

UCLA

UCLA Previously Published Works

Title

Shrinkage-based Random Local Clocks with Scalable Inference.

Permalink

<https://escholarship.org/uc/item/91v768df>

Journal

Molecular Biology and Evolution, 40(11)

Authors

Fisher, Alexander

Ji, Xiang

Nishimura, Akihiko

et al.

Publication Date




2023-11-03

DOI

10.1093/molbev/msad242

Peer reviewed

Shrinkage-based Random Local Clocks with Scalable Inference

Alexander A. Fisher,¹ Xiang Ji,² Akihiko Nishimura,³ Guy Baele ,⁴ Philippe Lemey ,⁴ and Marc A. Suchard ^{5,6,7,*}

¹Department of Statistical Science, Duke University, Durham, NC, USA

²Department of Mathematics, School of Science & Engineering, Tulane University, New Orleans, LA, USA

³Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

⁴Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium

⁵Department of Computational Medicine, University of California, Los Angeles, CA, USA

⁶Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

⁷Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, CA, USA

*Corresponding author: E-mail: msuchard@ucla.edu.

Associate editor: Aida Ouangraoua

Abstract

Molecular clock models undergird modern methods of divergence-time estimation. Local clock models propose that the rate of molecular evolution is constant within phylogenetic subtrees. Current local clock inference procedures exhibit one or more weaknesses, namely they achieve limited scalability to trees with large numbers of taxa, impose model misspecification, or require a priori knowledge of the existence and location of clocks. To overcome these challenges, we present an autocorrelated, Bayesian model of heritable clock rate evolution that leverages heavy-tailed priors with mean zero to shrink increments of change between branch-specific clocks. We further develop an efficient Hamiltonian Monte Carlo sampler that exploits closed form gradient computations to scale our model to large trees. Inference under our shrinkage clock exhibits a speed-up compared to the popular random local clock when estimating branch-specific clock rates on a variety of simulated datasets. This speed-up increases with the size of the problem. We further show our shrinkage clock recovers known local clocks within a rodent and mammalian phylogeny. Finally, in a problem that once appeared computationally impractical, we investigate the heritable clock structure of various surface glycoproteins of influenza A virus in the absence of prior knowledge about clock placement. We implement our shrinkage clock and make it publicly available in the BEAST software package.

Key words: Bayesian phylogenetics, shrinkage clock, random local clock, divergence time estimation, Hamiltonian Monte Carlo.

Introduction

Molecular clock models are ubiquitous phylogenetic instruments for divergence-time estimation with applications ranging from timing placental mammal radiation (Springer et al. 2003) to estimating influenza diversity (Davidson et al. 2014). To capture clock rate variation along the lineages of a phylogeny, Thorne et al. (1998) and Rannala and Yang (2007) propose an autocorrelated, or “heritable” rate model, while others (Yoder and Yang 2000; Drummond and Suchard 2010) assume there exist, at most, a small number of “local” clocks on any given tree. Separately, Sanderson (2002) induces clock rate smoothing via a penalized likelihood approach. In each case, closely related lineages maintain similar or even identical evolutionary rates. Autocorrelated rate models are computationally appealing due to the induced smooth transition in rate from parent to child node along the

tree but may inappropriately shrink large rate changes between adjacent nodes (Smith et al. 2010). On the other hand, local clock models allow large rate changes to exist but can be computationally unpalatable on large problems due to the combinatorial complexity of choosing (or learning) the number and location of local clocks. When these quantities are simultaneously learned with the tree, Drummond and Suchard (2010) call this the random local clock (RLC) model.

Due to these complications, some authors employ uncorrelated relaxed clocks such as the uncorrelated log-normal relaxed molecular clock (Drummond et al. 2006), but this generates excessive rate heterogeneity in cases where clock rate changes are thought to be more punctuated, for example between HIV subtypes (Bletsa et al. 2019). Indeed, in a variety of settings, autocorrelation between clock rates appears crucial for accurate divergence-time estimation (Lepage et al. 2007). For an

Received: May 31, 2023. Revised: October 23, 2023. Accepted: November 07, 2023

© The Author(s) 2023. 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.

Open Access

in-depth review of various molecular clock models, see [Ho and Duchêne \(2014\)](#). Here, we propose an autocorrelated clock model where we place a Bayesian bridge shrinkage prior on the increment between parent and child log branch rates. Among various shrinkage priors in the literature, the Bayesian bridge has a unique advantage in having both a collapsed spike-and-slab representation as well as a Gaussian scale-mixture form. The first representation intuitively places large mass near zero reflecting our a priori belief that most increments should be zero but has heavy tails that allow for estimating large rate changes in an approximately unbiased manner. Like many other shrinkage priors, the Bayesian bridge includes a “global scale” nuisance parameter. In the absence of prior information, learning about this nuisance parameter typically limits the speed of inference. [Polson et al. \(2014\)](#) develop a framework to facilitate efficient Gibbs sampling of this nuisance parameter in a regression context and we utilize their approach here. On the other hand, the second representation of the Bayesian bridge as Gaussian scale-mixture is differentiable almost everywhere. We exploit this feature and develop an efficient Hamiltonian Monte Carlo (HMC) sampler over the space of increments that employs recent work on closed form gradient representations ([Ji et al. 2020](#)) to make our shrinkage-clock inference scalable to large trees. Crucial to our inference, we define recursive algorithms to compute the requisite joint gradient of the log-posterior in our transformed increment space with computational complexity that scales only linearly with the number of tips in the tree. We implement our method in BEAST ([Suchard et al. 2018](#)), a popular software package for reconstructing rooted, time-measured phylogenies. Due to our efficient inference machinery, our shrinkage clock achieves the tractable benefits of the autocorrelated rate model and simultaneously maintains the flexibility of more punctuated local clock models.

