New Phylogenetic Models Incorporating Interval-Specific Dispersal Dynamics Improve Inference of Disease Spread

Jiansi Gao ,*[,1] Michael R. May,[1,2] Bruce Rannala,[1] and Brian R. Moore[1]

[1]Department of Evolution and Ecology, University of California, Storer Hall, Davis, CA 95616, USA
[2]Department of Integrative Biology, University of California, 3060 VLSB, Berkeley, CA 94720-3140, USA

***Corresponding author:** E-mail: jsigao@gmail.com.

Associate editor: Rasmus Nielsen

Abstract

Phylodynamic methods reveal the spatial and temporal dynamics of viral geographic spread, and have featured prominently in studies of the COVID-19 pandemic. Virtually all such studies are based on phylodynamic models that assume—despite direct and compelling evidence to the contrary—that rates of viral geographic dispersal are constant through time. Here, we: (1) extend phylodynamic models to allow both the average and relative rates of viral dispersal to vary independently between pre-specified time intervals; (2) implement methods to infer the number and timing of viral dispersal events between areas; and (3) develop statistics to assess the absolute fit of discrete-geographic phylodynamic models to empirical datasets. We first validate our new methods using simulations, and then apply them to a SARS-CoV-2 dataset from the early phase of the COVID-19 pandemic. We show that: (1) under simulation, failure to accommodate interval-specific variation in the study data will severely bias parameter estimates; (2) in practice, our interval-specific discrete-geographic phylodynamic models can significantly improve the relative and absolute fit to empirical data; and (3) the increased realism of our interval-specific models provides qualitatively different inferences regarding key aspects of the COVID-19 pandemic—revealing significant temporal variation in global viral dispersal rates, viral dispersal routes, and the number of viral dispersal events between areas—and alters interpretations regarding the efficacy of intervention measures to mitigate the pandemic.

Key words: **phylodynamic models, biogeographic history, epidemiology, phylogeography.**

Introduction

Phylodynamic methods encompass a suite of models for inferring various aspects of pathogen biology, including: (1) patterns of variation in demography through time (Drummond et al. 2005; Minin et al. 2008; Gill et al. 2013, 2016); (2) the history of geographic spread either over continuous space (Lemey et al. 2010; Pybus et al. 2012; Gill et al. 2017) or among a set of discrete-geographic areas (Lemey et al. 2009; Edwards et al. 2011), and; (3) the interaction between demography and geographic history (De Maio et al. 2015; Kühnert et al. 2016; Müller et al. 2017, 2019). Our focus here is on discrete-geographic phylodynamic models (Lemey et al. 2009; Edwards et al. 2011). These phylodynamic methods have been used extensively to understand the spatial and temporal spread of disease outbreaks and have played a central role for inferring key aspects of the COVID-19 pandemic, such as the geographic location and time of origin of the disease, the rates and geographic routes by which it spread, and the efficacy of various mitigation measures to limit its geographic expansion (Bedford et al. 2020; Candido et al. 2020; Fauver et al. 2020; Worobey et al. 2020; Alpert et al. 2021; Davies et al. 2021; Dellicour et al. 2021; Douglas et al. 2021; du Plessis et al. 2021; Kraemer et al. 2021; Lemey et al. 2021; Müller et al. 2021; Nadeau et al. 2021; Tegally et al. 2021; Washington et al. 2021; Wilkinson et al. 2021).

These phylodynamic methods adopt an explicitly probabilistic approach that model the process of viral dispersal among a set of discrete-geographic areas (Baele et al. 2017). The observations include the times and locations of viral sampling, and the genomic sequences of the sampled viruses. These data are used to estimate the parameters of phylodynamic models, which include a dated phylogeny of the viral samples, the global dispersal rate (the average rate of dispersal among all geographic areas), and the relative dispersal rates (the dispersal rate between each pair of geographic areas).

The vast majority (651 of 666, 97.7%; supplementary fig. S1, Supplementary Material online) of discrete-geographic phylodynamic studies are based on the earliest models (Lemey et al. 2009; Edwards et al. 2011), which assume that viral dispersal dynamics—including the average and relative rates of viral dispersal—remain constant over time (see the caption of fig. S1 for the search query we used to identify these studies). However, real-world observations indicate that the average and/or relative rates of viral dispersal inevitably vary during disease outbreaks. For example, relative rates of viral dispersal typically change as a disease is introduced to (and becomes

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Open Access

prevalent in) new areas, and begins dispersing from those areas to other areas. Dispersal dynamics are also generally impacted by the initiation (or alteration or cessation) of area-specific mitigation measures (e.g., domestic shelter-in-place policies) that change the rate of viral transmission within an area and the relative rate of dispersal to other areas. Similarly, average rates of viral dispersal may change in response to the initiation (or alteration or cessation) of more widespread intervention efforts—e.g., multiple area-specific mitigation measures, international-travel bans—that collectively impact the overall viral dispersal rate.

In this paper, we: (1) extend discrete-geographic phylodynamic models to allow both the average and relative dispersal rates to vary independently across pre-specified time intervals; (2) enable stochastic mapping under these models to estimate the number and timing of viral dispersal events between areas, and; (3) develop statistics to assess the absolute fit of discrete-geographic phylodynamic models to empirical datasets. We first validate the theory and implementation of our new phylodynamic methods using analyses of simulated data, and then provide an empirical demonstration of these methods with analyses of a SARS-CoV-2 dataset from the early phase of the COVID-19 pandemic.

Extending Phylodynamic Models

Anatomy of Interval-specific Phylodynamic Models

Phylodynamic models of dispersal include two main components (fig. 1): a *phylogenetic model* that allows us to estimate a dated phylogeny for the sampled viruses, Ψ, and a *biogeographic model* that describes the history of viral dispersal over the tree as a continuous-time Markov chain. For a geographic history with k discrete areas, this stochastic process is fully specified by a $k \times k$ instantaneous-rate matrix, $\mathbf{Q}$, where an element of the matrix, q_{ij}, is the instantaneous rate of change between state i and state j (i.e., the instantaneous rate of dispersal from area i to area j). We rescale the $\mathbf{Q}$ matrix such that the average rate of dispersal between all areas is μ; this represents the average rate of viral dispersal among all areas (Yang 2014).

We could specify alternative biogeographic models based on the assumed constancy of the dispersal process. For example, the simplest possible model assumes that the average dispersal rate, μ, and the relative dispersal rates, $\mathbf{Q}$, remain constant over the entire history of the viral outbreak. Typically, viral outbreaks are punctuated by events that are likely to impact the average rate of viral dispersal (e.g., the onset of an international-travel ban) and/or the relative rates of viral dispersal between pairs of areas (e.g., the initiation of localized mitigation measures). We can incorporate information on such events into our phylodynamic inference by specifying interval-specific models. That is, the investigator specifies the number of intervals, the boundaries between each interval, and the parameters that are specific to each interval according to the presumed changes in the history of viral dispersal. For example, we might specify an interval-specific model (Membrebe et al. 2019) that assumes that the average rate of viral dispersal varies among two or more intervals (while assuming that the relative rates of viral dispersal remain constant across intervals). Conversely, an interval-specific model (Bielejec et al. 2014) might allow the relative rates of viral dispersal to vary among two or more time intervals (while assuming that the average rate of viral dispersal remains constant across intervals). Alternatively, a more complex interval-specific model might allow both the average rate of viral dispersal and the relative rates of viral dispersal to vary among two or more intervals. We extend interval-specific phylodynamic models to allow *both* the relative *and* average dispersal rates to vary independently across two or more pre-defined intervals. Here, we describe how to compute transition probabilities, perform inference, simulate histories, and assess the absolute fit of the interval-specific models.

Computing Transition Probabilities

The transition-probability matrix, $\mathbf{P}$, describes the probability of transitioning from state i to state j (i.e., dispersing from area i to area j) along a branch with a finite duration; importantly, a branch may span two or more intervals with different relative and/or absolute dispersal rates.

Allowing average dispersal rates to vary across intervals

Under a constant-rate model, the transition-probability matrix for a branch is $\mathbf{P} = \exp(\mathbf{Q}\nu)$, where $\nu = \mu t$ represents the expected number of dispersal events on a branch of duration t with an average dispersal rate μ. However, under a phylodynamic model with interval-specific average dispersal rates (Membrebe et al. 2019)—which allows the average dispersal rate to vary among intervals, but assumes that relative dispersal rates are constant across all intervals—a given branch in a phylogeny may span two or more intervals with different average dispersal rates ("average-rate intervals"). The transition-probability matrix for the branch is then computed as the matrix exponential:

$$\mathbf{P} = \exp\left(\mathbf{Q}\sum_{l=1}^{n}\nu_l\right), \quad (1)$$

where $\mathbf{Q}$ is the instantaneous-rate matrix, n is the number of average-rate intervals spanned by the branch, and ν_l is the expected number of dispersal events on the branch in average-rate interval l. Recall that $\nu_l = \mu_l t_l$, where μ_l is the average dispersal rate during interval l and t_l is the time spent in interval l.

