Syst. Biol. 70(2):340–359, 2021

DOI:10.1093/sysbio/syaa061
Advance Access publication July 29, 2020

Phylogeographic Estimation and Simulation of Global Diffusive Dispersal

STILIANOS LOUCA[1,2,*]

[1]*Department of Biology, University of Oregon, USA; and* [2]*Institute of Ecology and Evolution, University of Oregon, USA*
[*]*Correspondence to be sent to: Stilianos Louca*
E-mail: louca.research@gmail.com

Received 12 April 2020; reviews returned 19 July 2020; accepted 20 July 2020
Associate Editor: Vincent Savolainen

Abstract.—The analysis of time-resolved phylogenies (timetrees) and geographic location data allows estimation of dispersal rates, for example, for invasive species and infectious diseases. Many estimation methods are based on the Brownian Motion model for diffusive dispersal on a 2D plane; however, the accuracy of these methods deteriorates substantially when dispersal occurs at global scales because spherical Brownian motion (SBM) differs from planar Brownian motion. No statistical method exists for estimating SBM diffusion coefficients from a given timetree and tip coordinates, and no method exists for simulating SBM along a given timetree. Here, I present new methods for simulating SBM along a given timetree, and for estimating SBM diffusivity from a given timetree and tip coordinates using a modification of Felsenstein's independent contrasts and maximum likelihood. My simulation and fitting methods can accommodate arbitrary time-dependent diffusivities and scale efficiently to trees with millions of tips, thus enabling new analyses even in cases where planar BM would be a sufficient approximation. I demonstrate these methods using a timetree of marine and terrestrial Cyanobacterial genomes, as well as timetrees of two globally circulating Influenza B clades. My methods are implemented in the R package "castor." [Independent contrasts; phylogenetic; random walk; simulation; spherical Brownian motion.]

Phylogeographic methods are widely used for reconstructing the dispersal dynamics of plant and animal species (Lemmon and Lemmon 2008; Landis *et al.* 2013), infectious diseases (Lemey *et al.* 2009, 2010; Pybus *et al.* 2012; Faria *et al.* 2012), and culture (Bouckaert *et al.* 2012; Currie *et al.* 2013), especially when the dispersal process cannot be directly observed (e.g., because it happened in the past) or when the detection rate is unknown (as is the case in many infectious diseases). The most common aims of such studies is to estimate rates of dispersal and/or to estimate the likely geographic locations of ancestral nodes (Lemmon and Lemmon 2008; Bouckaert and Cartwright 2016). Knowing the dispersal rates of invasive species can help prioritize management actions in the case of budgetary constraints, or weigh proactive management costs versus risk of dispersal. If a species has already spread substantially, then dispersal rate estimates could give insight into the possible mechanisms of dispersal or rule out possible dispersal mechanisms. Very fast dispersal of an invasive species may, for example, rule out certain animals as potential dispersal vectors, while very slow dispersal may, for example, rule out human movement as the dispersal mechanism. Comparative analyses of dispersal rates between taxa could also yield insight into how morphology, behavior, or environmental conditions affect dispersal. Further, a reconstruction of historical dispersal rates of invasive species or infectious diseases over time can yield insight into the effects of past policies, as well as insight into potential evolutionary changes (e.g., towards greater or lower transmissivity). Knowledge of dispersal rates is also important for properly interpreting biogeographic diversity patterns—for example, a highly debated question in microbial ecology is whether patterns of geographic restriction are fully attributable to environmental filtering or whether dispersal limitation is an important factor (Whitfield 2005; De Wit and Bouvier 2006; van der Gast 2015).

A widely used mathematical model of dispersal is planar Brownian motion (BM), which describes the dispersal of each lineage as an independent continuous-time continuous-space diffusion process or "random walk" on a 2D plane (Bloomquist *et al.* 2010; Lemey *et al.* 2010; Faria *et al.* 2011, 2012), similarly to the evolution of a continuous bivariate trait along a phylogeny. BM models offer a powerful alternative to phylogeographic models where species ranges are treated as discrete traits (i.e., geographic locations are grouped into discrete regions), especially when the number of possible geographic regions is large and computational bottlenecks become a serious issue (Ree and Smith 2008; Goldberg *et al.* 2011; Landis *et al.* 2013; Matzke 2014). The planar BM process is characterized by a single numeric parameter known as diffusion coefficient (henceforth simply "diffusivity" and denoted D), which in principle could depend on time although in most practical cases is assumed to be constant for simplicity. Given a time-calibrated phylogeny (henceforth "timetree") and geographic locations of its tips, the diffusivity of planar BM can be efficiently estimated using Felsenstein's independent contrasts (Felsenstein 1985; Freckleton 2012) or generalized linear models (Tung Ho and Ané 2014), both of which are equivalent to a maximum likelihood estimation.

When the examined geographical range is large (e.g., several thousands of km), planar BM can no longer be used to model the full dispersal process even if locally (i.e., at small spatial and temporal scales) the dispersal process is adequately described by planar BM. This is because geographic coordinates (latitude and longitude) cannot be mapped to a single 2D Cartesian

system while preserving distances, and because the transition probabilities of diffusion on a sphere differ from those of planar BM. This limitation is particularly relevant for viral epidemics, many invasive species, human migration, and language. In these cases, a modification of the process is needed that accounts for Earth's spherical geometry, known as spherical Brownian motion (SBM) (Perrin 1928; Brillinger 2012); however, tools able to fit SBM models are rare and tools for simulating phylogeographic data under the SBM model are non-existent. Efficient simulations of phylogeographic models are needed for evaluating the accuracy of estimation tools, for obtaining confidence intervals via bootstraps, for hypothesis testing and for approximate Bayesian computations (Janzen *et al.* 2015). To my knowledge, the only existing phylogeographic tool able to handle SBM is a package developed by Bouckaert and Cartwright (2016) for estimating ancestral geographical locations within the BEAST software environment (Bouckaert *et al.* 2014), which uses an approximation of the transition probability density by Ghosh *et al.* (2012) as well as an approximation of the likelihood of the tree (Eq. 2 in Bouckaert and Cartwright 2016). The approximation used for the transition density, however, is inaccurate when geographical distances between sister tips span global scales, that is, if dispersal is fast compared to lineage branching rates in the timetree, and the approximation used for the likelihood was not justified and remains poorly understood. Further, the tool provided by Bouckaert and Cartwright (2016) is part of BEAST's molecular sequence analysis pipeline, and is thus not suitable for situations where a timetree has already been obtained by other means.

This article makes the following contributions: first, I present a method for efficiently simulating SBM models along any given timetree. Second, I present a method for efficiently and accurately calculating the likelihood of a phylogeographic data set (timetree + tip coordinates) under an SBM model that is accurate both for slow as well as fast dispersal. Third, I present methods for estimating the SBM diffusivity via maximum likelihood. Fourth, I present a method for evaluating the extent to which a fitted SBM model is a good description of the dispersal process underlying a given data set, based on a data-consistency approach similar to Lindholm *et al.* (2019). In contrast to existing tools, the presented methods can accommodate models with arbitrary time-variable diffusivities and scale well to large trees with millions of tips, thus enabling analysis of modern massive phylogenetic data sets (Louca *et al.* 2018; Smith and Brown 2018). My methods are implemented in the package `castor` (Louca and Doebeli 2018) for the computational environment R (R Core Team, 2019), and can thus be easily integrated into other workflows. I examine the accuracy of these methods through simulations and demonstrate their use using a timetree of Cyanobacterial genomes, as well as two Influenza B hemagglutinin timetrees sampled between 1987 and 2015 (Langat *et al.* 2017). The Cyanobacterial example, in particular, illustrates how using planar BM instead of SBM to model global dispersal can lead to different biological conclusions.

Mathematical Formalism

Transition Density and Cumulative Distribution of the SBM

The methods presented here consider an SBM process with diffusivity D on a sphere with radius r. I assume that diffusion is isotropic and homogeneous, that is, no particular direction is favored and diffusivity is the same at all locations. For now I will also assume that D does not depend on time, although I later discuss how a time-dependent D can be accounted for. Note that locally (i.e., over short time scales and small distances) the diffusion process resembles a bivariate planar BM process with diffusivity D, that is, after some short time step t the expected squared geographical distance traversed is $4tD$. Note that D is related to the infinitesimal variance σ^2 as $D=\sigma^2/2$ (Lange 2010). The probability density for the central angle $\omega \in [0,\pi]$ between the initial and final point reached by a diffusing particle after some time step t is given by:

$$\rho(\omega;t) = \frac{\sin(\omega)}{2}\sum_{n=0}^{\infty}(2n+1)P_n(\cos\omega)e^{-n(n+1)tD/r^2}, \quad (1)$$

where P_n is the Legendre polynomial of order n (Perrin 1928; Brillinger 2012). Observe that the probability density of the central angle ω (henceforth "transition density," see Fig. 1a) depends on the unitless quantity tD/r^2, henceforth denoted by β for brevity. When β is large (i.e., $\beta \gtrsim 0.1$), the series in Eq. (1) converges rapidly and can be well-approximated by a truncated version:

$$\rho(\omega;t) \approx \frac{\sin(\omega)}{2}\sum_{n=0}^{N}(2n+1)P_n(\cos\omega)e^{-n(n+1)\beta}, \quad (2)$$

where N is chosen sufficiently high to ensure that only terms below a certain error threshold are omitted. For small β (i.e., much smaller than 0.1), the convergence of the series is slow and its truncations exhibit strong oscillations over ω that are not present in the full series (i.e., the true density ρ). This makes the truncated series a bad choice for purposes of likelihood calculations in practice. In that case, an alternative approximation developed by Ghosh *et al.* (2012) may be used instead:

$$\rho(\omega;t) \approx \frac{\mathcal{N}(\beta)}{2\beta}\sqrt{\omega\sin(\omega)}e^{-\frac{\omega^2}{4\beta}}, \quad (3)$$

where $\mathcal{N}(\beta)$ is an appropriately chosen normalization factor. Note that $\mathcal{N}(\beta)$ depends on β and must be calculated numerically. This approximation formula is used in the phylogeographic analysis package by Bouckaert and Cartwright (2016), and while fairly accurate for small β, it becomes inaccurate when β is large (which is where the truncated series becomes accurate). To address this shortcoming, here I calculate

likelihoods using either the truncated series in Eq. (2) or the approximation in Eq. (3) depending on whether β is larger or smaller than 0.1, respectively.

To simulate an SBM model in forward time, the cumulative distribution function (CDF) of the central angle is needed. The CDF can be obtained by integrating the transition density in Eq. (1) over ω and using well-known properties of Legendre polynomials:

$$F(\omega;t)=\frac{1}{2}\sum_{n=0}^{\infty}e^{-n(n+1)\beta}[P_{n-1}(\cos\omega)-P_{n+1}(\cos\omega)]. \quad (4)$$

As with the transition density, for $\beta \gtrsim 0.1$ the series in Eq. (4) can be well-approximated by truncation, whereas for much smaller β the convergence of the series is slow and an accurate truncation would require an impractically large number of terms. Hence, for small β I construct the CDF using the approximate density in Eq. (3). Because an analytical form for the integral of Eq. (3) is not known, and to avoid costly numerical integrations, I approximate the formula in Eq. (3) using a Taylor expansion of $\sqrt{\omega\sin(\omega)}$ in ω:

$$\sqrt{\omega\sin(\omega)}\approx\omega-\frac{\omega^3}{12}+\frac{\omega^5}{1440}-\frac{\omega^7}{24192}+\frac{67\omega^9}{29030400}. \quad (5)$$

Inserting this Taylor expansion into Eq. (3) and integrating over ω yields the following approximate CDF for small β:

$$\begin{aligned}F(\omega;t)\approx&\frac{\mathcal{N}(\beta)}{1814400}\cdot\\&\Bigg[-384\cdot(-4725+1575\beta-105\beta^2+75\beta^3+67\beta^4)\\&+5760e^{-\frac{\omega^2}{4\beta}}(-315+105\beta-7\beta^2+5\beta^3)\\&+15e^{-\frac{\omega^2}{4\beta}}\left[96(105-7\beta+5\beta^2)\omega^2+(-84+60\beta)\omega^4+5\omega^6\right]\\&+1072e^{-\frac{\omega^2}{4\beta}}\left(24\beta^4+6\omega^2\beta^3+\frac{3\omega^4\beta^2}{4}+\frac{\omega^6\beta}{16}+\frac{\omega^8}{256}\right)\Bigg].\end{aligned} \quad (6)$$

This approximation is accurate for $\beta<0.1$ because in that case $e^{-\frac{\omega^2}{4\beta}}$ decays rapidly for increasing ω (in other words, large ω are rare) and hence any errors of the Taylor expansion in Eq. (5) due to large ω are negligible when multiplied by $e^{-\frac{\omega^2}{4\beta}}$ (Appendix Fig. 2).

