Cambridge (MA), MIT Press.

Woolhouse M., Rambaut A., Kellam P. 2015. Lessons from Ebola: Improving infectious disease surveillance to inform outbreak management. Sci. Transl. Med. 7:307rv5.

Yang C., Li M., Mokili J., Winter J., Lubaki N., Mwandagalirwa K., Kasali M., Losoma A., Quinn T., Bollinger R., Lal R. 2005. Genetic diversification and recombination of HIV type 1 group M in Kinshasa, Democratic Republic of Congo. AIDS Res. Hum. Retroviruses 21:661–666.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.

Zhou T., Georgiev I., Wu X., Yang Z., Dai K., Finzi A., Kwon Y., Scheid J., Shi W., Xu L., Yang Y., Zhu J., Nussenzweig M., Sodroski J., Shapiro L., Nabel G., Mascola J., Kwong P. 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science 329:811–817.