Allowing relative dispersal rates to vary across intervals

Under a phylodynamic model with interval-specific relative dispersal rates (Bielejec et al. 2014)—which allows

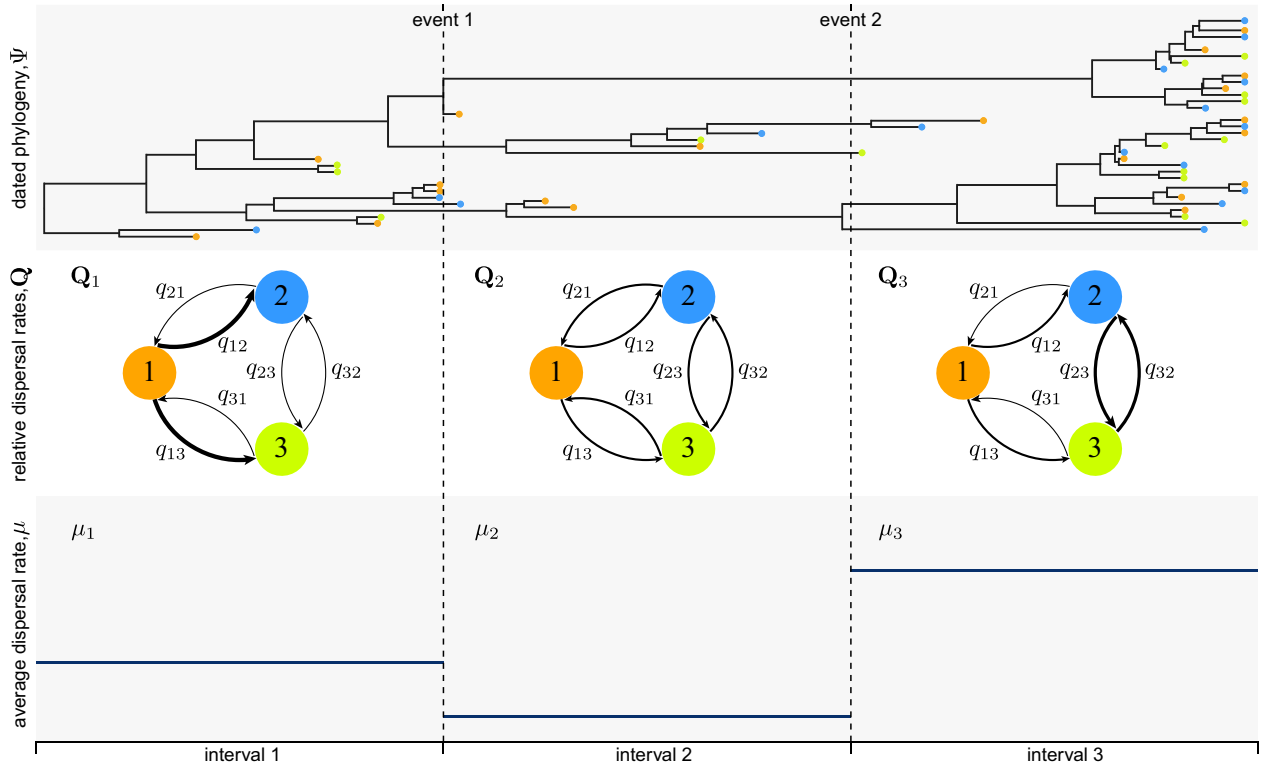

Fig. 1. Interval-specific phylodynamic models accommodate variation in the process of viral dispersal. Phylodynamic models include two main components: a *phylogenetic model* that specifies the relationships and divergence times of the sampled viruses, Ψ (top panel), and a *biogeographic model* that describes the history of viral dispersal among a set of discrete-geographic areas—here, areas 1 (orange), 2 (blue), and 3 (green)—from the root to the tips of the dated viral tree. Parameters of the biogeographic model include an instantaneous-rate matrix, **Q**, that specifies relative rates of viral dispersal between each pair of areas (here, each element of the matrix, q_{ij}, is represented as an arrow that indicates the direction and relative dispersal rate from area *i* to area *j*; middle panel), and a parameter that specifies the average rate of viral dispersal between all areas, μ (lower panel). Although most phylodynamic studies assume that the process of viral dispersal is constant through time, disease outbreaks are typically punctuated by events that impact the average and/or relative rates of viral dispersal among areas. Here, for example, the history involves two events (e.g., mitigation measures) that define three intervals, where both **Q** and μ are impacted by each of these events, such that the interval-specific parameters are ($\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3$) and (μ_1, μ_2, μ_3). Our framework allows investigators to specify discrete-geographic phylodynamic models with two or more intervals, where each interval has independent relative and/or average dispersal rates, which are then estimated from the data.

the instantaneous-rate of dispersal between each pair of areas to vary among intervals, but assumes that the average dispersal rate is constant across all intervals—a given branch may span two or more intervals with different **Q** matrices ("relative-rate intervals"). In this case, the transition-probability matrix for each relative-rate interval *l*, $\mathbf{P}_l$, is computed as:

$$\mathbf{P}_l = \exp(\mathbf{Q}_l \nu_l), \tag{2}$$

where $\mathbf{Q}_l$ is the instantaneous-rate matrix in relative-rate interval *l*, and $\nu_l = \mu t_l$ is the average dispersal rate multiplied by the time spent in interval *l*. The transition-probability matrix for the entire branch is then computed as the matrix product of interval-specific transition-probability matrices:

$$\mathbf{P} = \prod_{l=1}^{m} \mathbf{P}_l, \tag{3}$$

where *m* is the number of relative-rate intervals spanned by the branch.

Allowing average and relative dispersal rates to vary across intervals

We combine the two approaches described above to compute transition-probability matrices under an interval-specific model that allows both the average dispersal rate and the relative dispersal rates to vary independently among intervals. Let a given branch span *m* relative-rate intervals. The expected number of dispersal events in each such interval *l*, ν_l, is computed as:

$$\nu_l = \sum_{p=1}^{n} \mu_p t_p, \tag{4}$$

where *n* is the number of average-rate intervals spanned by interval *l*, μ_p is the dispersal rate in average-rate interval *p*, and t_p is the time spent in average-rate interval *p*. We then substitute equation (4) into equation (2), and apply equation (3) as normal to compute the transition-probability matrix for the entire branch. An example computation is illustrated in figure 2 for a scenario in which a branch spans two different relative-rate intervals and three different average-rate intervals.

$$\mathbf{P} = \mathbf{P}_1 \times \mathbf{P}_2 = \exp[\mathbf{Q}_1(\mu_1 t_1 + \mu_2 t_2)] \times \exp[\mathbf{Q}_2(\mu_2 t_3 + \mu_3 t_4)]$$

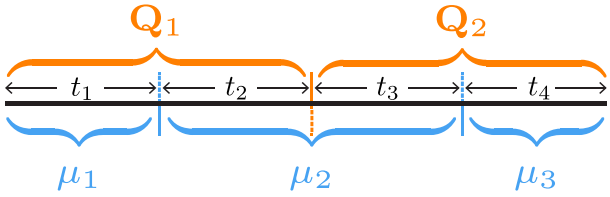

FIG. 2. Computing the transition-probability matrix for a branch spanning intervals where both the average and relative dispersal rates vary. An example illustrating the transition-probability matrix computation for a branch spanning two relative-rate intervals ($\mathbf{Q}_1$ and $\mathbf{Q}_2$) and three average-rate intervals (μ_1, μ_2, μ_3).

We modified BEAST source code to implement the above equation for computing $\mathbf{P}$ under our interval-specific phylodynamic models that allow both μ and $\mathbf{Q}$ to vary independently among two or more pre-specified intervals.

Inference under Interval-specific Phylodynamic Models

We estimate parameters of the interval-specific models within a Bayesian statistical framework. Specifically, we use numerical algorithms—Markov chain Monte Carlo (MCMC) simulation—to approximate the joint posterior probability distribution of the phylodynamic model parameters—the dated phylogeny, Ψ, the set of relative dispersal rates, $\mathbf{Q}$, and the average dispersal rates, μ—from the study data (i.e., the location and times of viral sampling, and the genomic sequences of the sampled viruses).

Simulating Dispersal Histories under Interval-specific Phylodynamic Models

We have also implemented numerical algorithms—stochastic mapping—to simulate histories of viral dispersal under the interval-specific models; these methods allow us to estimate the number of dispersal events between a specific pair of areas, the number of dispersal events from one area to a set of two or more areas, and the total number dispersal events among all areas. Stochastic mapping—initially proposed by Nielsen (2002; see also Huelsenbeck et al. 2003; Bollback 2006; Minin and Suchard 2008; Hobolth and Stone 2009)—is commonly used to sample dispersal histories over branches of a phylogeny. Here, we extend this approach to sample dispersal histories under our interval-specific models.