Likelihood Calculations from Independent Contrasts

In order to construct a likelihood function that can be used for model fitting, I extract independent realizations of the diffusion process from the timetree and tip coordinates as follows. Because the diffusion process is isotropic and homogeneous, the central angle between the locations of any two tips is a realization of the diffusion process with $\beta=tD/r^2$, where t is the length of the shortest phylogenetic path connecting the two tips, also known as their patristic distance (see Appendix 1 for explanation). For any other pair of tips, whose shortest connecting path does not overlap with the shortest connecting path of the first tip pair, the central angle is independently distributed from that of the first tip pair. Thus, if we have N tip pairs whose shortest connecting paths are all disjoint, we have at hand N independent realizations of the diffusion process (Fig. 1b). Denoting by $t_1,..,t_N$ their respective patristic distances and by $\omega_1,..,\omega_N$ their central angles, one can estimate the diffusivity via maximum likelihood, that is, by numerically maximizing the expression:

$$L=\prod_{n=1}^{N}\rho(\omega_n;t_n) \quad (7)$$

through choice of D. This idea is analogous to the independent contrasts for a numerical trait evolving according to 1D or multi-dimensional planar BM, where independent realizations (or "contrasts") are obtained by subtracting the trait values of tip pairs (Felsenstein, 1985). The difference is that instead of simply subtracting the coordinates of two tips, here the great-circle distance (or equivalently, the central angle) is calculated between the tips. For any bifurcating tree with M tips one can extract $\lfloor M/2\rfloor$ such independent contrasts (example in Fig. 1b; proof and explicit algorithm in Appendix 3). The likelihood in Eq. (7) is the joint probability density of observing the geographic distances between the tips in the N extracted independent contrasts.

Note that, in contrast to planar BM, it is unknown whether and how additional independent realizations of the SBM process could be efficiently extracted from the tree, for example, by comparing nodes as done by Felsenstein (1985) for planar BM, because one cannot simply "add" or "subtract" geographic locations similarly to planar coordinates. Consequently, here only about half the number of independent contrasts are used compared to Felsenstein (1985). Also note that in the present implementation of the likelihood I consider the transition density for the transformed variable $x=\cos(\omega)$ rather than ω, which is equal to $\tilde{\rho}(x;t)=\rho(\arccos(x);t)/\sin(\arccos(x))$. The reason is that $\rho(0;t)=0$ when $t>0$, which can cause issues during maximum likelihood estimation when multiple tips have been sampled from the same geographic location (in contrast, $\tilde{\rho}(1;t)>0$).

Simulating SBM

Simulation of SBM along a timetree proceeds from root to tips, generating random transitions on the sphere along each edge. Random transitions from an ancestral node to its child along an edge of length t are generated by drawing a random central angle ω according to the SBM transition density and choosing a random direction

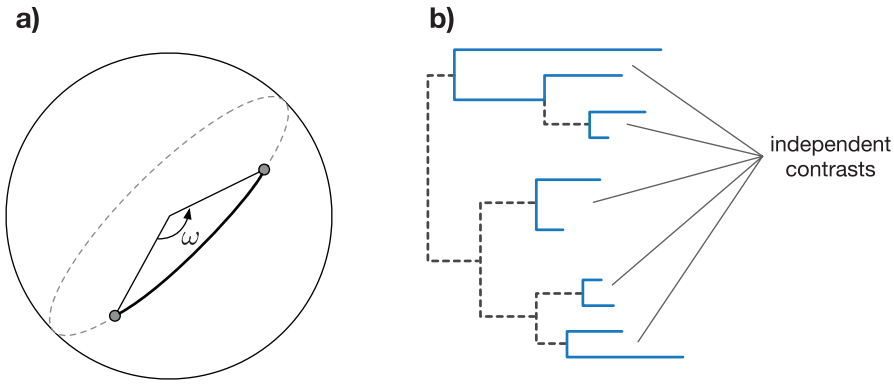


FIGURE 1. Spherical Brownian motion algorithm. a) Illustration of the central angle ω between two points on a sphere (small circles), traversed by a Spherical Brownian Motion during a given time step t. The probability distribution of ω depends solely on tD/r^2, where D is the diffusivity and r the sphere radius. b) Illustration of five independent contrasts (continuous line segments) representing independent realizations of SBM, extracted from a timetree.

uniformly within $\alpha \in [-\pi, \pi]$ (e.g., a random longitude if the ancestral node was located at the North pole). The central angle can be drawn using the cumulative distribution function, $F(\omega; t)$, by drawing a random variable u uniformly from $[0,1]$ and then solving $F(\omega; t) = u$ numerically for ω. Given ω and α, and given the latitude $\vartheta_a \in [-\pi/2, \pi/2]$ and longitude $\varphi_a \in [-\pi, \pi]$ of the ancestral node, the latitude ϑ_c and longitude φ_c of the child can be calculated as follows. Define the vector:

$$x = \begin{pmatrix} \sin(\omega)\cos(\alpha) \\ \sin(\omega)\sin(\alpha) \\ \cos(\omega) \end{pmatrix}. \quad (8)$$

Note that x would be the location of the child in a 3D Cartesian coordinate system if the ancestral node was located at the North pole. Let $\mathbb{A}$ be the unique 3×3 rotation matrix that takes the North pole to the ancestral node, and calculate $y = \mathbb{A} \cdot x$. Then y is the location of the child in 3D Cartesian coordinates, adjusted for the true location of the ancestral node. The corresponding latitude and longitude are:

$$\vartheta_c = \arcsin(y_3), \quad \varphi_c = \arccos(y_1/\cos(\vartheta_c)) \cdot \mathrm{sgn}(y_2), \quad (9)$$

where y_1, y_2, y_3 are the three components of y and sgn() is the sign function (note that if $\vartheta_c = \pm\frac{\pi}{2}$, i.e., the child landed at one of the poles, the longitude φ_c is irrelevant and can be arbitrarily set to 0).

Time-Dependent Diffusivity

So far I assumed that D is constant over time. However, the transition density in Eq. (1) and the presented approximations can be easily generalized to the case where D varies over time. This can be done by replacing any occurrence of tD with the integral $\int_{t_1}^{t_2} D(t)\, dt$, where t_1 and t_2 are the start and end time of the transition ($t = t_2 - t_1$) (explanation in Appendix 2). Hence, the probability density for the central angle ω traversed by a diffusing particle from time t_1 to time t_2 is:

$$\rho(\omega; t_1, t_2) = \frac{\sin(\omega)}{2} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\omega) \exp\left[-n(n+1) \frac{1}{r^2} \int_{t_1}^{t_2} D(t)\, dt \right]. \quad (10)$$

To simulate a time-dependent SBM model one thus simply needs to calculate the antiderivative of D over time, $\tilde{D}(t) = \int_0^t D(s) ds$, and for any edge generate random transitions according to the standard SBM density, using $\tilde{D}(t_c) - \tilde{D}(t_a)$ instead of $(t_c - t_a)D$, where t_a and t_c are the times of the ancestral and child node, respectively.

To calculate the likelihood of a set of independent contrasts one can proceed similarly, as follows. For any two tips defining an independent contrast, let $t_{c,1}$ and $t_{c,2}$ be the times of those tips (i.e., their distances from the root) and let t_a be the time of their most recent common ancestor. Then the probability density of their central angle can be obtained using the transition density of the time-independent SBM (Eq. 1) by replacing tD with the expression:

$$\tilde{D}(t_{c,1}) - \tilde{D}(t_a) + \tilde{D}(t_{c,2}) - \tilde{D}(t_a). \quad (11)$$

See Appendix 1 for explanation. In the present implementation, the antiderivative $\tilde{D}$ is calculated only once at the beginning of a simulation or likelihood calculation and stored internally as a piecewise polynomial function; thus SBM models with time-dependent D do not require substantially more computation than SBM models with constant D. I mention that methods for examining time-dependent diffusivities in the context of planar BM have been developed previously (Eastman *et al.* 2011; Stack *et al.* 2011).

Scalable Algorithms for Simulating and Fitting SBM Models

Based on the mathematical approaches outlined above, I implemented efficient computer code for simulating SBM models with arbitrary time-dependent diffusivity on any given timetree, that is, generating random geographical coordinates for the tree's tips and nodes. The root's coordinates, that is, the start location, can be explicitly specified or randomly drawn from the equilibrium distribution of the SBM process (i.e., the uniform density on the sphere). The algorithm scales well to large trees, with computation time scaling asymptotically linearly with the number of edges. For example, simulating an SBM model with constant diffusivity takes about 0.005 s on a tree with 1000 tips and about 5 s on a tree with 1 million tips, when tested on a 2015 MacBook Pro (Fig. 2a). Simulating a model with time-dependent diffusivity does not introduce a noticeable overhead at these tree sizes, since the antiderivative of D only needs to be pre-computed once prior to the simulation and typically only takes a few milliseconds thanks to internally used symbolic representations (previously described in Louca 2020). The simulation method is available as function `simulate_sbm` in the R package `castor` v1.6.2 (Louca and Doebeli 2018).

Using the above approaches, I also implemented efficient code for calculating the likelihood of phylogeographic data (timetree+tip coordinates) under SBM models with arbitrary time-dependent diffusivity, as well as for fitting such models via maximum likelihood. In the case of time-dependent models, the user can provide an arbitrary functional form for D, parameterized by an arbitrary number of free parameters to be estimated from the data. For example, if D is suspected to vary approximately exponentially with time, $D(t)=\alpha e^{\beta t}$, my methods can be used to fit the parameters α and β. Alternatively, if D is suspected to be proportional to some continuous environmental variable (e.g., temperature T), $D(t)=\alpha T(t)$, my methods can be used to fit the free parameter α. The implemented methods scale well to large trees, and computation time scales asymptotically linearly with the number of tips. For example, fitting a constant-diffusivity SBM model (including 100 parametric bootstraps) takes about 5 s on a tree with 1000 tips and about 260 s on a tree with 1 million tips (Fig. 2b). My methods are also much faster than existing R packages for fitting planar BM (by multiple orders of magnitude in the case of large trees), that is, they present a performance improvement even in cases where planar BM would be an adequate approximation (Appendix Fig. 7).

The availability of efficient methods for simulating and fitting SBM models opens the door to parametric bootstrapping for calculating confidence intervals of parameter estimates, to the assessment of model adequacy through posterior predictive simulations (Brown and Thomson 2018) (see "data consistency" approach described below), and to statistical hypothesis testing. For example, one could fit a separate SBM model to each of two taxa and then use simulations to examine the statistical significance of their fitted diffusivities, that is, estimate the probability that the fitted diffusivities would be as different as observed, if the two taxa actually had identical diffusivities (see Cyanobacterial example below). For convenience, I have implemented this binary diffusivity comparison, as well as parametric bootstrapping, in `castor`. In the case of linear models, that is, where $D(t)=\alpha+\beta t$, one can also estimate the statistical significance of the slope, that is, the probability that $|\beta|$ would be at least as large as observed if the data had been generated by a constant-D model (where the constant D used is fitted via maximum likelihood). Hence, the presented methods can be used to test hypotheses regarding a decline or increase of geographic dispersal rates over time. This question is of particular interest in epidemiology, where the efficacy of implemented containment policies often needs to be determined in retrospect.

I mention that methods for examining temporal variation in the diffusivity using planar BM have been developed previously (Eastman *et al.* 2011; Stack *et al.* 2011; Clavel *et al.* 2015). None of these methods, however, can fit arbitrary functional forms of D over time, and instead require that discrete "selective regimes" with distinct diffusivities be specified beforehand. In contrast, the methods presented here can fit arbitrary temporal profiles of D with an arbitrary number of free parameters. This allows, for example, modeling potential effects of continuously varying environmental variables on dispersal rates over time (as illustrated above).

To assess the accuracy of my methods and to examine the improvements of SBM over planar BM, I performed two types of simulations of SBM models on timetrees and investigated the ability to correctly reconstruct the true diffusivities. In the first type of simulations, I generated multiple timetrees using a Yule model (i.e., with constant speciation rate λ and no extinction) and then simulated SBM models with constant diffusivity on those trees. It is worth noting that in such simple scenarios the identifiability of the diffusivity is largely determined by the magnitude of D compared to the radius of the sphere and the typical time scales separating sister tips in the timetree. In other words, if τ is the average temporal scale of independent contrasts ($\tau \approx 1.37/\lambda$ in the case of a Yule tree) and D is the diffusivity, then the identifiability of D (given a certain number of tips) is largely determined by the unitless ratio $D/(r^2/\tau)$, since one can always change distance and time units so that r and τ are 1. The related quantity $\delta=\sqrt{\pi D/(r^2/\tau)}$ (henceforth "diffusion scale") is the expected Euclidean distance that a lineage would traverse after time τ, counted in sphere radii r, if the diffusion process took place on a 2D plane. Hence, a larger diffusion scale δ means that on average compared tips tend to be geographically further apart. My simulations of SBM models with constant D on Yule trees showed that, when tip locations are exactly known, the accuracy of the estimated diffusivity increases for

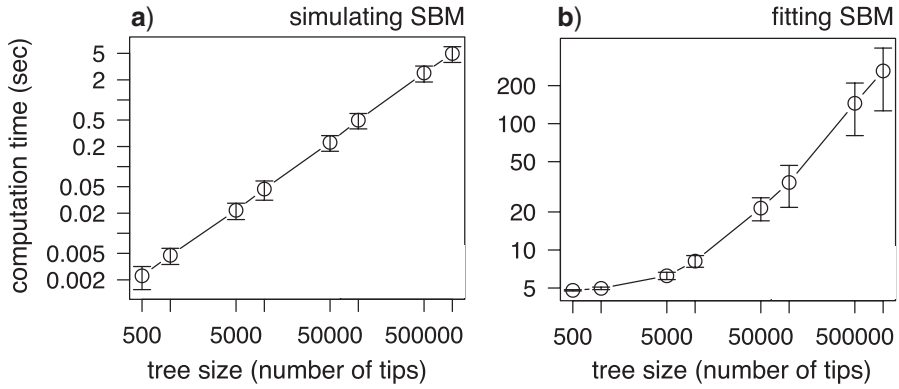


FIGURE 2. Computational efficiency. Benchmarks of computation time for simulating and fitting SBM models to phylogenies of varying sizes. a) Simulating an SBM model with time-independent D along a timetree. b) Fitting a time-independent D via maximum likelihood, including 100 parametric bootstraps. In both figures, circles show mean computation times and vertical error bars extend one standard deviation above and below the mean, calculated based on 100 independent repeats. See Fig. 7 for a comparison to other R packages in the case of planar BM.