Let a given branch start at time T_0 with state i and end at time T_m with state k. Further, let the dispersal process change (either by changing the average or relative dispersal rates) $m-1$ times on the branch at times $\{T_1, \ldots, T_{m-1}\}$, resulting in m intervals. For interval l, denote the average dispersal rate as μ_l, the instantaneous-rate matrix as $\mathbf{Q}_l$, and the duration as t_l. We simulate a dispersal history along this branch using a two-step procedure: (1) we first sample the state at each of the $m-1$ time points, and; (2) we then simulate the history between each time point, conditional on the states sampled in the first step.

To simulate the states at each time point, we first compute a transition-probability matrix for each interval:

$$\mathbf{P}_l = \exp(\mathbf{Q}_l \mu_l t_l).$$

We then calculate the probability of state j at the first time point, T_1, given that the branch begins in state i and ends in state k, as:

$$P(j \mid i, k) \propto \mathbf{P}_{ij,1} \times \left[\prod_{l=2}^{m} \mathbf{P}_l\right]_{jk},$$

where the first term is the probability of transitioning from state i (the state at the beginning of the branch) to state j at the first time point, and the second term is the probability of transitioning from state j to state k (the state at the end of the branch) over the remaining time intervals. We compute this for each state j, and sample the state in proportion to these probabilities. We then repeat this process for each remaining time point, recursively conditioning on the state sampled at the previous time point and the state at the end of the branch.

Second, we simulate histories within each interval. For a given time interval, we simulate histories conditional on the start and end states generated in the first step using the uniformization algorithm described by Hobolth and Stone (2009).

Assessing the Absolute Fit of Interval-specific Phylodynamic Models

For a given phylodynamic study, we might wish to consider several candidate interval-specific models (where each candidate model specifies a unique number of intervals, set of interval boundaries, and/or interval-specific parameters). Comparing the fit of these competing phylodynamic models to the data offers two benefits: (1) confirming that our inference model adequately describes the process that gave rise to data will improve the accuracy of the corresponding inferences (i.e., estimates of relative and/or average dispersal-rate parameters and viral dispersal histories), and; (2) comparing alternative models provides a means to objectively test hypotheses regarding the impact of events on the history of viral dispersal (i.e., by assessing the relative fit of data to competing models that include/exclude the impact of a putative event on the average and/or relative viral dispersal rates). We can assess the *relative* fit of two or more candidate phylodynamic models to a given dataset using Bayes factors; this requires that we first estimate the marginal likelihood for each model (which represents the average fit of a model to a dataset), and then compute the Bayes factor as twice the difference in the log marginal likelihoods of the competing models (Kass and Raftery 1995).

However, even the best candidate model may fail to provide an adequate description of the process that gave

rise to our study data. We can leverage our ability to simulate histories under the interval-specific models to develop new methods to assess the *absolute* fit of a candidate model using posterior-predictive assessment (Gelman et al. 1996). This Bayesian approach for assessing model adequacy is based on the following premise: if our inference model provides an adequate description of the process that gave rise to our observed data, then we should be able to use that model to simulate datasets that resemble our original data. The resemblance between the observed and simulated datasets is quantified using a summary statistic. Accordingly, posterior-predictive simulation requires: (1) the ability to simulate geographic datasets under interval-specific models for a given set of parameter values, and; (2) summary statistics that allow us to compare the resulting simulated datasets to the observed dataset. We describe each of these components below.

Simulating under interval-specific phylodynamic models

We draw m random samples from the joint posterior distribution of the model; each sample i consists of a fully specified phylodynamic model, $\theta_i = \{\Psi_i, \mathbf{Q}_i, \mu_i\}$. For each sample, we simulate a new geographic dataset on the sampled tree, Ψ_i, given the sampled parameters of the geographic model, $\{\mathbf{Q}_i, \mu_i\}$; we label the newly simulated dataset G_i^{sim}. Under a constant-rate phylodynamic model, we simulate full dispersal histories forward in time over a tree using the `sim.history()` function in the `R` package `phytools` (Revell 2012). We implemented an extension of the `sim.history()` function to simulate dispersal histories under interval-specific phylodynamic models. These functions allow us to perform posterior-predictive simulation to assess the adequacy of both the constant-rate and the interval-specific models.

Summary statistics

We define a summary statistic, which we generically denote $T(G \mid \theta_i)$, where G is either the simulated or observed dataset. For each simulated dataset, we compute a discrepancy statistic,

$$D_i = T(G_i^{\text{sim}} \mid \theta_i) - T(G^{\text{obs}} \mid \theta_i),$$

where G^{obs} is the observed geographic dataset and G^{sim} is a simulated dataset. We developed two summary statistics to assess the adequacy of interval-specific phylodynamic models: (1) the *parsimony statistic*, and; (2) the *tipwise-multinomial statistic*. The parsimony statistic is calculated as the difference in the parsimony score for the observed areas and the simulated areas across the tips of the tree (where the parsimony score is the minimum number of dispersal events required to explain the distribution of

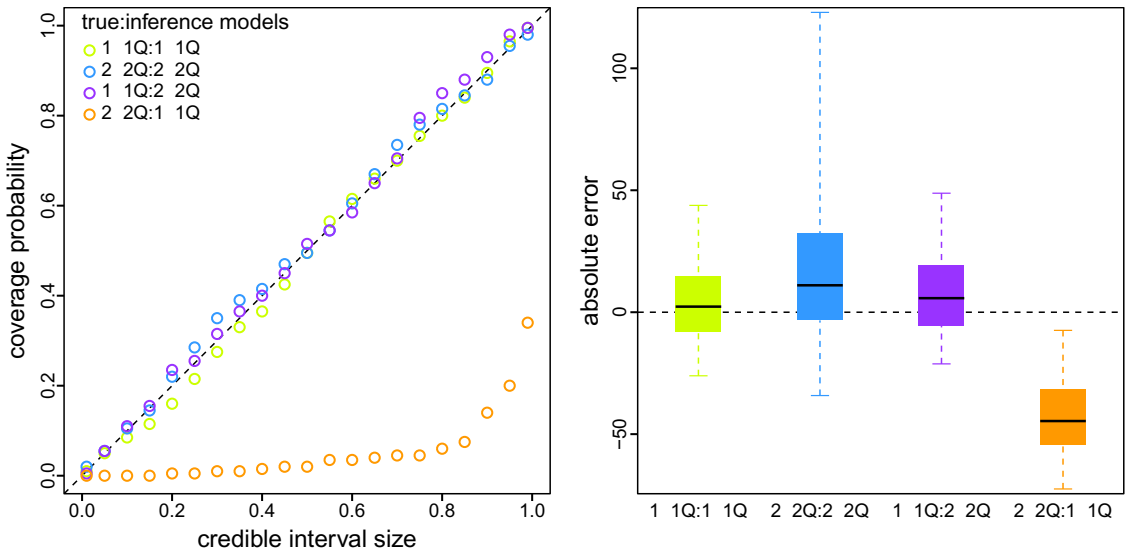

Fig. 3. Simulation demonstrates that reliable inference of viral dispersal history requires a correctly specified phylodynamic model. We simulated 200 geographic datasets under each of two models: one that assumed a constant μ and $\mathbf{Q}$ ($1\mu 1\mathbf{Q}$), and one that allowed μ and $\mathbf{Q}$ to vary over two intervals ($2\mu 2\mathbf{Q}$). For each simulated dataset, we separately inferred the total number of dispersal events under each model, resulting in four true:inference model combinations ($1\mu 1\mathbf{Q}$:$1\mu 1\mathbf{Q}$, $2\mu 2\mathbf{Q}$:$2\mu 2\mathbf{Q}$, $1\mu 1\mathbf{Q}$:$2\mu 2\mathbf{Q}$, and $2\mu 2\mathbf{Q}$:$1\mu 1\mathbf{Q}$). (Left) For each combination of true and inference model, we computed the coverage probability (the frequency with which the true number of dispersal events was contained in the corresponding X% credible interval; y-axis) as a function of the size of the credible interval (x-axis). When the model is true, we expect the coverage probability to be equal to the size of the credible interval (Cook et al. 2006). As expected, coverage probabilities fall along the one-to-one line when the model is correctly specified (green and blue). Moreover, coverage probabilities are also appropriate when the inference model is overspecified (i.e., the inference model includes interval-specific parameters not included in the true model; purple). However, coverage probabilities are extremely unreliable when the inference model is underspecified (i.e., the inference model excludes interval-specific parameters of the true model; orange). (Right) For each true:inference model combination, we summarized the absolute error (estimated minus true number of dispersal events) as boxplots (median [horizontal bar], 50% probability interval [boxes], and 95% probability interval [whiskers]). Again, when the model is underspecified (orange) inferences are strongly biased compared to those under the correctly specified (green and blue) and overspecified (purple) models.

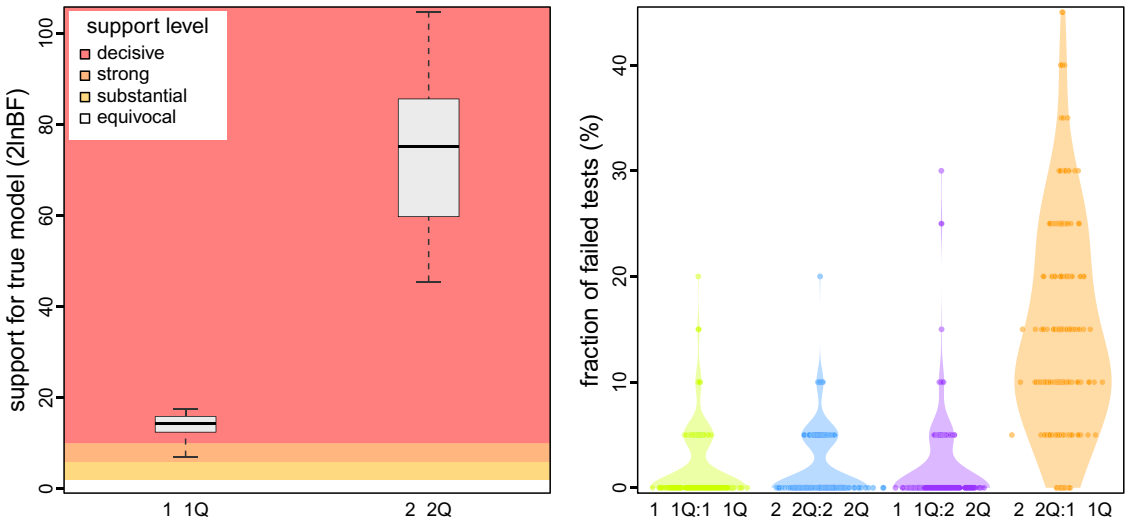

Fig. 4. Simulation demonstrates our ability to accurately identify a correctly specified phylodynamic model. We assessed the relative and absolute fit of alternative models to the simulated datasets described in figure 3. (Left) For each simulated dataset, we compared the relative fit of the true and alternative models using Bayes factors. The boxplots summarize Bayes factors for datasets simulated under the constant-rate ($1\mu 1\mathbf{Q}$, left) and interval-specific ($2\mu 2\mathbf{Q}$, right) models, which demonstrate that we are able to decisively identify the true phylodynamic model. (Right) For each combination of true:inference model, we assessed absolute model fit using posterior-predictive simulation with a set of 20 summary statistics. Each dot represents the fraction of those 20 summary statistics for which the corresponding inference model provides an inadequate fit to a single simulated dataset. The violin plots summarize the distribution of these values for all datasets under each true:inference model combination. As expected, the true model is overwhelmingly inferred to be adequate (green and blue). Encouragingly, model overspecification appears to have a negligible impact on model adequacy (purple). By contrast, an underspecified model severely impacts model adequacy (orange).

areas across the tips of a tree). We compute parsimony scores using the `parsimony()` function in the `R` package, `phangorn` (Schliep 2010). The tipwise-multinomial statistic is inspired by the multinomial statistic that was proposed by Goldman (1993) and later used by Bollback (2002) to assess the adequacy (absolute fit) of substitution models to sequence alignments. Our tipwise statistic treats the set of states (areas) across the tips of the tree as an outcome of a multinomial trial. Specifically, we calculate the tipwise-multinomial statistic as the difference in the multinomial probabilities for the observed set of areas versus the simulated set of areas across the tips of the tree. We calculate each multinomial probability as:

$$T(G \mid \theta_i) = \sum_{i=1}^{k} n_i \ln (n_i/n),$$

where n is the number of tips in the tree, and n_i is the number of tips that occur area i.

Time-slice summary statistics

To assess the ability of phylodynamic models to describe the temporal distribution of dispersal events, we extend the parsimony and tipwise-multinomial summary statistics to assess time slices of the geographic history. (Note that the time slices that we define for summary statistics are distinct from the intervals specified in an interval-specific model. The time slices are motivated to better assess the adequacy of a phylodynamic model, whereas the intervals are motivated to accommodate variation in dispersal dynamics in the empirical data. Accordingly, we might use time-slice summary statistics to assess the adequacy of both constant-rate and interval-specific phylodynamic models.) We calculate these summary statistics for k pre-specified time slices, resulting in k parsimony statistics and k tipwise-multinomial statistics for each simulated dataset. We compute the time-slice variant of the parsimony summary statistic as follows: (1) we first infer the most-parsimonious dispersal history (i.e., the minimum number of dispersal events) for a given simulated dataset and the observed dataset using the `ancestral.pars()` function in the `R` package, `phangorn` (Schliep 2010); (2) we then assign each inferred dispersal event to one of the k time slices based on the time span of the branch along which the dispersal event was inferred (when a dispersal event is inferred to occur along a branch that spans two or more time slices, we locate the event uniformly along the branch, and then assign it to the corresponding slice), and finally; (3) we compute the difference in the number of dispersal events between the simulated and observed dataset for each time slice. We compute the time-slice variant of the tipwise-multinomial summary statistic in a similar manner; i.e., we first find the set of tips in each time slice, and then compute the tipwise-multinomial statistic for that time slice (as described above) for the corresponding set of tips. Further details regarding the computation of these summary statistics are available in an `R` script provided in our GitHub and Dryad repositories.

Simulation Study

We performed a simulation study to explore the statistical behavior of the interval-specific phylodynamic models.

Specifically, the goals of this simulation study are to assess: (1) our ability to perform reliable inference under interval-specific models; (2) the impact of model misspecification, and; (3) our ability to identify the correct model. To this end, we simulated 200 geographic datasets under each of two models: the first assumes a constant μ and **Q** ($1\mu1$**Q**), and the second allows μ and **Q** to vary over two intervals ($2\mu2$**Q**). The parameter values used in the simulation are from empirical analyses of a SARS-CoV-2 dataset (with 1271 viral sequences sampled from three coarsely aggregated geographic areas) under each corresponding model. For each simulated dataset, we separately inferred the history of viral dispersal under each model, resulting in four true:inference model combinations: $1\mu1$**Q**:$1\mu1$**Q**, $2\mu2$**Q**:$2\mu2$**Q**, $1\mu1$**Q**:$2\mu2$**Q**, and $2\mu2$**Q**:$1\mu1$**Q**. We provide detailed descriptions of the simulation analyses and results in supplementary section S2, Supplementary Material online.

Ability to Reliably Estimate Parameters of Interval-specific Phylodynamic Models

Interval-specific phylodynamic models are inherently more complex than their constant-rate counterparts, and therefore contain many more parameters that must be inferred from geographic datasets that contain minimal information; these datasets only include a single observation (i.e., the area in which each virus was sampled). These considerations raise concerns about our ability to reliably estimate parameters of interval-specific phylodynamic models. Encouragingly, when the inference model is correctly specified (i.e., where both the true and inference models include [or exclude] interval-specific parameters, $2\mu2$**Q**:$2\mu2$**Q** and $1\mu1$**Q**:$1\mu1$**Q**), our simulation study demonstrates that estimates under interval-specific models are as reliable as those under constant-rate models (fig. 3, green, blue; see the online version for color). Moreover, when the inference model is overspecified (i.e., it includes interval-specific parameters not included in the true model) inferences are comparable to those under correctly specified models (fig. 3, purple). However, when the inference model is underspecified (i.e., it excludes interval-specific parameters of the true model) inferences are severely biased estimates (fig. 3, orange).

Ability to Accurately Identify an Appropriately Specified Phylodynamic Model

Our simulation study demonstrates the importance of identifying scenarios where an inference model is underspecified; failure to accommodate interval-specific variation in the study data can severely bias parameter estimates. Fortunately, our simulation study demonstrates that we can reliably identify when a given model is correctly specified, overspecified, or underspecified using a combination of Bayes factors (to assess the relative fit of competing models to the data; fig. 4, left) and posterior-predictive simulation (to assess the absolute fit of each candidate model to the data; fig. 4, right). Using a combination of Bayes factors and posterior-predictive simulation allows us to not only identify the best of the candidate models, but also to ensure that the best model provides an adequate description of the true process that gave rise to our study data.

Empirical Application

We demonstrate our new phylodynamic methods with analyses of all publicly available SARS-CoV-2 genomes sampled during the early phase of the COVID-19 pandemic (with 2598 viral genomes collected from 23 geographic areas between December 24, 2019 and March 8, 2020 [downloaded from GISAID, Shu and McCauley 2017]). We used our study dataset to estimate the parameters of —and assess the relative and absolute fit to—nine candidate phylodynamic models. These models assign interval-specific parameters—for the average rate of viral dispersal, μ, and/or relative rates of viral dispersal, **Q**—to one, two, four, or five pre-specified time intervals; i.e., $1\mu1$**Q**, $2\mu1$**Q**, $1\mu2$**Q**, $2\mu2$**Q**, $4\mu1$**Q**, $1\mu4$**Q**, $4\mu4$**Q**, $5\mu5$**Q**, and $5\mu5$**Q***. We specified interval boundaries based on external information regarding events within the study period that might plausibly impact viral dispersal dynamics, including: (A) start of the Spring Festival travel season in China (the highest annual period of domestic travel, January 12); (B) onset of mitigation measures in Hubei province, China (January 26); (C) onset of international air-travel restrictions against China (February 2), and; (D) relaxation of domestic travel restrictions in China (February 16). Phylodynamic models with two intervals include event C, models with four intervals include events A, C, and D, and the $5\mu5$**Q** model includes all four events. The final candidate model, $5\mu5$**Q***, includes five arbitrary and uniform (14-day) intervals. We provide detailed descriptions of our empirical data collection, analyses, and results in supplementary section S3, Supplementary Material online.