larger trees and decreases for larger diffusivities (or equivalently, larger diffusion scales) (Fig. 3a–f and Appendix Fig. 4a–c). For example, for a true $D = 0.01\ r^2/\tau$ (i.e., $\delta = 0.17$) and a phylogeny of 1000 tips, the width of the 95% confidence interval for the estimated D is about $0.0020\ r^2/\tau$, for a true $D = 0.1\ r^2/\tau$ (i.e., $\delta = 0.56$) the 95% confidence interval has a width of about $0.022\ r^2/\tau$, and for a true $D = 1\ r^2/\tau$ (i.e., $\delta = 1.8$) the 95% confidence interval has a width of $\sim 0.35\ r^2/\tau$ (Fig. 3d–f). For smaller trees (e.g., with 10 or 100 tips) the confidence intervals naturally tend to increase (Fig. 3a–c and Appendix Fig. 4a–c). At the scales examined here the estimated diffusivity does not appear to be substantially biased, as becomes evident from the strong agreement between the median estimated and true D (Fig. 3c,f). In contrast, when using an approximation for the likelihood under the assumption that pairwise geographic tip distances are generated by a planar BM, the diffusivity tends to be underestimated especially for larger D (Fig. 3b,e). Alternatively, when fitting a planar BM model directly using latitudes and longitudes as Cartesian coordinates (Fig. 3a,d), diffusivity estimates are either positively biased (for intermediate diffusivities, $D \sim 0.1$–$1\ r^2/\tau$) or negatively biased (for high diffusivities, $D \gtrsim 10\ r^2/\tau$). This highlights the importance of accounting for spherical geometry when reconstructing global or continental-scale dispersal processes. For much higher diffusivities (i.e., $\delta \gg 1$), the accuracy of the estimated diffusivity deteriorates rapidly and much larger trees are needed for reasonable estimates. The reason is that under very fast diffusion the distribution of individual transitions over the time scale τ approaches the uniform distribution on the sphere (i.e., every point is approximately equally likely to be reached after time τ), and hence the distribution of tip coordinates contains little usable information for fitting the model. In conclusion, it should be remembered that reconstructive accuracy ultimately depends on a variety of factors, including tree size, phylogenetic distances between sister tips and diffusivity; consequently, the reliability of diffusivity estimates and biological inferences should be assessed on a case-by-case basis, for example in terms of statistical significances and confidence intervals.

In practice, tip coordinates may not be exactly known. This situation is common in epidemiological data sets, where sampling locations may only be known at the city or county scale (e.g., due to rules regarding patient privacy). To examine the effects of inaccurate tip locations, I repeated the above analysis using randomly varied tip coordinates, that is, after replacing each tip's geographic location with a random point within a certain "error radius" ε from its true location (Fig. 3g–l and Appendix Figs. 3 and 4). I considered various error radii, ranging from $\varepsilon = 0.016 \times r$ (i.e., 100 km) up to $\varepsilon = 0.16 \times r$ (i.e., 1000 km). These analyses revealed that the resulting error in the estimated diffusivity is fundamentally different from the error stemming from a planar approximation. Indeed, inaccurate tip locations mainly led to a strong positive bias in diffusivity estimates when the true diffusivity was low, that is, when the diffusion scale δ was comparable to or smaller than ε. For larger diffusion scales (i.e., $\delta \gg \varepsilon$), the resulting bias was negligible. This pattern is not too surprising, since errors in geographic location are expected to matter little if they are much smaller than the geographic distances between compared tips.

In the second type of simulations, I generated timetrees of various sizes (10–10,000 tips) using a birth–death model with constant speciation rate and an extinction rate equal to 90% of the speciation rate, roughly resembling typical diversification scenarios (Marshall, 2017), and simulated SBM models with diffusivities that depend linearly on time (either increasing, decreasing, or constant). I then fitted the two free parameters of an SBM model with linearly varying diffusivity to the simulated data, calculated confidence intervals using parametric bootstrapping, estimated the statistical significance of the slopes and compared the results to the true diffusivity profile over time (examples

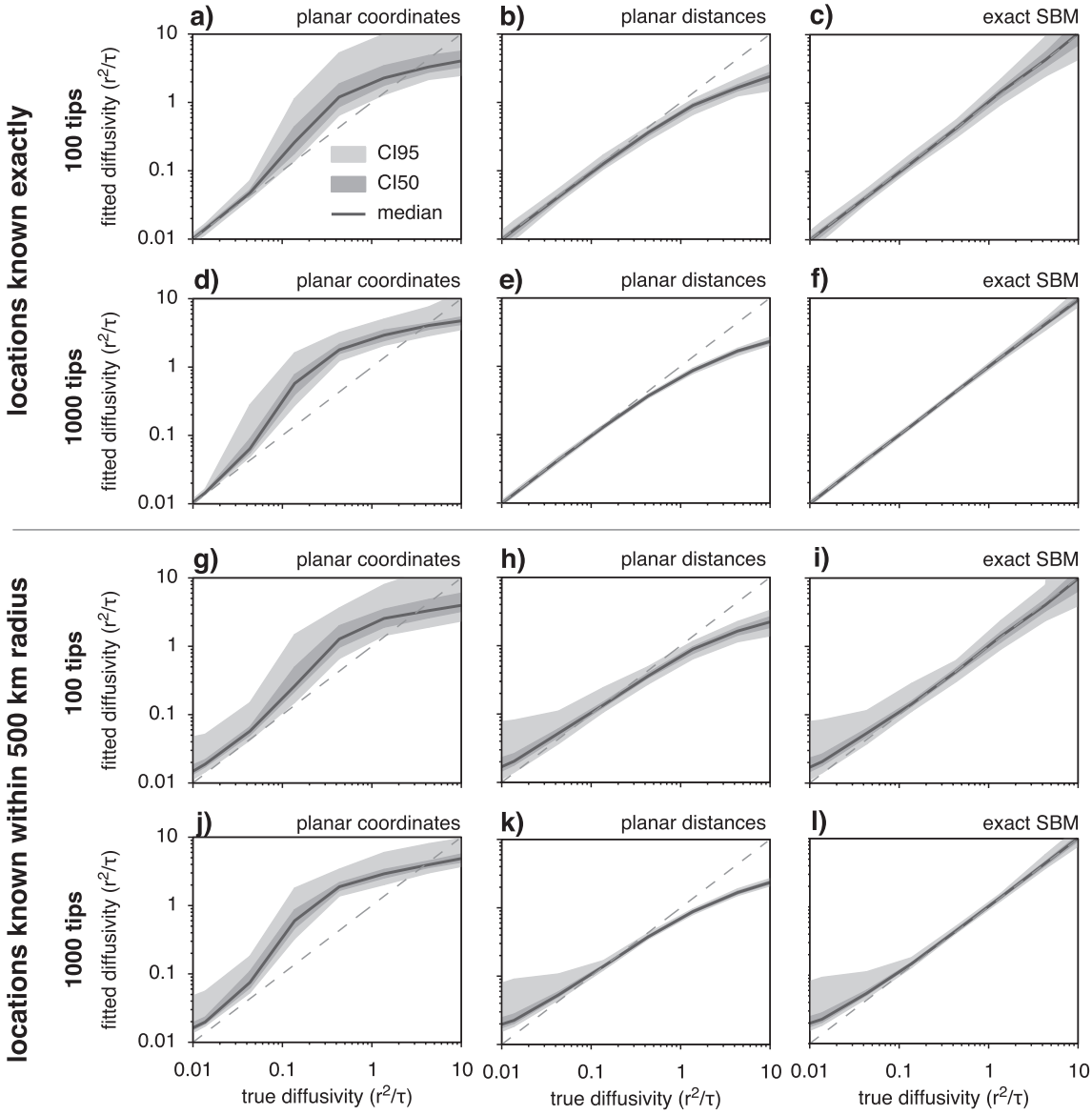


FIGURE 3. Evaluation of accuracy. Diffusivities (D) fitted to simulated phylogeographic data (timetrees generated by a Yule model, tip coordinates generated by an SBM model with constant diffusivity). Each figure shows the distribution of fitted diffusivities across a large number of simulations and for various true diffusivities. Horizontal axis: true diffusivity used in the simulations. Vertical axis: maximum likelihood-fitted diffusivity. Black curves show median fitted diffusivities. Shaded areas show 95% and 50% percentiles of fitted diffusivities. Dashed diagonals are shown for reference. a–c) Considering trees with 100 tips, and with exactly known tip locations. Diffusivities are either estimated (a) using planar BM, considering latitudes and longitudes as Cartesian coordinates, or (b) by correctly calculating geodesic distances between tips and fitting D under the assumption that these distances are generated by planar BM, or c) by fitting an SBM model. d–f) Similar to (a–c), but considering trees with 1000 tips. g–l): Similar to (a–f), but with inaccurately known tip locations; each tip location was replaced by a random location within a 500 km radius (i.e., $\lesssim 0.078 \times r$) from its true location. All diffusivities are expressed in terms of squared sphere radii per τ, where τ is the average temporal scale of independent contrasts. See Appendix Fig. 3 for alternative location error sizes.

in Fig. 4 and Appendix Figs. 5 and 6). I found that in about 85–90% of cases the 95% confidence interval contained the full true diffusivity profile over time. In nearly all cases (99–100%) where the fitted slope was estimated to be statistically significant, the fitted diffusivity profile had the same trend over time (i.e., positive or negative) as the true diffusivity profile, although I mention that for very small trees (10 tips) only about 2% of trees yielded statistically significant slopes. The width of the confidence intervals generally decreased towards later time points (i.e., closer to the present); this is not surprising, since independent contrasts used for fitting are constructed exclusively from pairs of tips and in the simulated trees most tips are located near the present. In order to accurately reconstruct diffusivities at older times, samples (i.e., tips and their coordinates) from

these time periods generally need to be included in the tree. This requirement could be satisfied for example in cases where pathogen strains are sampled continuously over the course of an epidemic (Stadler *et al.*, 2013), thus yielding a timetree whose tips cover multiple time points rather than just the present.

The above examples confirm the accuracy of the implemented simulation procedures and likelihood calculations, and highlight the issues associated with using planar BM to model large-scale dispersal. These examples do not, however, imply that the diffusivity is always fully identifiable as a function of time no matter how complex its true profile is; more systematic examinations of various diffusivity scenarios are needed to understand what scenarios (e.g., which functional forms) are identifiable and under what a priori constraints, which is beyond the scope of this article. Indeed, recent work on birth–death models has shown that just because one can accurately calculate the likelihood of a phylodynamic model this does not imply that all parameters of the model are fully identifiable (Louca and Pennell, 2020).

Caveats

The methods presented here are subject to a number of caveats, largely stemming from a potential inadequacy of the SBM model and similar to common caveats encountered in other trait-evolution models. First, it is assumed that dispersal is diffusive and, crucially, isotropic and homogeneous across space. In some cases, however, dispersal may occur at different rates in various regions, or may not occur at all in some regions (e.g., no dispersal across the ocean in the case of most terrestrial organisms). Further, dispersal may occur mainly along certain routes, for example, along shipping routes or animal migration routes. If the diffusivity is suspected to be non-isotropic or non-homogeneous, the estimated diffusivity should be considered an "effective" diffusivity, averaged across the globe. A more realistic model would of course be one that takes into account variation in dispersal rates between regions or sub-clades as well as geographic barriers; similar frameworks have been proposed previously for planar BM (O'Meara *et al.* 2006; Eastman *et al.* 2011). However, this would come at the cost of a much larger parameter space (potentially subject to serious identifiability issues) and a substantially increased computational cost both for simulation and fitting. Hence, whether diffusion is a reasonable model for understanding a clade's dispersal dynamics, or whether more complex models need to be deployed, should be judged on a case-by-case basis. To facilitate this decision process, I have implemented a criterion similar to the "data consistency" approach introduced by Lindholm *et al.* (2019), which quantifies the deviation of a data set from expectations based on the fitted model. Specifically, I consider the deviation of the log-likelihood of the observed data, denoted $\hat{Z}$, from the expected log-likelihood of hypothetical data generated by the model, denoted $\mathbb{E}\{Z\}$ (where Z is a random variable, representing the log-likelihood of data generated by the fitted model, under that same model). I define the "consistency" of the data as the probability Q that $|Z-\mathbb{E}\{Z\}|$ would be larger than $|\hat{Z}-\mathbb{E}\{Z\}|$. Hence, an improbably strong deviation $|\hat{Z}-\mathbb{E}\{Z\}|$, that is, a low consistency (e.g., $Q<0.05$), suggests that the fitted model should be rejected as a model for the observed data. Reciprocally, a log-likelihood Z close to the expectation $\mathbb{E}\{Z\}$, that is, a high consistency ($Q>0.05$), means that the data are at least consistent with the model's distribution of log-likelihoods.