An Interval-specific Model Best Describes Viral Dispersal in the Early Phase of the Pandemic

Our phylodynamic analyses of the SARS-CoV-2 dataset reveal that the early phase of the COVID-19 pandemic exhibits significant variation in both the average and relative rates of viral dispersal over four time intervals. Bayes factor comparisons (fig. 5, left) demonstrate that the $4\mu4$**Q** interval-specific model is decisively preferred both over all less complex candidate models—including models that allow *either* the average dispersal rate *or* relative dispersal rates to vary over the same four intervals ($4\mu1$**Q** and $1\mu4$**Q**, respectively)—and also over more complex candidate models ($5\mu5$**Q**, and $5\mu5$**Q***). Posterior-predictive analyses (fig. 5, right) demonstrate that the preferred model, $4\mu4$**Q**, also provides an adequate description of the process that gave rise to our SARS-CoV-2 dataset. Below, we will use the preferred ($4\mu4$**Q**) interval-specific phylodynamic model to explore various aspects of viral dispersal during the early phase of the COVID-19 pandemic and—for the purposes of comparison—we also present corresponding results inferred using the (underspecified) constant-rate ($1\mu1$**Q**) phylodynamic model.

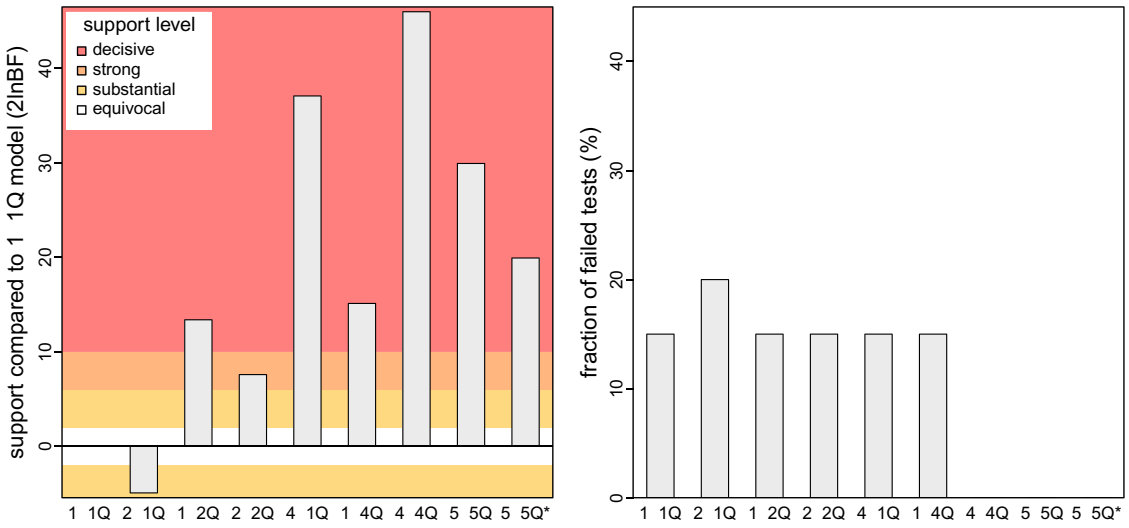

Fig. 5. An interval-specific model provides the best relative and absolute fit to our SARS-CoV-2 dataset. We assessed the relative and absolute fit of nine candidate phylodynamic models to our study dataset (comprised of all publicly available SARS-CoV-2 genomes from the early phase of the COVID-19 pandemic). (Left) We compared the relative fit of each candidate model to the constant-rate ($1\mu1\mathbf{Q}$) phylodynamic model using Bayes factors, which indicate that the $4\mu4\mathbf{Q}$ interval-specific model outcompetes both less complex and more complex models. (Right) We performed posterior-predictive simulation for each candidate model using 20 summary statistics, plotting the fraction of those summary statistics indicating that a given candidate model was inadequate. Our results indicate that three candidate models ($4\mu4\mathbf{Q}$, $5\mu5\mathbf{Q}$, and $5\mu5\mathbf{Q}^*$) provide an adequate fit to our SARS-CoV-2 dataset. The simplest of these adequate models ($4\mu4\mathbf{Q}$) also provides the best relative fit. Collectively, these results identify the $4\mu4\mathbf{Q}$ model as the clear choice for phylodynamic analyses of our study dataset.

Variation in Global Viral Dispersal Rates

Between late 2019 and early March, 2020, COVID-19 emerged (in Wuhan, China) and established a global distribution—with reported cases in 83% of the study areas by this date (WHO 2020)—despite the implementation of numerous intervention efforts to slow the spread of the causative SARS-CoV-2 virus (Hsiang et al. 2020). This crucial early phase of the pandemic provides a unique opportunity to explore the dispersal dynamics that led to the worldwide establishment of the virus and to assess the efficacy of key public-health measures to mitigate the spread of COVID-19. The constant-rate ($1\mu1\mathbf{Q}$) model infers a static rate of global viral dispersal throughout the study period (fig. 6, orange). By contrast, inferences under the preferred ($4\mu4\mathbf{Q}$) model reveal significant variation in global viral dispersal rates over four intervals, exhibiting both increases and decreases over the early phase of the pandemic (fig. 6, dark blue). The significant decrease in the global viral dispersal rate between the second and third interval (with a boundary at February 2) coincides with the initiation of international air-travel bans with China (imposed by 34 countries and nation states by this date). To further explore the possible impact of the air-travel ban on the global spread of COVID-19, we inferred daily rates of global viral dispersal under a more granular interval-specific model ($71\mu4\mathbf{Q}$; fig. 6, light blue). Our estimates of daily rates of global viral dispersal are significantly correlated with independent information on daily global air-travel volume (fig. 6, dashed) over the interval from Jan. 31 (when the virus first achieved a cosmopolitan distribution; WHO 2020) to the end of our study period (see supplementary section S3.3, Supplementary Material online for detailed descriptions of the correlation test and results).

Variation in Viral Dispersal Routes

In addition to revealing differences in the global viral dispersal rate, our interval-specific phylodynamic models allow us to explore how relative dispersal rates vary through time. Specifically, our analyses allow us to identify the dispersal routes by which the SARS-CoV-2 virus achieved a global distribution during the early phase of the COVID-19 pandemic. We focus on dispersal routes involving China both because it was the point of origin, and because it was the area against which travel bans were imposed. Inferences under the constant-rate ($1\mu1\mathbf{Q}$) and preferred ($4\mu4\mathbf{Q}$) phylodynamic models imply strongly contrasting viral dispersal dynamics (fig. 7). In contrast to the invariant set of dispersal routes identified by the constant-rate model, the preferred interval-specific model reveals that the number and intensity of dispersal routes varied significantly over the four intervals, with a sharp decrease in the number of dispersal routes following the onset of air-travel bans on February 2. Moreover, the constant-rate model infers one spurious dispersal route, while failing to identify six significant dispersal routes; the preferred model implies a more significant role for Hubei as a source of viral spread in the first and second intervals and reveals additional viral dispersal routes originating from China in the third and fourth intervals. The patterns of variation in dispersal routes among all 23 study areas are similar to—but more pronounced than—those involving China; e.g., where the constant-rate model infers a total of nine spurious dispersal