Another caveat is that tips are assumed to be sampled randomly across the globe, that is, the probability of sampling a tip from the full phylogeny does not depend on geographic location. In reality, sampling may be geographically heterogeneous, for example in the case of infectious diseases sampling may be strongly determined by the medical and scientific infrastructure available in a country. In the extreme case where sampling is only performed in a small region of the world that does not cover the full dispersal range, the presented methods will tend to underestimate the true diffusivity. This issue may be mitigated by only considering a sub-clade that is known to have not (yet) dispersed past the sampled region.

It should also be noted that, because the presented likelihood function only uses independent contrasts between tips and not between internal nodes, reliably estimating the diffusivity at older times (i.e., close to the root) requires that the timetree include extinct tips or tips sampled at older times (e.g., as is often the case in viral phylogenies). For example, if the timetree only consists of extant species, then the diffusivity can at most be estimated near present day (i.e., over a recent time range in the order of τ), unless it is assumed that D was constant throughout time.

Application to Cyanobacterial Dispersal

A fundamental question in microbial ecology is whether microbial dispersal rates are sufficiently slow to cause geographic endemism and other biogeographic patterns (Finlay and Clarke 1999; Whitaker *et al.* 2003; Whitfield 2005; De Wit and Bouvier 2006; van der Gast 2015). A heavily debated hypothesis, for example, is that "all microbes are everywhere" and that "the environment selects" (De Wit and Bouvier, 2006). Past global surveys suggested that marine bacterial species are indeed distributed globally, in other words, the time it takes for a marine bacterial lineage to disperse around the world is shorter than the typical time scales involved in bacterial speciation (Gibbons *et al.* 2013; Gonnella *et al.* 2016). Surveys strongly supporting or refuting a similar conclusion for terrestrial microorganisms are currently lacking. While for free-living terrestrial microorganisms

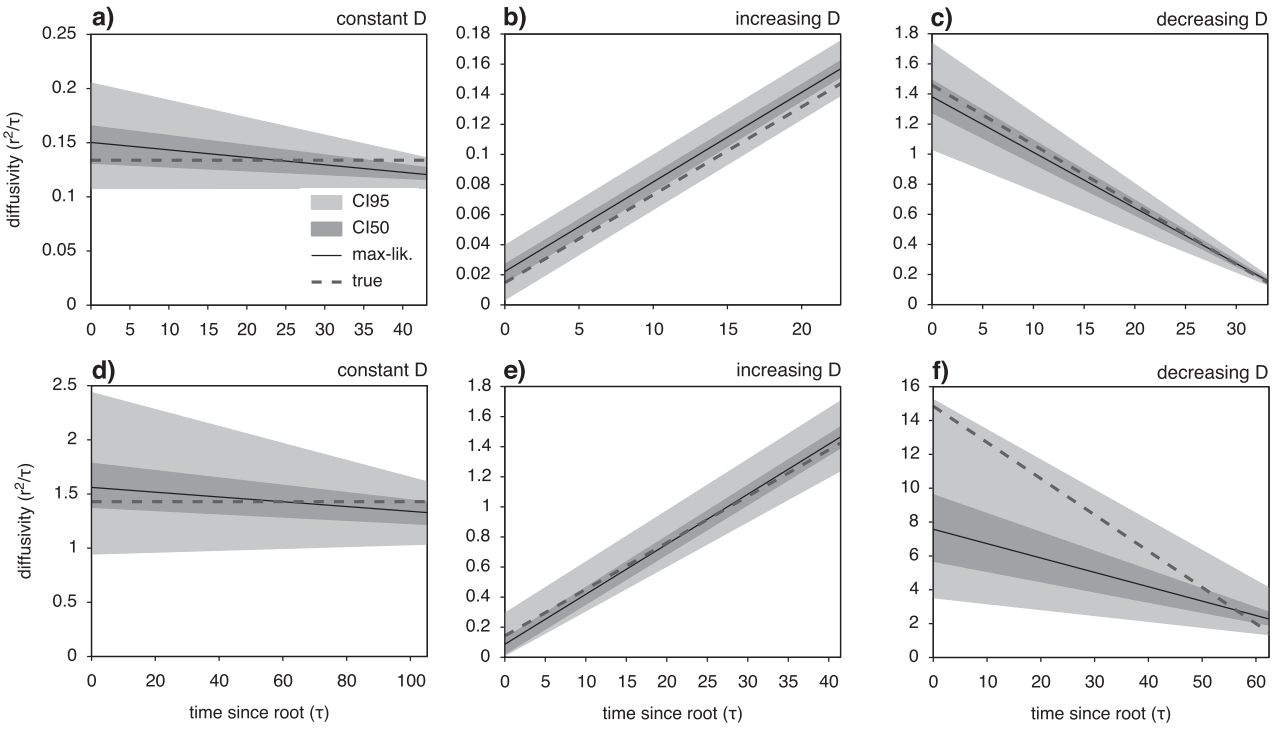


FIGURE 4. Fitting SBM models with linearly varying diffusivity. Diffusivities fitted to simulated phylogeographic data (1000 tips) simulated under an SBM model. In both the simulation and the fitted models the diffusivity was assumed to vary linearly over time. The timetrees were generated using a birth–death model with extinction rate equal to 90% of the speciation rate, and both extant and extinct lineages were kept. Geographic coordinates were either simulated with a constant diffusivity (i.e., zero slope, a and d), a linearly increasing diffusivity (b and e) or a linearly decreasing diffusivity (c and f). Examples in the bottom row consider a roughly 10-fold higher diffusivity than in the top row. The diffusivity is measured in units of r^2/τ, where r is the sphere radius and τ is the average temporal scale of the independent contrasts. Dashed lines show maximum likelihood fitted diffusivities, shaded areas show 50% and 95% confidence intervals, and the solid lines show the true diffusivities. For analogous simulations using trees with 100 or 10,000 tips, see Appendix Figs. 5 and 6 , respectively.

dispersal is suspected to be generally slower than for marine microorganisms, the actual dispersal rates are largely unknown in either case.

Here, I demonstrate how my methods can be used to examine this question, using a time-calibrated phylogeny of 371 georeferenced Cyanobacterial genomes obtained from GenBank (Clark *et al.* 2015). Specifically, I fitted an SBM model with constant diffusivity via maximum likelihood separately to marine and terrestrial Cyanobacteria. Only considering Cyanobacteria allows controlling broadly for taxonomy, while assessing the effects of a terrestrial versus marine lifestyle on dispersal rates. For marine Cyanobacteria I estimated a diffusivity of 535 $km^2 \times yr^{-1}$ (50% confidence interval 141–723 $km^2 \times yr^{-1}$), while for terrestrial Cyanobacteria I estimated a diffusivity of only 1.47 $km^2 \cdot yr^{-1}$ (50% confidence interval 1.27–1.66 $km^2 \cdot yr^{-1}$). The difference between the two diffusivities was highly statistically significant ($P = 0.0006$ on a linear axis, $P < 0.0001$ on a logarithmic axis). For both fitted models, the consistency was far above the significance threshold of 0.05 (marine: 0.96, terrestrial: 0.54), indicating that the data are—at least in terms of the expected log-likelihoods—consistent with the fitted models. These diffusivity estimates imply that within 1500 years a marine Cyanobacterial lineage is expected to traverse on average ~1582 km, while a terrestrial lineage is expected to only traverse on average ~83 km during that time. It is worth mentioning that the ocean's average circulation time is in the order of 1500 years (although strong variation exists between regions and depths) (Frank, 2002), which is roughly consistent with the estimated average marine cyanobacterial dispersal distances over that time interval. Further, the strong difference in diffusivities is consistent with the hypothesis that free-living terrestrial microorganisms tend to disperse much slower than marine microorganisms.

Note that if one were to use standard planar BM-based methods to estimate Cyanobacterial diffusivities, for example, as commonly used for clades with smaller dispersal ranges (Lemey *et al.* 2010; Faria *et al.* 2011; Monjane *et al.* 2011; Faria *et al.* 2012), one would partly obtain very different results. For example, fitting a bivariate BM model of continuous trait evolution using the R package `mvMORPH` (Clavel *et al.*, 2015), while treating latitudes and longitudes as two numerical traits

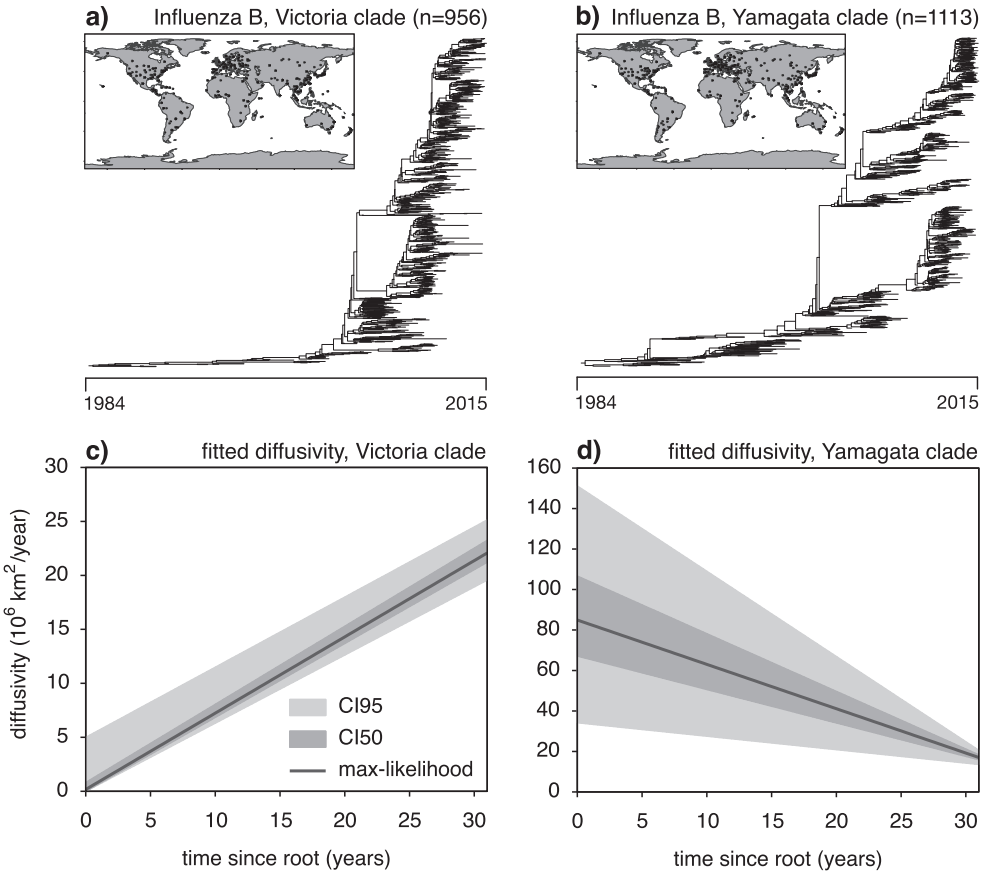


FIGURE 5. Analysis of Influenza B phylogeographic data. Top row: Hemagglutinin (HA) timetrees for two globally circulating Influenza B clades and strain locations (a: Victoria clade, b: Yamagata clade), used in this analysis. Timetrees were obtained from Langat *et al.* (2017). Bottom row: SBM diffusivities fitted to the two timetrees shown in the top row, assuming that the diffusivity varies linearly with time (black curves: maximum likelihood estimates, shaded areas: 50% and 95% confidence intervals). The statistical significances of the slopes are estimated to be 0.12 for the Victoria clade and 0.006 for the Yamagata clade.

with equal diffusivity, yields a diffusivity of 3.6 km$^2\cdot$ yr^{-1} for marine and 1.35 km$^2\cdot$yr^{-1} for terrestrial Cyanobacteria. Hence, while for the slowly dispersing terrestrial Cyanobacteria spherical and planar BM yield comparable diffusivity estimates, for marine Cyanobacteria the diffusivity estimates differ by a factor of $\sim$150, that is, by about two orders of magnitude. Not only does the planar diffusivity estimate contradict expectations based on typical ocean circulation rates, it also fails to reveal the fact that marine Cyanobacterial dispersal is much faster than terrestrial.

APPLICATION TO INFLUENZA B DISPERSAL

To further demonstrate the use of the presented methods, I also examined the dispersal rates of Influenza B over time. Influenza B viruses are an important human pathogen that repeatedly causes serious epidemics throughout the world. Influenza B viruses are grouped into two major antigenically and phylogenetically distinct clades, with reference strains "B/Victoria/2/87" (henceforth "Victoria clade") and "B/Yamagata/16/88" (henceforth "Yamagata clade"), which diverged before 1983 and since then co-circulate globally (Rota *et al.*, 1990; Yang *et al.*, 2012). Here, I separately examine the geographic diffusivity of these two clades using timetrees based on the hemagglutinin gene and obtained from Langat *et al.* (2017), comprising 956 (Victoria) and 1113 (Yamagata) strains sampled between the years 1987 and 2015 (note that I only consider a subset of the strains from Langat *et al.* (2017) for which reasonably accurate geographic coordinates could be determined). The analyzed timetrees and tip coordinates are provided as Supplemental Material and shown in Figure 5a,b. To each clade, I fitted an SBM model with linearly varying diffusivity, estimated the statistical significance of the fitted slope, and assessed the "consistency" of the data with the fitted model as described in the previous section.