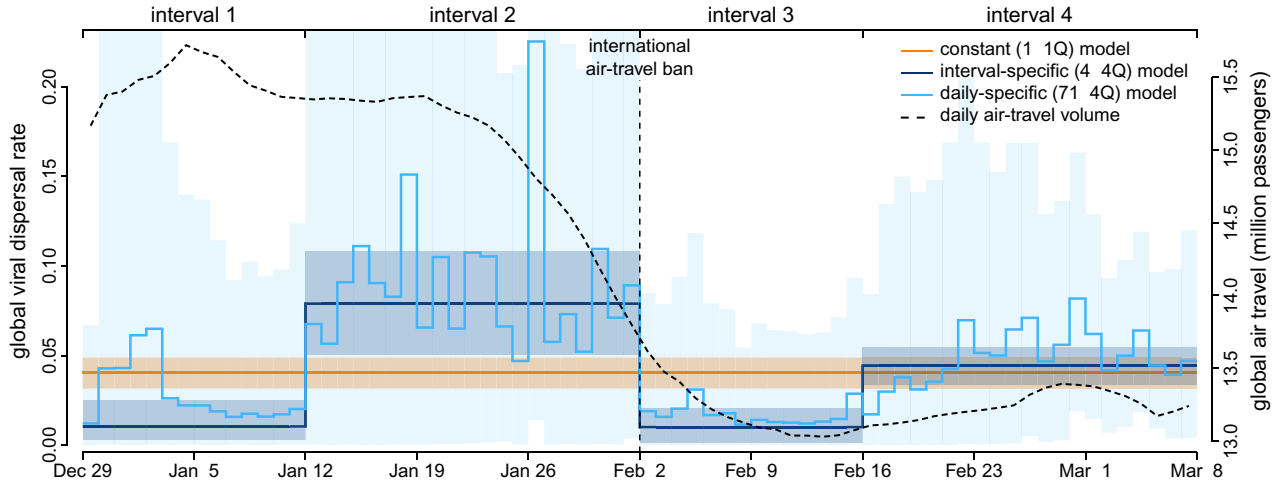

Fig. 6. Patterns and correlates of variation in global viral dispersal rate early in the COVID-19 pandemic. The COVID-19 pandemic emerged in Wuhan, China, in late 2019, and established a global distribution by March 8, 2020. Our phylodynamic analyses of this critical early phase of the pandemic provide estimates of the average rate of viral dispersal across all 23 study areas, μ (posterior mean [solid lines], 95% credible interval [shaded areas]). By assumption, the constant-rate ($1\mu 1\mathbf{Q}$) model infers a static rate of global viral dispersal (orange). By contrast, the preferred interval-specific ($4\mu 4\mathbf{Q}$) model reveals significant variation in the global viral dispersal rate (dark blue). Notably, the global viral dispersal rate decreases sharply on February 2, which coincides with the onset of international air-travel bans with China. The efficacy of these air-travel restrictions is further corroborated by estimates of daily global viral dispersal rates (light blue)—inferred under a more granular, interval-specific ($71\mu 4\mathbf{Q}$) model—that are significantly correlated with independent information on daily global air-travel volume (dashed line, obtained from FlightAware).

routes, and the interval-specific model reveals a total of ten significant dispersal routes that were not detected by the constant-rate model (supplementary figs. S16–S17, Supplementary Material online).

Variation in the Number of Viral Dispersal Events

Our phylodynamic analyses also allow us to infer the number of SARS-CoV-2 dispersal events between areas during the early phase of the COVID-19 pandemic. Here, we focus on the number of viral dispersal events originating from China because it was the point of origin and primary source of SARS-CoV-2 spread early in the pandemic. The constant-rate ($1\mu 1\mathbf{Q}$) and preferred ($4\mu 4\mathbf{Q}$) phylodynamic models infer distinct trends in—and support different conclusions regarding the impact of mitigation measures on—the number of viral dispersal events out of China. The constant-rate model infers a gradual decrease in the number of dispersal events from late Jan. through mid-February (fig. 8, orange). By contrast, the preferred interval-specific model reveals a sharp decrease in the number of dispersal events on February 2, which coincides with the onset of air-travel bans imposed against China (fig. 8, blue). Moreover, the preferred phylodynamic model infers an uptick in the number of viral dispersal events on February 17 (not detected by the constant-rate model), which coincides with the lifting of domestic travel restrictions within China (except for Hubei, where the travel restrictions were enforced through late March).

Discussion

Phylodynamic methods increasingly inform our understanding of the spatial and temporal dynamics of viral spread. The vast majority of discrete-geographic phylodynamic studies assume—despite direct (and compelling) evidence to the contrary—that disease outbreaks are intrinsically constant: ≈98% of all such studies are based on the constant-rate models. These considerations have motivated previous extensions of phylodynamic models that allow *either* the average (Membrebe et al. 2019) *or* relative (Bielejec et al. 2014) dispersal rates to vary, and our development of more complex phylodynamic models that allow *both* the average and relative dispersal rates to vary independently over two or more pre-specified intervals. By accommodating ubiquitous temporal variation in the dynamics of disease outbreaks—and by allowing us to incorporate independent information regarding events that may impact viral dispersal—our new interval-specific phylodynamic models are more realistic (providing a better description of the processes that gives rise to empirical datasets), thereby enhancing the accuracy of our epidemiological inferences based on these models.

Our simulation study demonstrates that (in principle): (1) we are able to accurately identify when phylodynamic models are correctly specified, overspecified, or underspecified (fig. 4); (2) when the phylodynamic model is correctly specified, we are able to reliably estimate parameters of these more complex interval-specific models (fig. 3), and; (3) when the phylodynamic model is underspecified, failure to accommodate interval-specific variation in the study data can bias parameter estimates and mislead inferences about viral dispersal history based on those biased estimates (fig. 3).

Our empirical study of SARS-CoV-2 data from the early phase of the COVID-19 pandemic demonstrates that (in practice): (1) our interval-specific phylodynamic model

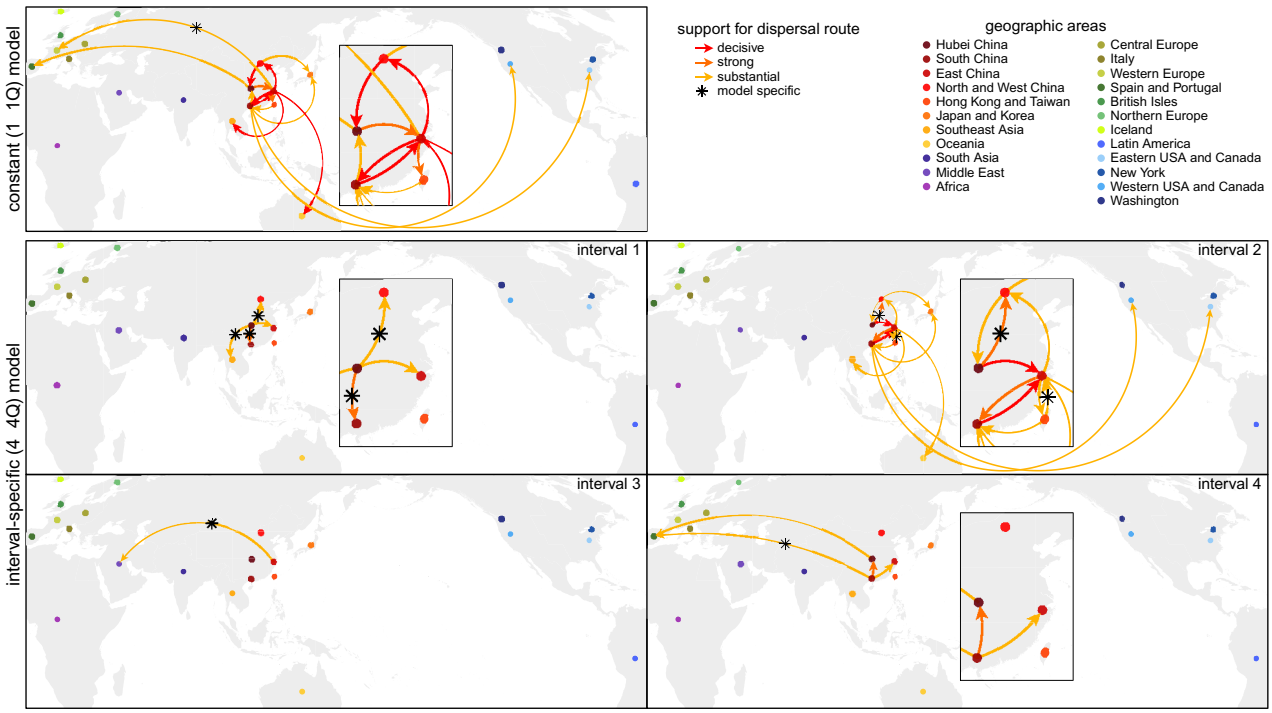

Fig. 7. Variation in viral dispersal routes involving China during the early phase of the pandemic. Arrows indicate routes inferred to play a significant role in viral dispersal to/from China during the early phase of the COVID-19 pandemic; colors indicate the level of evidential support for each dispersal route (as 2 ln Bayes factors). We focus on dispersal routes involving China both because it was the point of origin, and because it was the area against which travel bans were imposed. The number, duration, and significance of dispersal routes inferred under the constant-rate ($1\mu1\mathbf{Q}$) model differ strongly from those inferred under the preferred ($4\mu4\mathbf{Q}$) interval-specific model. By assumption, the constant-rate ($1\mu1\mathbf{Q}$) model implies an invariant set of dispersal routes. By contrast, the preferred ($4\mu4\mathbf{Q}$) interval-specific model reveals that the number and intensity of dispersal routes varied over the four intervals. The first interval (November 17–January 12) is dominated by dispersal from Hubei to other areas in China, and the second interval (January 12–February 2) exhibits more widespread international dispersal originating from China. The third interval (February 2–16)—immediately following the onset of international air-travel bans with China—exhibits a sustained reduction in the number of dispersal routes. Note that the constant-rate model infers a spurious dispersal route from East China to West Europe. Conversely, the preferred interval-specific model reveals six significant dispersal routes (not detected under the constant-rate model) that imply a more significant role for Hubei as a source of viral spread in the first and second intervals, and also reveals additional dispersal routes emanating from China (to the Middle East in the third interval and to Spain/Portugal in the fourth interval).

(where *both* the global rate of viral dispersal *and* the relative rates of viral dispersal vary over four distinct intervals) significantly improves the relative and absolute fit to our study dataset compared to constant-rate phylodynamic models (Lemey et al. 2009; Edwards et al. 2011) and to phylodynamic models that allow *either* the average dispersal rate (Membrebe et al. 2019) *or* the relative dispersal rates (Bielejec et al. 2014) to vary over the same four intervals; (2) the preferred interval-specific phylodynamic model provides qualitatively different insights on key aspects of viral dynamics during the early phase of the pandemic—on global rates of viral dispersal (fig. 6), viral dispersal routes (fig. 7), and the number of viral dispersal events (fig. 8)—compared to conventional estimates based on constant-rate (and underspecified) phylodynamic models, and; (3) inferences under the preferred interval-specific phylodynamic model support qualitatively different conclusions regarding the impact of mitigation measures to limit the spread of the COVID-19 pandemic; e.g., the variation in global viral dispersal rate, viral dispersal routes, and number of viral dispersal events revealed by the interval-specific model (but masked by the constant-rate model) collectively support the efficacy of the international air-travel bans in slowing the progression of the COVID-19 pandemic.

Our interval-specific models promise to enhance the accuracy of phylodynamic inferences not only by virtue of their increased realism, but also by allowing us to incorporate additional information (related to events in the history of disease outbreaks) in our phylodynamic inferences. The ability to incorporate independent/external information is particularly valuable for phylodynamic inference—where many parameters must be estimated from datasets with limited information—which has also motivated the development of other innovative phylodynamic approaches for incorporating external information (Lemey et al. 2014; Bielejec et al. 2016). The potential benefit of harnessing external information is evident in our empirical study: our inference model —$4\mu4\mathbf{Q}$, with four intervals that we specified based on external evidence regarding events that might plausibly impact viral dispersal dynamics—is decisively preferred (2 ln BF = 27.3) over a substantially more complex model, $5\mu5\mathbf{Q}^*$, with five *arbitrarily* specified (14-day) intervals.

Importantly, comparison of alternative phylodynamic models provides a powerful framework for testing

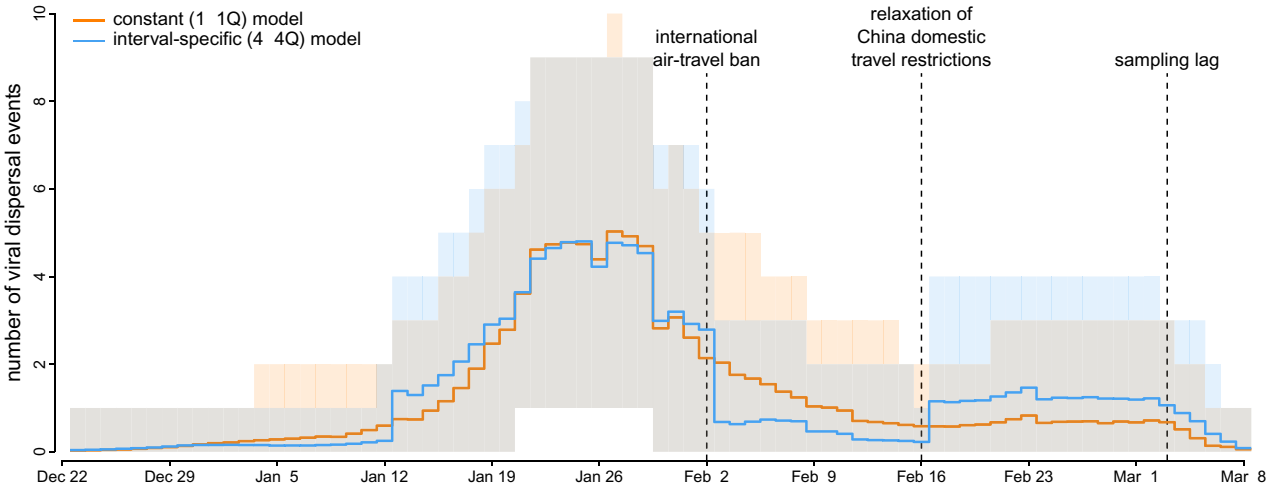

Fig. 8. Variation in the number of viral dispersal events out of China early in the COVID-19 pandemic. Our phylogenetic analyses of SARS-CoV-2 genomes sampled during the early phase of the COVID-19 pandemic allow us to estimate the number of viral dispersal events from China to all other study areas (posterior mean [solid lines], 95% credible interval [shaded areas]). The constant-rate ($1\mu 1\mathbf{Q}$) model implies that the number of viral dispersal events emanating from China remained relatively high following the onset of international air-travel bans on February 2 (orange). By contrast, the preferred interval-specific ($4\mu 4\mathbf{Q}$) model reveals that the number of viral dispersal events emanating from China decreased sharply on February 2 (blue), which supports the efficacy of these international air-travel restrictions. The preferred model also infers an uptick in the number of viral dispersal events on February 17 (not detected by the constant-rate model), which coincides with the relaxation of domestic travel restrictions in China. Note that sampling lag causes the number of dispersal events near the end of the sampling period to be underestimated.

hypotheses about the impact of various events (i.e., assessing the efficacy of mitigation measures) on viral dispersal dynamics. Our empirical study allows us, for example, to assess the impact of domestic mitigation measures imposed in the Hubei province of China. This simply involves comparing the relative fit of our data to two candidate phylodynamic models: $4\mu 4\mathbf{Q}$ and $5\mu 5\mathbf{Q}$. The $5\mu 5\mathbf{Q}$ model adds an interval (corresponding to the onset of the Hubei lockdown on January 26) to the otherwise identical $4\mu 4\mathbf{Q}$ model. In contrast to the international air-travel ban, this domestic mitigation measure does not appear to have significantly impacted global SARS-CoV-2 dispersal dynamics: the $5\mu 5\mathbf{Q}$ model is decisively rejected when compared to the $4\mu 4\mathbf{Q}$ model ($2 \ln \mathrm{BF} = -15.9$).

We have focused on interval-specific models where each interval involves a change in both the average and relative dispersal rates. For example, the scenario depicted in figure 1 involves two events that define three intervals, where both $\mathbf{Q}$ and μ are impacted by each event, such that the interval-specific parameters are $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$ and (μ_1, μ_2, μ_3). However, our interval-specific models also allow the average and relative dispersal rates to vary *independently* across intervals. For example, under an alternative scenario for figure 1, the first event may have impacted both the relative and average dispersal rates, $\mathbf{Q}$ and μ, whereas the second event may have only changed the relative dispersal rates, $\mathbf{Q}$; in this case, the interval-specific parameters would be $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$ and (μ_1, μ_2, μ_2). Allowing dispersal rates to vary independently enables these models to accommodate more complex patterns of variation in empirical datasets (and thereby improve estimates from these more realistic models), and also provides tremendous flexibility for testing hypotheses about the impact of various mitigation measures on *either* the relative and/or average rates of viral dispersal.

Nevertheless, this flexibility comes at a cost: interval-specific models are inherently more complex than their constant-rate counterparts, with many parameters that must be estimated from minimal data (i.e., the geographic location of each virus). Accordingly, careful model selection and validation is necessary to avoid specification of an over-parameterized model (although the Kullback–Leibler divergence between the posterior and the prior reveals similar amounts of information gain under the constant-rate and interval-specific models; see supplementary section S3.3, Supplementary Material online for detailed descriptions of the KL-divergence computation and results). Moreover, the space of phylodynamic models expands rapidly as we increase the number of intervals. For a model with three intervals, for example, we can specify five allocations for the average dispersal rate parameter, μ—(μ_1, μ_1, μ_1), (μ_1, μ_1, μ_2), (μ_1, μ_2, μ_1), (μ_1, μ_2, μ_2), and (μ_1, μ_2, μ_3)—and, similarly, five allocations for the relative dispersal rate parameter, $\mathbf{Q}$: $(\mathbf{Q}_1, \mathbf{Q}_1, \mathbf{Q}_1)$, $(\mathbf{Q}_1, \mathbf{Q}_1, \mathbf{Q}_2)$, $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_1)$, $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_2)$, and $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$. We can therefore specify 25 unique three-interval phylodynamic models (representing all combinations of the two parameter-allocation vectors), 225 unique four-interval models, 2704 unique five-interval models, 41,209 unique six-interval models, etc. Accordingly, the effort required to identify the best interval-specific phylodynamic model quickly becomes prohibitive, particularly because this search requires that we estimate the marginal likelihood for each candidate model using computationally intensive methods (Xie et al. 2011; Baele et al. 2012). Nevertheless, our interval-specific models establish a foundation for developing

more computationally efficient methods; e.g., we could pursue a finite-mixture approach (Kazmi and Rodrigue 2019) that averages inferences of dispersal dynamics over the space of all possible interval-specific phylodynamic models with a given number of intervals.