For both fitted models, the consistency was far above the significance threshold of 0.05 (Victoria: 0.306, Yamagata: 0.684), indicating that the data are—in terms of the expected log-likelihoods—consistent with the fitted models. For the Victoria clade, I estimated that the diffusivity has been increasing over time from $D=0.17\times 10^6$ km$^2\cdot$year^{-1} (i.e., a diffusion scale of $\delta=0.15$, based on an average temporal scale $\tau\approx 1.7$ years of independent contrasts) at the root to $D=23\times 10^6$ km$^2\cdot$year^{-1} (i.e., $\delta=1.7$) at the youngest tips (Fig. 5c). The slope of the fitted linear profile was not found to be statistically significant given the data at hand ($P=0.12$), although the 95% confidence intervals strongly suggest at least some increase occurred over time (Fig. 5a). In contrast, for the Yamagata clade I estimated that the diffusivity has been decreasing over time from $D=85\times 10^6$ km$^2\cdot$year^{-1} (i.e., $\delta=3.0$ based on $\tau\approx 1.4$ year) at the root to $D=21\times 10^6$ km$^2\cdot$year^{-1} (i.e., $\delta=1.5$) at the youngest tips (Fig. 5d). This decrease was found to be statistically significant ($P=0.006$) and is also strongly supported by the 95% confidence intervals (Fig. 5b). These estimates suggest that the two co-circulating Influenza B clades have historically differed substantially in their geographic dispersal rates, although at present both clades appear to exhibit similar diffusivities. The magnitude of the estimated diffusivities can be put into perspective by considering the expected distance that would be traversed by a lineage within a specific time interval: Based on the estimated present-day diffusivities, an Influenza B lineage traverses on average about 2400 km within 1 month. It should be kept in mind that this is an approximate estimate under the idealized scenario of purely diffusive dispersal; in reality, non-diffusive dispersal (e.g., through long-distance flights) likely also occurs in addition to diffusion-like dispersal (Bonabeau *et al.*, 1998).

Conclusions

I have presented methods for efficiently simulating and fitting spherical Brownian motion (diffusion) models along a phylogeny, using a modification of Felsenstein's independent contrasts. In the case of fast global dispersal, accounting for spherical geometry is essential for obtaining accurate diffusivity estimates. Even for small dispersal ranges, where planar BM and SBM become equivalent, the methods presented here enable new types of analyses not previously possible, such as simulating and fitting arbitrary time-dependent diffusivity profiles with an arbitrary number of free parameters, for example depending on continuously varying environmental variables. My methods scale well to large phylogenies comprising thousands to millions of tips, thus enabling efficient large-scale analyses, including parametric bootstrapping and simulation-based hypothesis testing. My methods provide improved means to investigate the dispersal process of bacteria, globally circulating viral pathogens, invasive species, and human culture.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.573n5tb4d.

Code Availability

Methods described in this article are available in the R package `castor` v1.6.2, as functions `fit_SBM_const`, `fit_SBM_parametric`, `fit_SBM_linear`, `fit_and_compare_sbm_const` and `simulate_SBM`. The package `castor` is available on The Comprehensive R Archive Network (CRAN).

Funding

S.L. was supported by a startup grant by the University of Oregon, USA.

Competing Financial Interests

The authors declare that they have no competing interests.

Materials & Correspondence

Correspondence and requests for materials should be addressed to S.L.

Appendix 1 Tip distances are independent realizations of SBM

In this section, I explain how the central angle ω between any two given tips can be interpreted as a realization of an SBM. I will consider the general case of time-dependent diffusivity $D(t)$. Specifically, I will show that if t_a is the time (distance from root) of the most recent common ancestor (MRCA) of the two tips, and $t_{c,1}$ and $t_{c,2}$ are the times of the two tips, then ω is distributed as if it were the central angle traversed by an SBM during the time interval $[0, t_1+t_2]$ (where $t_1 := t_{c,1}-t_a$ and $t_2 := t_{c,2}-t_a$) with time dependent diffusivity:

$$\overline{D}(t) := \begin{cases} D(t_a+t) & t\in[0,t_1) \\ D(t_a+t-t_1) & t\in[t_1,t_1+t_2]. \end{cases} \quad \text{(A1)}$$

Proof: Let p_a be the geographic location of the MRCA, and $p_{c,1}$ and $p_{c,2}$ the geographic locations of the two tips. In the language of stochastic processes, lineage 1 started at p_a at time t_a and diffused on the sphere according to an SBM with diffusivity D over the time interval $[t_a, t_{c,1}]$. Concurrently, lineage 2 started at p_a at time t_a and diffused on the sphere according to an

SBM with diffusivity D over the time interval $[t_a, t_{c,2}]$, independently of lineage 1. One would obtain the same probability distribution for $(p_{c,1}, p_{c,2})$ if lineage 1 started at p_a at time 0 and diffused according to the diffusivity $\overline{D}$ over the time interval $[0, t_1]$, and lineage 2 started at p_a at time t_1 and diffused according to the diffusivity $\overline{D}$ over the time interval $[t_1, t_1+t_2]$ (these modifications are simple time-shifts of the two processes). Since the SBM process is assumed to be isotropic and homogeneous, the distribution of the central angle ω between $p_{c,1}$ and $p_{c,2}$ remains unchanged if instead of lineage 2 diffusing, lineage 2 stayed at p_a and lineage 1 continued to diffuse according to the diffusivity $\overline{D}$ over the time interval $[t_1, t_1+t_2]$. Hence, ω has the same distribution as the central angle traversed by a lineage during the time interval $[0, t_1+t_2]$ with diffusivity $\overline{D}$.

Appendix 2 Time-dependent diffusivity

In this section, I explain why the transition density of a time-dependent SBM model can be obtained from the transition density of a time-independent SBM by simply replacing $(t_2-t_1)D$ with $\int_{t_1}^{t_2} D(t)dt$. Instead of the central angle ω, I will consider the transformed variable $x=\cos(\omega)$, whose transition density is simply that of ω divided by $\sin(\omega)$. According to Brillinger (2012, Eq. 3.10 therein), the stochastic differential equation for x can be written as:

$$dx = -2\frac{D(t)}{r^2}x\,dt - \frac{D(t)}{r^2}\sqrt{2(1-x^2)}\,dB, \tag{B2}$$

where B is the standard 1D Brownian Motion, also known as "Wiener process". Let y be a stochastic process satisfying the simpler stochastic differential equation with unit diffusivity and unit radius:

$$dy = -2y\,dt - \sqrt{2(1-y^2)}\,dB. \tag{B3}$$

Then the transition density of y after time t is that of a standard time-independent SBM:

$$\tilde{\rho}(y;t) = \frac{1}{2}\sum_{n=0}^{\infty}(2n+1)P_n(y)e^{-n(n+1)t}. \tag{B4}$$

Let $A(t) = \frac{1}{r^2}\int_0^t D(s)\,ds$. Then it is straightforward to verify that the modified process $x(t) = y(A(t))$, that is, with time t non-linearly rescaled according to A, is a solution to the stochastic differential equation (B2). Further, the transition density of x is the same as that of y after replacing t with $A(t)$, that is:

$$\tilde{\rho}(x;t) = \frac{1}{2}\sum_{n=0}^{\infty}(2n+1)P_n(x)e^{-n(n+1)\frac{1}{r^2}\int_0^t D(s)\,ds}. \tag{B5}$$

Corollary: The probability density of the central angle ω between two tips can be calculated using the transition density of the time-independent SBM (Eq. 1 in the main article) by replacing tD therein with the expression:

$$\tilde{D}(t_{c,1}) - \tilde{D}(t_a) + \tilde{D}(t_{c,2}) - \tilde{D}(t_a), \tag{B6}$$

where $\tilde{D}(t) = \int_0^t D(s)\,ds$ is the temporal antiderivative of D, t_a is the time of the MRCA between the two tips and $t_{c,1}$ and $t_{c,2}$ are the times of the two tips.

Proof of corollary: In Appendix 1, we have seen that ω is distributed as if it were the central angle traversed by a diffusing lineage over a time interval $[0, t_1+t_2]$ (where $t_k := t_{c,k} - t_a$) with diffusivity $\overline{D}$ given by Eq. (A1). By Appendix 2, Eq. (B5), the probability density of ω is thus:

$$\begin{aligned}
\rho(\omega;t) &= \frac{\sin(\omega)}{2}\sum_{n=0}^{\infty}(2n+1)P_n(x) \\
&\quad \exp\left[-n(n+1)\frac{1}{r^2}\int_0^{t_1+t_2}\overline{D}(s)\,ds\right] \\
&= \frac{\sin(\omega)}{2}\sum_{n=0}^{\infty}(2n+1)P_n(x) \\
&\quad \exp\left[-n(n+1)\frac{1}{r^2}\left[\int_0^{t_1}\overline{D}(s)\,ds + \int_{t_1}^{t_1+t_2}\overline{D}(s)\,ds\right]\right] \\
&= \frac{\sin(\omega)}{2}\sum_{n=0}^{\infty}(2n+1)P_n(x) \\
&\quad \exp\left[-n(n+1)\frac{1}{r^2}\left[\int_{t_a}^{t_{c,1}}D(s)\,ds + \int_{t_a}^{t_{c,2}}D(s)\,ds\right]\right].
\end{aligned} \tag{B7}$$

Appendix 3 Proof on the number of independent contrasts

In this section, I prove the following statement: For any rooted bifurcating and/or monofurcating tree with M tips one can find $\lfloor M/2 \rfloor$ independent contrasts, that is, tip pairs with disjoint shortest connecting paths. By "disjoint" I mean that any two such paths have no edge in common (although they may share nodes). I prove this statement by providing an explicit algorithm for finding $\lfloor M/2 \rfloor$ such tip pairs. In the following I say that a node is an "n-node" (for some $n \in \mathbb{N}$) if it has exactly n descending tips.

Lemma: Existence of nodes with exactly 2 descending tips Consider a rooted tree with only bifurcating or monofurcating nodes, and at least 2 tips. Then there must exist at least one 2-node.

Proof: One needs to show that the tree has at least one 2-node. The following algorithm will inevitably lead to a 2-node, moving from root to tips. The variable K denotes a node in the tree, and the algorithm guarantees that K always has ≥ 2 descending tips.

1. Set K to the root node. Note that by assumption K has ≥ 2 descending tips.

2. If K is a 2-node, we are done. Otherwise, since K has ≥ 3 descending tips, at least one of its child nodes (of which there are at most 2) must have ≥ 2 descending tips. Set K to that child node. Repeat step 2 until reaching a 2-node.

Observe that with every iteration the number of tips descending from K decreases, and hence the algorithm must inevitably reach a point where K is a 2-node.
□

Algorithm for extracting $\lfloor M/2 \rfloor$ independent contrasts Consider a rooted bifurcating and/or monofurcating phylogenetic tree with M tips. For any two tips i and j denote p_{ij} their shortest connecting path, that is, the minimal sequence of edges connecting tip i to tip j. Consider the following iterative procedure:

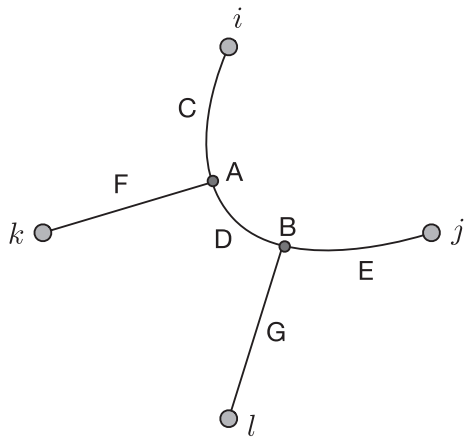
1. Pick a 2-node of the tree (according to any rule) with descending tips i and j and consider the path p_{ij} between those tips.

2. Remove the two tips from the tree and prune the tree accordingly, that is, removing any ancestral node that has no descending tips left.

3. Repeat steps 1 and 2 until no more 2-nodes are left.

Then the following holds: for any two tip pairs chosen in step 1, their shortest connecting paths are disjoint. Since the tree has a 2-node as long as it has at least 2 tips (see previous Lemma), this algorithm continues until there is exactly 1 tip left (if M is odd) or 0 tips left (if M is even). Thus the algorithm yields $\lfloor M/2 \rfloor$ tip pairs with disjoint shortest connecting paths.