We are optimistic that—by increasing (and providing a means to assess) model realism, incorporating additional information, and providing a powerful and flexible means to test alternative models/hypotheses—our phylodynamic methods will greatly enhance our ability to understand the dynamics of viral spread, and thereby inform policies to mitigate the impact of disease outbreaks.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Louis du Plessis, two anonymous reviewers, and the editor for providing thoughtful comments that greatly improved the manuscript. This research was supported by the National Science Foundation grants DEB-0842181, DEB-0919529, DBI-1356737, and DEB-1457835 awarded to B.R.M.; the National Institutes of Health grant RO1GM123306-S awarded to B.R.; and research awards from the UC Davis Center for Population Biology to J.G.

Data and Code Availability

GISAID accession IDs of the SARS-CoV-2 sequences used in this study, as well as the flight-volume data (obtained from FlightAware, LLC) and intervention-measure data, are maintained in the GitHub repository (https://github.com/jsigao/interval˙specific˙phylodynamic˙models˙supparchive) and archived in the Dryad repository (https://doi.org/10.25338/B89P9K). Our repositories also contain `BEAST` XML scripts used to perform the phylodynamic analyses, `R` scripts used to perform simulations and post processing, and a modified version of the `BEAST` program used for some of the analyses in this study.

References

Alpert T, Brito AF, Lasek-Nesselquist E, Rothman J, Valesano AL, MacKay MJ, Petrone ME, Breban MI, Watkins AE, Vogels CBF, *et al.* 2021. Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. *Cell* **184**:2595–2604.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. **29**:2157–2167.

Baele G, Suchard MA, Rambaut A, Lemey P. 2017. Emerging concepts of data integration in pathogen phylodynamics. *Syst Biol*. **66**: e47–e65.

Bedford T, Greninger AL, Roychoudhury P, Starita LM, Famulare M, Huang M-L, Nalla A, Pepper G, Reinhardt A, Xie H, *et al.* 2020. Cryptic transmission of SARS-CoV-2 in Washington state. *Science* **370**:571–575.

Bielejec F, Baele G, Rodrigo AG, Suchard MA, Lemey P. 2016. Identifying predictors of time-inhomogeneous viral evolutionary processes. *Virus Evol*. **2**:vew023.

Bielejec F, Lemey P, Baele G, Rambaut A, Suchard MA. 2014. Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. *Syst Biol*. **63**:493–504.

Bollback JP. 2002. Bayesian model adequacy and choice in phylogenetics. *Mol Biol Evol*. **19**:1171–1180.

Bollback JP. 2006. Simmap: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinform*. **7**:88.

Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, Mellan TA, du Plessis L, Pereira RHM, Sales FCS, *et al.* 2020. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* **369**:1255–1260.

Cook SR, Gelman A, Rubin DB. 2006. Validation of software for Bayesian models using posterior quantiles. *J Comput Graph Stat*. **15**:675–692.

WHO. 2020. Coronavirus disease (COVID-19) situation reports [cited 2020 Dec 19]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, *et al.* 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**:eabg3055.

Dellicour S, Durkin K, Hong SL, Vanmechelen B, Martí-Carreras J, Gill MS, Meex C, Bontems S, André E, Gilbert M, *et al.* 2021. A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. *Mol Biol Evol*. **38**:1608–1613.

De Maio N, Wu CH, O'Reilly KM, Wilson D. 2015. New routes to phylogeography: a Bayesian structured coalescent approximation. *PLoS Genet*. **11**:e1005421.

Douglas J, Mendes FK, Bouckaert R, Xie D, Jiménez-Silva CL, Swanepoel C, de Ligt J, Ren X, Storey M, Hadfield J, *et al.* 2021. Phylodynamics reveals the role of human travel and contact tracing in controlling the first wave of COVID-19 in four island nations. *Virus Evol*. **7**:veab052.

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

du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, Raghwani J, Ashworth J, Colquhoun R, Connor TR, *et al.* 2021. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**:708–712.

Edwards C, Suchard M, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N, *et al.* 2011. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr Biol*. **21**:1251–1258.

Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, Vogels CBF, Brito AF, Alpert T, Muyombwe A, *et al.* 2020. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell*. **181**:990–996.

Gelman A, Meng XL, Stern H. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Stat Sin*. **6**: 733–807.

Gill MS, Lemey P, Bennett SN, Biek R, Suchard MA. 2016. Understanding past population dynamics: Bayesian coalescent-based modeling with covariates. *Syst Biol*. **65**:1041–1056.

Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. 2013. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol*. **30**:713–724.

Gill MS, Tung Ho LS, Baele G, Lemey P, Suchard MA. 2017. A relaxed directional random walk model for phylogenetic trait evolution. *Syst Biol*. **66**:299–319.

Goldman N. 1993. Statistical tests of models of DNA substitution. *J Mol Evol*. **36**:182–198.

Hobolth A, Stone EA. 2009. Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution. *Ann Appl Stat*. **3**:1204.

Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, Druckenmiller H, Huang LY, Hultgren A, Krasovich E, *et al.* 2020. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**:262–267.

Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Syst Biol*. **52**:131–158.

Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc*. **90**: 773–795.

Kazmi SO, Rodrigue N. 2019. Detecting amino acid preference shifts with codon-level mutation-selection mixture models. *BMC Evol Biol*. **19**:62.

Kraemer MUG, Hill V, Ruis C, Dellicour S, Bajaj S, McCrone JT, Baele G, Parag KV, Battle AL, Gutierrez B., *et al.* 2021. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science* **373**:889–895.

Kühnert D, Stadler T, Vaughan TG, Drummond AJ. 2016. Phylodynamics with migration: a computational framework to quantify population structure from genomic data. *Mol Biol Evol*. **33**:2102–2116.

Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, Russell CA, Smith DJ, Pybus OG, Brockmann D, *et al.* 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H_3N_2. *PLoS Pathog*. **10**:e1003932.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol*. **5**:e1000520.

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

Lemey P, Ruktanonchai N, Hong SL, Colizza V, Poletto C, Van den Broeck F, Gill MS, Ji X, Levasseur A, Oude Munnink BB, *et al.* 2021. Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature* **595**:713–717.

Membrebe JV, Suchard MA, Rambaut A, Baele G, Lemey P. 2019. Bayesian inference of evolutionary histories under time-dependent substitution rates. *Mol Biol Evol*. **36**:1793–1803.

Minin VN, Bloomquist EW, Suchard MA. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol*. **25**:1459–1471.

Minin VN, Suchard MA. 2008. Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc B: Biol Sci*. **363**: 3985–3995.

Müller NF, Dudas G, Stadler T. 2019. Inferring time-dependent migration and coalescence patterns from genetic sequence and predictor data in structured populations. *Virus Evol*. **5**: vez030.

Müller NF, Rasmussen DA, Stadler T. 2017. The structured coalescent and its approximations. *Mol Biol Evol*. **34**:2970–2981.

Müller NF, Wagner C, Frazar CD, Roychoudhury P, Lee J, Moncla LH, Pelle B, Richardson M, Ryke E, Xie H, *et al.* 2021. Viral genomes reveal patterns of the SARS-CoV-2 outbreak in Washington state. *Sci Transl Med*. **13**.

Nadeau SA, Vaughan TG, Scire J, Huisman JS, Stadler T. 2021. The origin and early spread of SARS-CoV-2 in Europe. *Proc Natl Acad Sci*. **118**:e2012008118.

Nielsen R. 2002. Mapping mutations on phylogenies. *Syst Biol*. **51**: 729–739.

Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, Gray RR, Arinaminpathy N, Stramer SL, Busch MP, *et al.* 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci*. **109**: 15066–15071.

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

Schliep KP. 2010. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**:592–593.

Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data–from vision to reality. *Eurosurveillance* **22**:30494.

Tegally H, Wilkinson E, Giovanetti M, Fonseca V, Doolabh DS, James S, Mdlalose N, Mlisana K, von Gottberg A, Walaza S, *et al.* 2021. Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature* **592**:1–8.

Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET, Schiabor Barrett KM, Larsen BB, Anderson C, White S, Cassens T, *et al.* 2021. Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* **184**:2587–2594.

Wilkinson E, Giovanetti M, Tegally H, San JE, Lessells R, Cuadros D, Martin DP, Rasmussen DA, Zekri ARN, Sangare AK, *et al.* 2021. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science* **374**:423–431.

Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA, Wertheim JO, Lemey P. 2020. The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**: 564–570.

Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol*. **60**:150–160.

Yang Z. 2014. *Molecular evolution: a statistical approach*. Oxford (England): Oxford University Press.