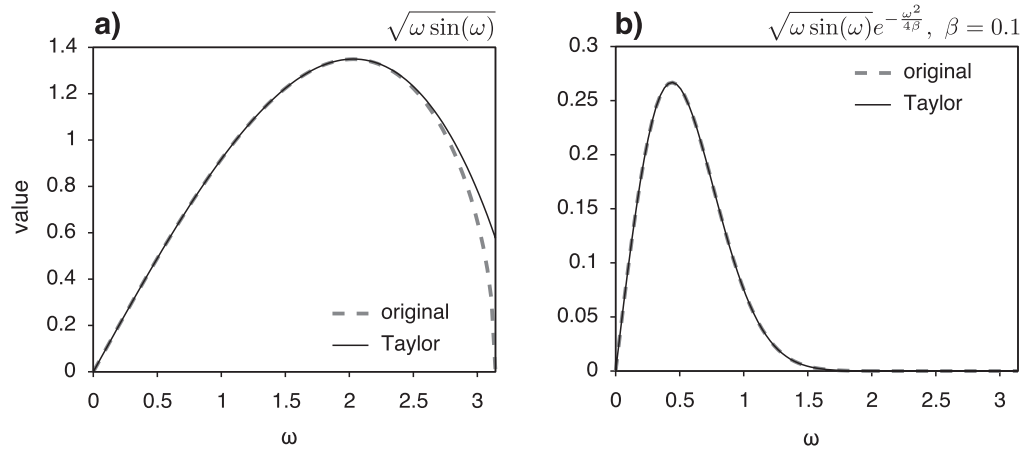
Proof: The proof proceeds by contradiction. Suppose that two tip pairs (i,j) and (k,l) were chosen in step 1 of the above algorithm such that p_{ij} and p_{kl} overlap (i.e., are non-disjoint). Then the 4 tips i,j,k,l must be topologically connected in the original tree as shown in Appendix Fig. 1 , where the letters A and B denote the two depicted nodes and the letters C–F denote path segments that may consist of one or more edges (i.e., may or may not include additional nodes). Note that no statement is made a priori about the direction (downstream vs. upstream) of the various segments. I will show that there cannot exist a 2-node whose sole descending tips are i and j (henceforth referred to as N). If N was at A, then j must also be descending from N and hence N cannot be a 2-node (it must be at least a 3-node). Similarly, if N was at B then i must also be descending from N and hence N cannot be a 2-node. If N was located somewhere along segment C, then both k and l must also be descending from N and hence N cannot be a 2-node. For a similar reason N cannot be located in segment E. If N was located in segment D then both k and l must also be descending from N, and hence again N cannot be a 2-node. If N was located along segment G then k must also be descending from N and hence N cannot be a 2-node. For a similar reason N cannot be located along segment F. In conclusion, there cannot be a 2-node in the tree whose sole descendants are i and j.
□



Appendix Figure 1. On the proof that $\lfloor M/2 \rfloor$ tip pairs with disjoint shortest connecting paths can be extracted from a rooted bifurcating and/or monofurcating tree with M tips (Appendix 3).

Appendix 4 Methods Details

To examine the accuracy of my simulation and fitting methods, I performed two types of simulations. In the first type (summarized in Fig. 3 and Appendix Fig. 3), I simulated Yule timetrees of various sizes with a constant speciation rate λ, where without loss of generality I set $\lambda = 1$. For any given tree size (e.g., 100 tips), I simulated SBM models with a constant diffusivity D, where various discrete values of D were considered. All SBM simulations were performed using the R function `simulate_sbm` in the R package `castor` v1.6.2 (Louca and Doebeli, 2018). Without loss of generality the sphere radius was set to $r = 1$, which essentially means that distances are measured in multiples of r. The root's location was set to latitude 0° and longitude 0°. The diffusivity was measured in units of r^2/τ, where τ is the expected temporal distance between sister tips in independent contrasts ($\sim 1.37/\lambda$ in the case of Yule trees). For any given tree size and diffusivity D, I performed 100 independent simulations of the SBM model along the tree, each time generating random tip coordinates, and then fitted a separate constant-diffusivity model to the data of each simulation. For fitting the SBM model, I extracted $\lfloor M/2 \rfloor$ independent contrasts (where M is the number of tips) and maximized the likelihood function described in the main article, using the `castor` function `fit_sbm_const`. For comparison, I also considered an alternative likelihood function (referred to as "planar distances" in Fig. 3), which assumes that pairwise tip distances are generated according to a 2D planar Brownian Motion process with constant diffusivity. For comparison, I also fitted the diffusivity using a classical planar BM model, by considering the (appropriately

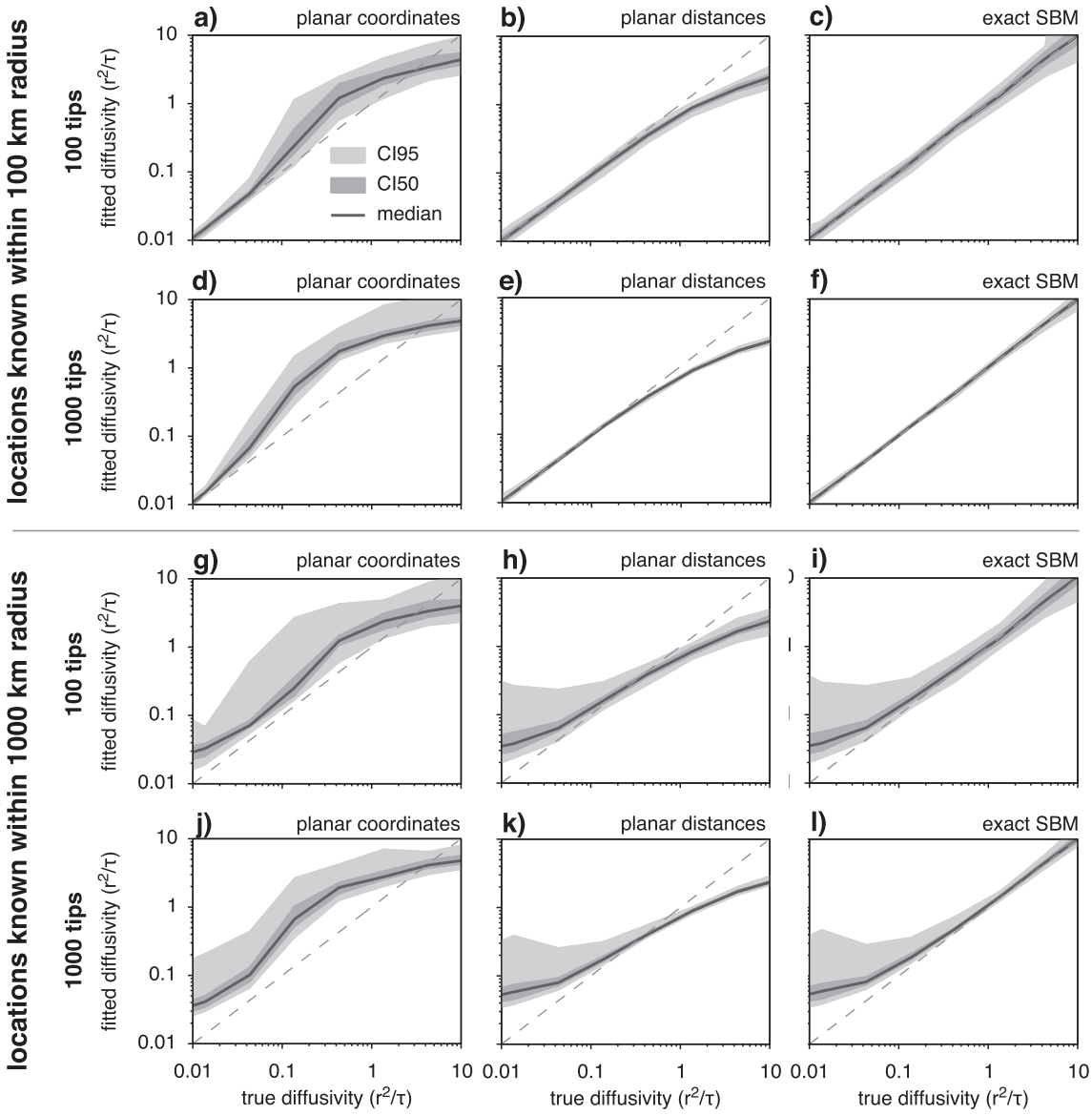


Appendix Figure 2. Taylor approximation of the transition density. a) The function $f(\omega)=\sqrt{\omega \sin(\omega)}$ ("original," dashed curve) compared to its Taylor expansion of 9th degree (continuous curve). b) The curves shown in a, multiplied by $e^{-\frac{\omega^2}{4\beta}}$, where $\beta=0.1$. Observe that any errors of the Taylor expansion at larger ω become negligible when multiplied by the rapidly decaying $e^{-\frac{\omega^2}{4\beta}}$.

rescaled) latitudes and longitudes as two numerical traits (`castor` function `fit_bm_model` with option "`isotropic=TRUE`"); this approximation is referred to as "planar coordinates" in Fig. 3. To examine the role of inaccurate tip coordinates, I repeated the above analyses using modified tip coordinates, which were chosen randomly within a specific "error radius" ε from their true location. I considered the error-radii $\varepsilon=0.016\times r$, $\varepsilon=0.078\times r$ and $\varepsilon=0.16\times r$, corresponding to geographic deviations up to 100 km, 500 km, and 1000 km, respectively. The distribution of fitted diffusivities across all 100 simulations, for any given tree size, error radius, and true diffusivity D, is shown in Fig. 3 and Appendix Figs. 3 and 4 .

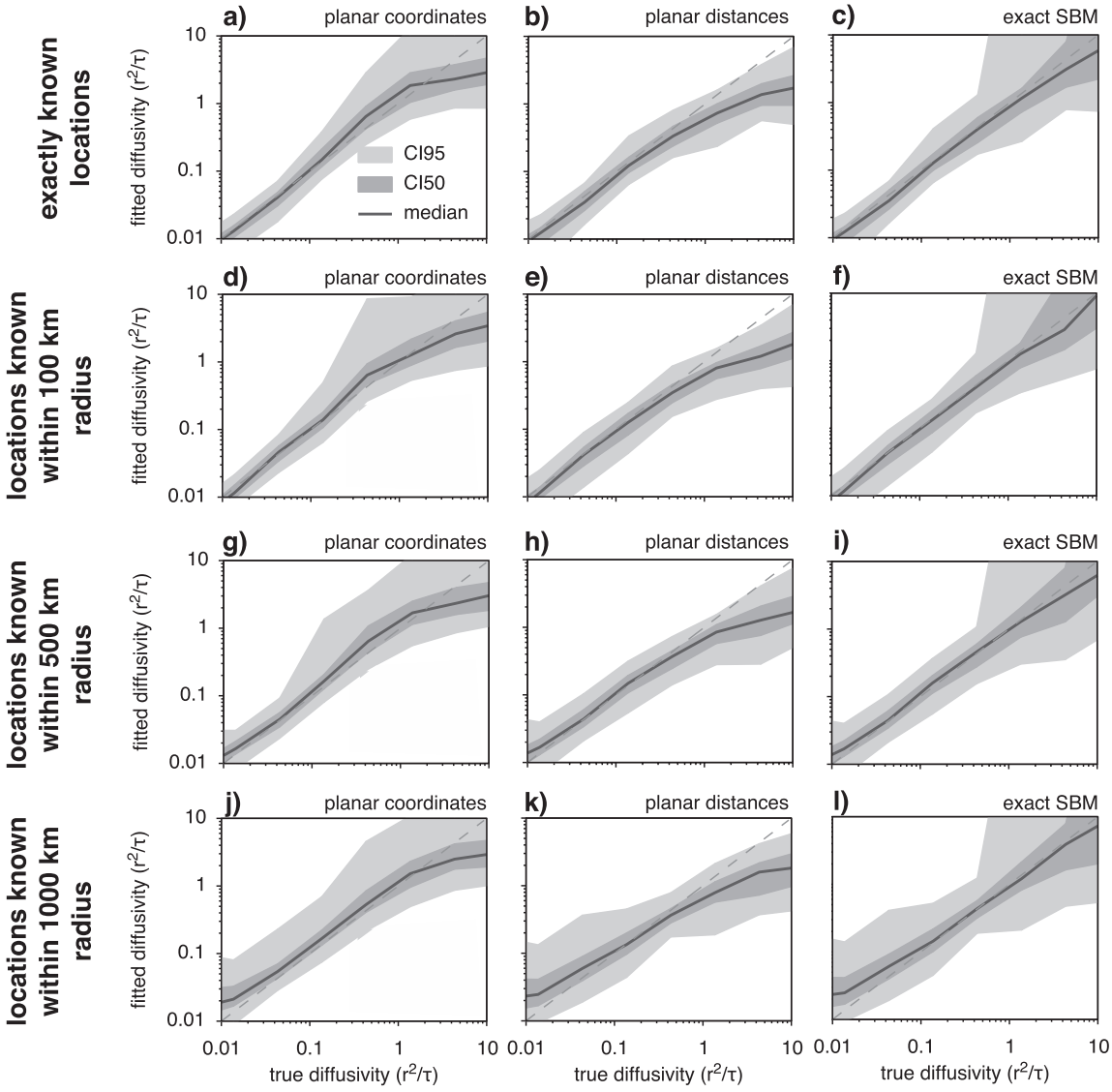
In the second simulation type, I generated timetrees using a birth–death model with constant speciation $\lambda=1$ and extinction rate $\mu=0.9\lambda$. On each timetree, I simulated an SBM model with a time-dependent diffusivity that varied linearly through time, from D_r at the root to D_p at the present (i.e., the tip farthest from the root), using the `castor` function `simulate_sbm`. As before, I assumed $r=1$. For each simulation, D_r and D_p were chosen randomly and uniformly between 0.01 and 1. I then fitted an SBM model with linearly varying diffusivity to the simulated data using the `castor` function `fit_sbm_linear`. Confidence intervals were estimated using 100 parametric bootstraps (option "`Nbootstraps=100`") and the statistical significance of the fitted slope was estimated using 100 simulations (option "`Nsignificance=100`"). I considered trees with 100 tips, 1000 tips, and 10,000 tips. For any given tree size, I counted the fraction of cases where the true diffusivity profile was completely contained in the estimated 95%-confidence interval. I also counted how many of the statistically significant fitted slopes had the same sign as the true diffusivity slope. Examples of simulated and fitted diffusivity profiles are shown in Fig. 4, Appendix Fig. 5 and Appendix Fig. 6 .

To demonstrate the computational efficiency of my simulation method (as implemented in `castor`), I proceeded as follows. I used a birth–death model with constant speciation rate $\lambda=1$ and extinction rate $\mu=0.9\lambda$ to generate extant timetrees of varying tree sizes (500–1,000,000 tips). For any given generated tree, I randomly chose a diffusivity between 0.01 $r^2\cdot\lambda$ and 100 $r^2\cdot\lambda$ (uniformly on a logarithmic scale) and simulated an SBM process along the tree. For any given tree size, this process was repeated 100 times, and the runtimes averaged over all 100 simulations and plotted as shown in Fig. 2a. To demonstrate the efficiency of my maximum likelihood fitting methods, I proceeded as follows. For any given tree size I generated 100 random extant timetrees and simulated an SBM process on each tree as described earlier in this passage. I then fitted the diffusivity and estimated confidence intervals through 100 parametric bootstraps, using the `castor` function `fit_sbm_const`. Runtimes were averaged over all 100 cases and are shown for various tree sizes in Fig. 2b. To compare the efficiency of my fitting method to that of other R packages capable of planar BM fitting (Appendix Fig. 7), that is, in the case where a planar approximation is adequate, I proceeded as follows. For any given tree size, I generated 100 random extant timetrees as before, and on each tree simulated an SBM process with a diffusivity chosen randomly between 0.001 $r^2\cdot\lambda$ and 0.01 $r^2\cdot\lambda$ (uniformly on a logarithmic scale) and with the root coordinates set to zero latitude and zero longitude. In each case, I fitted the diffusivity using the `castor` function `fit_sbm_const` (with option "`planar_approximation=TRUE`"), the function `mvBM` function in the package `mvMORPH` v1.1.3 (Clavel *et al.*, 2015) (with options "`model='BM1', method='rpf', param = list (constraint = 'equal', smean=TRUE, trend =FALSE), diagnostic=FALSE`"), the function `transformPhylo.ML` in the package



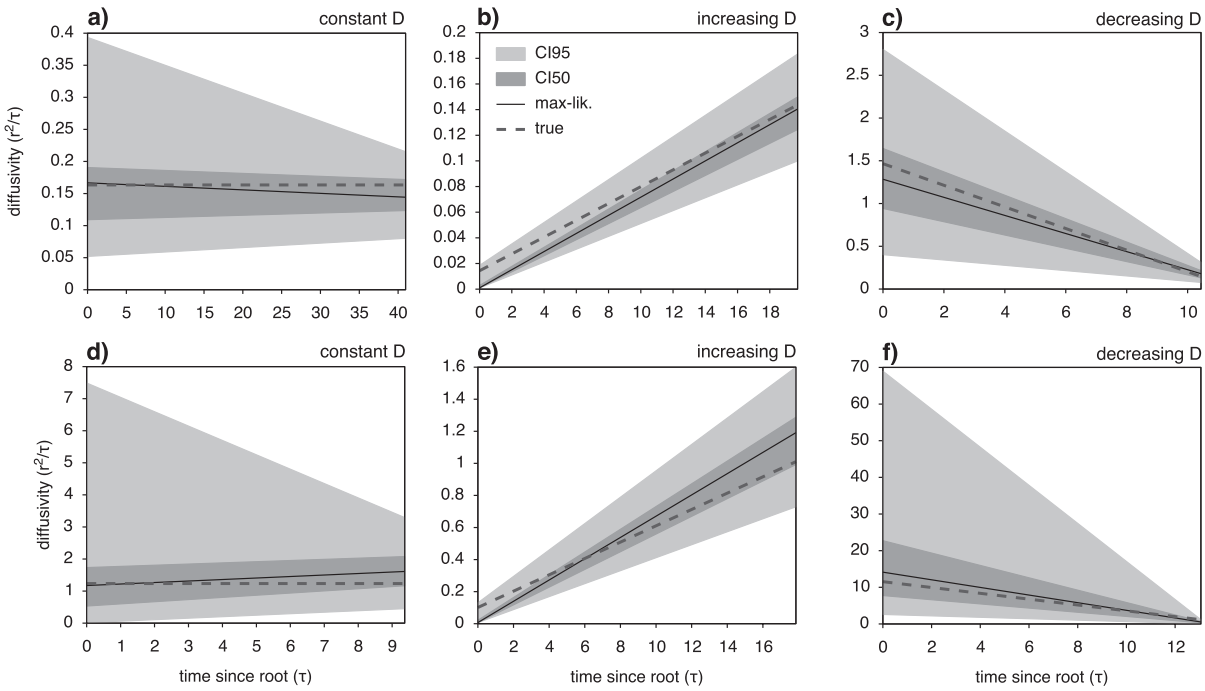
Appendix Figure 3. Evaluation of accuracy when tip locations are erroneous. Diffusivities (D) fitted to simulated phylogeographic data (timetrees generated by a Yule model, tip coordinates generated by an SBM model with constant diffusivity and subsequent random modification). Each figure shows the distribution of fitted diffusivities across a large number of simulations and for various true diffusivities. Horizontal axis: true diffusivity used in the simulations. Vertical axis: maximum likelihood-fitted diffusivity. Black curves show median fitted diffusivities. Shaded areas show 95% and 50% percentiles of fitted diffusivities. Dashed diagonals are shown for reference. a–c) Considering trees with 100 tips, and with tip locations replaced by random points within a 100 km radius (i.e., $\lesssim 0.016 \times r$) from the true locations. Diffusivities are either estimated (a) using planar BM, considering latitudes and longitudes as Cartesian coordinates, or (b) by correctly calculating geodesic distances between tips and fitting D under the assumption that these distances are generated by planar BM, or (c) by fitting an SBM model. d–f) Similar to (a–c), but considering trees with 1000 tips. (g–l): Similar to (a–f), but with tip locations replaced by random points within a 1000 km radius (i.e., $\lesssim 0.16 \times r$) from their true locations. All diffusivities are expressed in terms of squared sphere radii per τ, where τ is the average temporal scale of independent contrasts.

`motmot` v2.1.3 (Thomas and Freckleton, 2012) (with options "`model='bm', modelCIs=FALSE, returnPhy=FALSE`"), the function `brown.fit` in the package `slouch` v2.1.4 (Hansen *et al.*, 2008), the function `fitContinuous` in the package `geiger` v2.0.7 (Pennell *et al.*, 2014) (with option "`model='BM'`"), the functions `find.mle` and `make.bm` in the package `diversitree` v09-13 (FitzJohn, 2012) and the function `brownie.lite` in the package `phytools` v0.7-47 (Revell, 2012). For packages that could only handle univariate numerical traits I only provided tip longitudes. All benchmarks were performed on a 2015 MacBook Pro (2.9 GHz Intel Core i5) using a single core. For any given tree size and any given package, runtimes were averaged across all 100 simulations and are shown in Appendix Fig. 7 .



Appendix Figure 4. Evaluation of accuracy for small trees (10 tips). Diffusivities (D) fitted to simulated phylogeographic data (timetrees generated by a Yule model, 10 tips per tree, tip coordinates generated by an SBM model with constant diffusivity and subsequent random modification). Each figure shows the distribution of fitted diffusivities across a large number of simulations and for various true diffusivities. Horizontal axis: true diffusivity used in the simulations. Vertical axis: maximum likelihood-fitted diffusivity. Black curves show median fitted diffusivities. Shaded areas show 95% and 50% percentiles of fitted diffusivities. Dashed diagonals are shown for reference. a–c) Considering trees with exactly known tip coordinates. Diffusivities are either estimated (a) using planar BM, considering latitudes and longitudes as Cartesian coordinates, or (b) by correctly calculating geodesic distances between tips and fitting D under the assumption that these distances are generated by planar BM, or (c) by fitting an SBM model. d–f) Similar to (a–c), but considering trees with tip locations replaced by random points within a 100 km radius (i.e., $\lesssim 0.016 \times r$) from the true locations. g–i): Similar to (a–c), but with tip locations replaced by random points within a 500 km radius (i.e., $\lesssim 0.079 \times r$) from the true locations. j–l) Similar to (A–C), but with tip locations replaced by random points within a 1000 km radius (i.e., $\lesssim 0.16 \times r$) from the true locations. All diffusivities are expressed in terms of squared sphere radii per τ, where τ is the average temporal scale of independent contrasts.

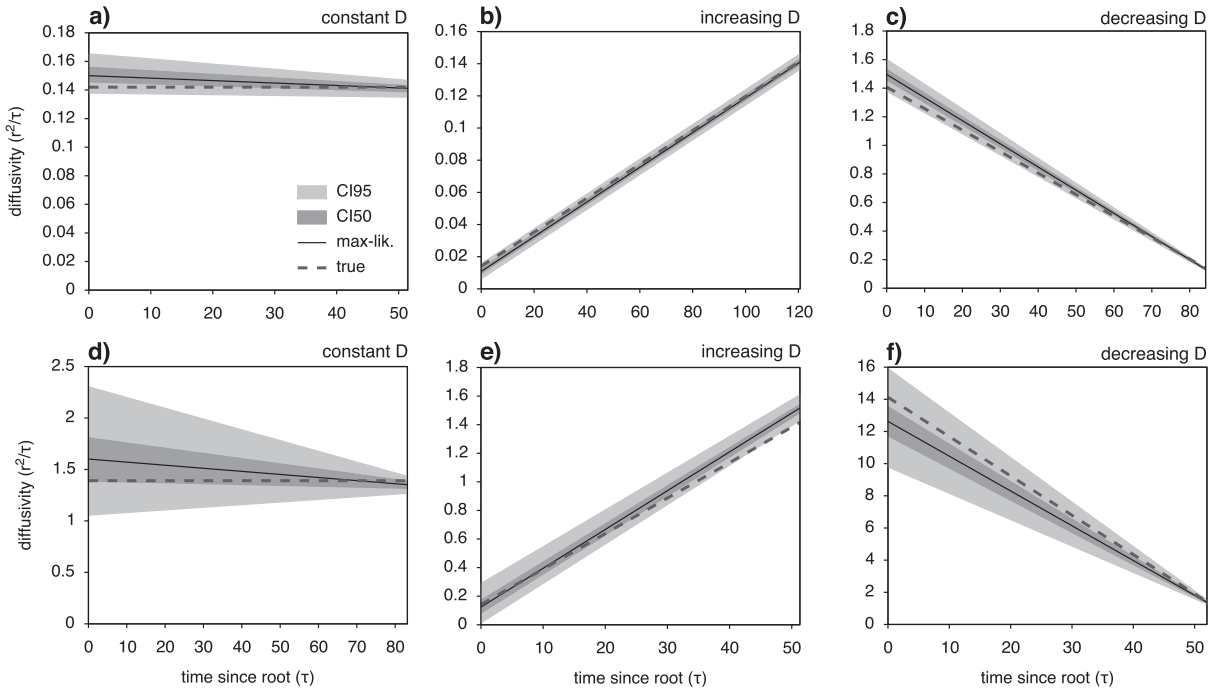
To analyze the geographic diffusivity of Cyanobacteria, I proceeded as follows. Genomes were downloaded from GenBank according to the following criteria: their contig-N50 value is below 5000, geographic coordinates are available for their associated BioSample, their completeness is $\geq$90% and their contamination is $\leq$1% (as inferred using the software checkM v1.1.2; Parks *et al.*, 2014). A phylogeny was constructed from the genomes based on multiple universal marker genes using GTDB-Tk v0.3.3 (Chaumeil *et al.*, 2019) (with default options) and FastTree v2.1.11 (Price *et al.*, 2010) (with options "`-wag -gamma`"). The tree was rooted using non-Cyanobacterial bacteria as outgroup. The tree was subsequently dated (time-calibrated) using PATHd8 v1.0 (Britton *et al.*, 2007), by fixing the root at 2.55 Ga (Shih *et al.*, 2017). Cyanobacteria were identified as either marine or terrestrial (defined here as anything non-marine) based on their geographic locations using



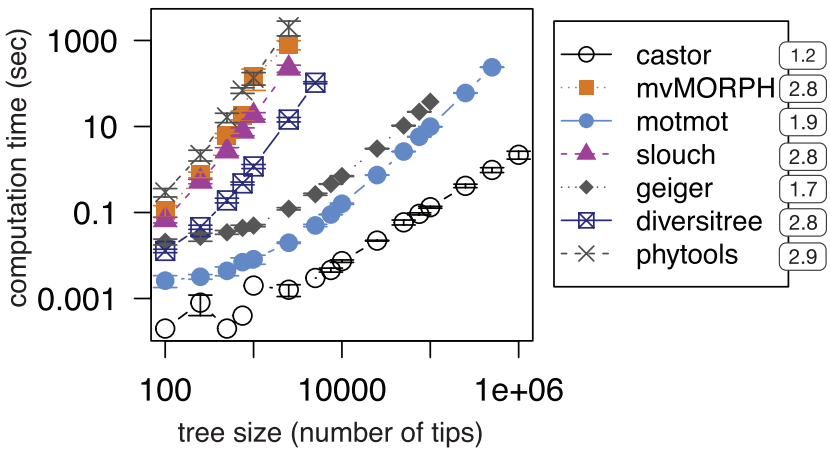
Appendix Figure 5. Fitting SBM models with linearly varying diffusivity (100 tips). Diffusivities fitted to simulated phylogeographic data (100 tips) simulated under an SBM model. In both the simulation and the fitted models, the diffusivity was assumed to vary linearly over time. The timetrees were generated using a birth–death model with extinction rate equal to 90% of the speciation rate, and both extant and extinct lineages were kept. Geographic coordinates were either simulated with a constant diffusivity (i.e., zero slope, a and d), a linearly increasing diffusivity (b and e) or a linearly decreasing diffusivity (c and f). Examples in the bottom row consider a roughly 10-fold higher diffusivity than in the top row. The diffusivity is measured in units of r^2/τ, where r is the sphere radius and τ is the average temporal scale of the independent contrasts. Dashed lines show maximum likelihood fitted diffusivities, shaded areas show 50% and 95% confidence intervals, and the solid lines show the true diffusivities. For analogous simulations using trees with 1000 or 10,000 tips, see Figure 4 and Appendix Figure 6 , respectively.

the python package `global_land_mask` v0.0.3 (van Rossum, 1995). To fit an SBM model separately to marine and terrestrial Cyanobacteria, the timetree was pruned to each of the two environment types prior to model fitting. Note that, strictly speaking, each pruned tree may include old ancestral nodes adapted to a different environment type than the tips; this, however, has a negligible effect on the fitted diffusivities since the independent contrasts are mostly constructed from pairs of closely related tips. The Cyanobacterial marine and terrestrial timetrees are provided as Supplementary Files S1 and S2 available on Dryad; associated tip metadata (including accession numbers and coordinates) are provided as Supplementary File S3 available on Dryad. SBM models with constant diffusivity were fitted to the two trees and compared using the `castor` function `fit_and_compare_sbm_const`, with options "`Nbootstraps=1000, Nsignificance=10000, radius=6371, only_distant_tip_pairs=TRUE`". Planar BM models with constant diffusivity were fitted using the function `mvBM` in the R package `mvMORPH` v1.1.0 (Clavel *et al.*, 2015), with options "`data=coordinates*(6371*pi/180), model='BM1', method='rpf',param = list(constraint = 'equal', smean = TRUE, trend=FALSE)`", where `coordinates` is a matrix listing the latitudes and longitudes of the tips. Example code demonstrating the SBM model fitting to the Cyanobacterial trees is provided as Supplementary File S4 available on Dryad.

To analyse the geographic diffusivity of Influenza B, I proceeded as follows. Separate timetrees for the Victoria and Yamagata clades based on the hemagglutinin gene were obtained from the supplementary material published by Langat *et al.* (2017), available on Dryad (`https://doi.org/10.5061/dryad.s1d37`) (maximum clade credibility trees "`unified_global_vic.HA.mcc.tre`" and "`unified_global_yam.HA.mcc.tre`"). Geographic coordinates for some of the strains were also obtained from the same location (file "`fluB_location_summary.txt`") and supplemented where needed and possible based on city or other administrative unit names. Models with linearly varying diffusivity over time were fitted using the `castor` function `fit_sbm_linear`, with options "`Ntrials=100, Nbootstraps=500, Nsignificance=500, radius=6371`". The timetrees of the Victoria and Yamagata clades are provided as Supplementary Files S5 and S6 available on Dryad, respectively; associated tip coordinates are provided as Supplementary Files S7 and S8 available on Dryad.



Appendix Figure 6. Fitting SBM models with linearly varying diffusivity (10,000 tips). Diffusivities fitted to simulated phylogeographic data (10,000 tips) simulated under an SBM model. In both the simulation and the fitted models the diffusivity was assumed to vary linearly over time. The timetrees were generated using a birth–death model with extinction rate equal to 90% of the speciation rate, and both extant and extinct lineages were kept. Geographic coordinates were either simulated with a constant diffusivity (i.e., zero slope, a and d), a linearly increasing diffusivity (b and e) or a linearly decreasing diffusivity (c and f). Examples in the bottom row consider a roughly 10-fold higher diffusivity than in the top row. The diffusivity is measured in units of r^2/τ, where r is the sphere radius and τ is the average temporal scale of the independent contrasts. Dashed lines show maximum likelihood fitted diffusivities, shaded areas show 50% and 95% confidence intervals, and the solid lines show the true diffusivities. For analogous simulations using trees with 1000 tips, see Figure 4 in the main article.



Appendix Figure 7. Comparison of computation times to other planar-BM software. Computation times (T) needed for fitting planar BM models with constant diffusivity to phylogenies of varying sizes (S), using `castor` and other R packages. Circles show mean computation times and vertical error bars extend one standard deviation above and below the mean, calculated based on 100 independent repeats. Numerical labels next to package names indicate fitted power-law exponents ($T \propto S^p$).

REFERENCES

Bloomquist, E.W., Lemey, P., and Suchard, M.A. (2010) Three roads diverged? Routes to phylogeographic inference. *Trends Ecol. Evol.* **25**: 626–632.

Bonabeau, E., Toubiana, L., and Flahault, A. (1998) The geographical spread of influenza. *Proc. R. Soc. Lond. B* **265**: 2421–2425.

Bouckaert, R. and Cartwright, R. (2016) Phylogeography by diffusion on a sphere: whole world phylogeography. *PeerJ* **4**: e2406.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., *et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**: 1–6.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S.J., Alekseyenko, A.V., Drummond, A.J., *et al.* (2012) Mapping the origins and expansion of the Indo-European language family. *Science* **337**: 957–960.

Brillinger, D.R. (2012) A particle migrating randomly on a sphere. In: Selected works of David Brillinger. Springer. p. 73–87.

Britton, T., Anderson, C.L., Jacquet, D., Lundqvist, S., and Bremer, K. (2007) Estimating divergence times in large phylogenetic trees. *Syst. Biol.* **56**: 741–752.

Brown, J.M. and Thomson, R.C. (2018) Evaluating model performance in evolutionary biology. *Annu. Rev. Ecol. Evol. Syst.* **49**: 95–114.

Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**: 1925–1927.

Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2015) Genbank. *Nucleic Acids Res.* **44**: D67–D72.

Clavel, J., Escarguel, G., and Merceron, G. (2015) mvmorph: an r package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**: 1311–1319.

Currie, T.E., Meade, A., Guillon, M., and Mace, R. (2013) Cultural phylogeography of the Bantu languages of sub-Saharan Africa. *Proc. R. Soc. B: Biol. Sci.* **280**: 20130695.

De Wit, R. and Bouvier, T. (2006) 'Everything is everywhere, but, the environment selects': what did Baas Becking and Beijerinck really say? *Environ. Microbiol.* **8**: 755–758.

Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L., Harmon, L.J. (2011) A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* **65**: 3578–3589.

Faria, N.R., Suchard, M.A., Abecasis, A., Sousa, J.D., Ndembi, N., Bonfim, I., *et al.* (2012) Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. *Infect. Genet. Evol.* **12**: 453–460.

Faria, N.R., Suchard, M.A., Rambaut, A., and Lemey, P. (2011) Toward a quantitative understanding of viral phylogeography. *Curr. Opin. Virol.* **1**: 423–429.

Felsenstein, J. (1985) Phylogenies and the comparative method. *Am. Nat.* **125**: 1–15.

Finlay, B.J. and Clarke, K.J. (1999) Ubiquitous dispersal of microbial species. *Nature* **400**: 828.

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**: 1084–1092.

Frank, M. (2002) Radiogenic isotopes: Tracers of past ocean circulation and erosional input. *Rev. Geophys.* **40**: 1–1–1–38.

Freckleton, R.P. (2012) Fast likelihood calculations for comparative analyses. *Methods Ecol. Evol.* **3**: 940–947.

Ghosh, A., Samuel, J., and Sinha, S. (2012) A "Gaussian" for diffusion on the sphere. *Europhys. Lett.* **98**: 30003.

Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R., and Gilbert, J.A. (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proc. Natl. Acad. Sci. USA* **110**: 4651–4655.

Goldberg, E.E., Lancaster, L.T., and Ree, R.H. (2011) Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.* **60**: 451–465.

Gonnella, G., Böhnke, S., Indenbirken, D., Garbe-Schönberg, D., Seifert, R., Mertens, C., *et al.* (2016) Endemic hydrothermal vent species identified in the open ocean seed bank. *Nat. Microbiol.* **1**: 16086 EP.

Hansen, T.F., Pienaar, J., and Orzack, S.H. (2008) A comparative method for studying adaptation to a randomly evolving environment. *Evolution* **62**: 1965–1977.

Janzen, T., Höhna, S., and Etienne, R.S. (2015) Approximate bayesian computation of diversification rates from molecular phylogenies: introducing a new efficient summary statistic, the nltt. *Methods Ecol. Evol.* **6**: 566–575.

Landis, M.J., Matzke, N.J., Moore, B.R., and Huelsenbeck, J.P. (2013) Bayesian analysis of biogeography when the number of areas is large. *Syst. Biol.* **62**: 789–804.

Langat, P., Raghwani, J., Dudas, G., Bowden, T.A., Edwards, S., Gall, A., *et al.* (2017) Genome-wide evolutionary dynamics of influenza B viruses on a global scale. *PLoS Pathogens* **13**: e1006749.

Lange, K. (2010) Diffusion processes. In: Applied probability, chapter 11, pp. 269–295. New York, NY: Springer New York.

Lemey, P., Rambaut, A., Drummond, A.J., and Suchard, M.A. (2009) Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**: e1000520.

Lemey, P., Rambaut, A., Welch, J.J., and Suchard, M.A. (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**: 1877–1885.

Lemmon, A.R. and Lemmon, E.M. (2008) A likelihood framework for estimating phylogeographic history on a continuous landscape. *Syst. Biol.* **57**: 544–561.

Lindholm, A., Zachariah, D., Stoica, P., and Schön, T.B. (2019) Data consistency approach to model validation. *IEEE Access* **7**: 59788–59796.

Louca, S. (2020) Simulating trees with millions of species. *Bioinformatics* **36**: 2907–2908.

Louca, S. and Doebeli, M. (2018) Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**: 1053–1055.

Louca, S. and Pennell, M.W. (2020) Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**: 502–505.

Louca, S., Shih, P.M., Pennell, M.W., Fischer, W.W., Parfrey, L.W., and Doebeli, M. (2018) Bacterial diversification through geological time. *Nat. Ecol. Evol.* **2**: 1458–1467.

Marshall, C.R. (2017) Five palaeobiological laws needed to understand the evolution of the living biota. *Nat. Ecol. Evol.* **1**: 165.

Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst. Biol.* **63**: 951–970.

Monjane, A.L., Harkins, G.W., Martin, D.P., Lemey, P., Lefeuvre, P., Shepherd, D.N., *et al.* (2011) Reconstructing the history of maize streak virus strain a dispersal to reveal diversification hot spots and its origin in southern africa. *J. Virol.* **85**: 9623–9636.

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

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2014) Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**: 1043–1055.

Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G., *et al.* (2014) geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**: 2216–2218.

Perrin, F. (1928) Étude mathématique du mouvement brownien de rotation. In Annales scientifiques de l'École Normale Supérieure, vol. 45. p. 1–51.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2: approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Pybus, O.G., Suchard, M.A., Lemey, P., Bernardin, F.J., Rambaut, A., Crawford, F.W., *et al.* (2012) Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc. Natl. Acad. Sci. USA* **109**: 15066.

R Core Team (2019) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

Ree, R.H. and Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* **57**: 4–14.

Revell, L.J. (2012) phytools: an r package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**: 217–223.

Rota, P.A., Wallis, T.R., Harmon, M.W., Rota, J.S., Kendal, A.P., and Nerome, K. (1990) Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. *Virology* **175**: 59–68.

Shih, P.M., Hemp, J., Ward, L.M., Matzke, N.J., and Fischer, W.W. (2017) Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* **15**: 19–29.

Smith, S.A. and Brown, J.W. (2018) Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**: 302–314.

Stack, J.C., Harmon, L.J., and O'Meara, B. (2011) Rbrownie: an r package for testing hypotheses about rates of evolutionary change. *Methods Ecol. Evol.* **2**: 660–662.

Stadler, T., Kühnert, D., Bonhoeffer, S., and Drummond, A.J. (2013) Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. USA* **110**: 228–233.

Thomas, G.H. and Freckleton, R.P. (2012) MOTMOT: models of trait macroevolution on trees. *Methods Ecol. Evol.* **3**: 145–151.

Tung Ho, L.S. and Ané, C. (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**: 397–408.

van der Gast, C.J. (2015) Microbial biogeography: the end of the ubiquitous dispersal hypothesis? *Environ. Microbiol.* **17**: 544–546.

van Rossum, G. (1995) Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**: 976–978.

Whitfield, J. (2005) Is everything everywhere? *Science* **310**: 960–961.

Yang, J.R., Huang, Y.P., Chang, F.Y., Hsu, L.C., Lin, Y.C., Huang, H.Y., *et al.* (2012) Phylogenetic and evolutionary history of Influenza B viruses, which caused a large epidemic in 2011–2012, Taiwan. *PLoS One* **7**: e47179.