We examine our model under examples where identifying local clock structure is vital to accurate divergence-time estimation. For this reason, we compare our method to the state-of-the-art RLC. We demonstrate the inference speed gains of our approach versus the RLC across 90 different simulated data sets comprised of 40, 80, and 160 taxa trees. We additionally compare the accuracy of our shrinkage clock to the RLC by studying the adaptive radiation of rodents and other mammals, and demonstrate utility of the heavy-tailed Bayesian bridge shrinkage prior by comparing it to the more ubiquitous Laplace prior. Finally, we deploy our shrinkage clock to estimate the existence, location, and magnitude of host-specific clock rates in surface glycoproteins of the influenza A virus.

Materials and Methods

Shrinkage-based RLCs

Setup

Consider a rooted, bifurcating tree F with N tips and $N - 1$ internal (ancestral) nodes. We index tips $i = 1, \dots, N$ and

internal nodes $i = N + 1, \dots, 2N - 2$. We designate node $2N - 1$ to be the root of the tree. Let $\text{pa}(i)$ denote the parent of the i th node and let branch length t_i connect node i with its parent.

The Relaxed Clock

Aligned molecular sequence data \mathbf{S} evolve according to a continuous time Markov process defined by infinitesimal rate matrix \mathbf{Q} . In our examples, \mathbf{Q} is a 4×4 matrix that describes the relative substitution process between nucleotides along the branches in F , but in general, \mathbf{Q} may be of larger dimension to accommodate alignments at the codon or amino acid resolution, see [Yang \(2014\)](#) for reference. Each site l of \mathbf{S} evolves independently and identically according to \mathbf{Q} but may have its own site-specific rate of evolution s_l . A priori we specify that $\mathbf{E}[s_l] = 1$. Under the relaxed clock model, the transition probability matrix for branch i ,

$$\mathbf{P}_i = \exp \{ \rho_i t_i s_l \mathbf{Q} \}, \quad (1)$$

where branch-rate multiplier ρ_i is the number of expected substitutions per unit time. To resolve identifiability issues between the height of the tree and branch-rate multipliers $\boldsymbol{\rho} = \{ \rho_1, \dots, \rho_{2N-2} \}$, we employ the rescaling proposed by [Drummond and Suchard \(2010\)](#). Under this transform, each branch-rate multiplier is the product of branch-specific clock rate r_i and global substitution rate γ scaled by the inverse of total expected substitutions per total tree time,

$$\rho_i = \gamma r_i \frac{\sum_k t_k}{\sum_k r_k t_k}. \quad (2)$$

This results in one fewer degree of freedom since

$$\frac{\sum_i \rho_i t_i}{\sum_i t_i} = \gamma. \quad (3)$$

For heterochronous data, we estimate γ , but for ultrametric studies where the height of the tree is not identifiable we fix $\gamma = 1$.

Autocorrelated Shrinkage Clock

We assume clock rates $\mathbf{r} = \{ r_1, \dots, r_{2N-2} \}$ are autocorrelated and model the incremental difference ϕ_i between branch i 's clock rate and its parent lineage clock rate,

$$\begin{aligned} \log r_i - \log r_{\text{pa}(i)} &= \phi_i, \\ \text{for } i &\in \{ 1, \dots, 2N - 2 \} \text{ and } r_{2N-1} = 1. \end{aligned} \quad (4)$$

Under this parameterization, the increments $\boldsymbol{\phi} = \{ \phi_1, \dots, \phi_{2N-2} \} \in \mathbb{R}^{2N-2}$ are a linear transformation of $\log \mathbf{r}$. To shrink the total number of rate changes along the tree, we let $\phi_i \text{ iid} \sim P_\phi$ such that $\mathbf{E}[\phi_i] = 0$. Typically, P_ϕ may follow a Gaussian ([Thorne et al. 1998](#)) or Laplace

distribution. We choose the flexible, heavy-tailed, Bayesian bridge prior (Polson et al. 2014) on the increments,

$$P_\phi \propto \exp\left\{-\left|\frac{\phi_i}{\mu}\right|^\alpha\right\}, \quad (5)$$

where $\mu > 0$ is termed the “global scale” and $\alpha \in (0, 1]$ changes the shape of P_ϕ where smaller α places more mass near zero. See Fig. 1 for a comparison of the bridge to common shrinkage priors. Choosing α to be close to 0 forces the Bayesian bridge prior closer to best subset selection when used in a regression setting, while $\alpha = 1$ matches the Laplace prior. In all examples, we set $\alpha = \frac{1}{4}$ to enforce slightly stronger shrinkage than the default 0.5 employed by Polson et al. (2014). Since increments are independent, the joint prior is simply the product

$$p(\phi | \mu) \propto \prod_{i=1}^{2N-2} \exp\left\{-\left|\frac{\phi_i}{\mu}\right|^\alpha\right\}. \quad (6)$$

Inference

We follow the computationally efficient sampling approach outlined by Polson et al. (2014) and view the prior on the increments as a scale mixture of normals (West 1987),

$$p(\phi_i | \mu) = \int p(\phi_i | \lambda_i, \mu) d\lambda_i, \quad (7)$$

where the local scale of branch i , $\lambda_i > 0$ and draws from a stable distribution with positive support, see Hofert

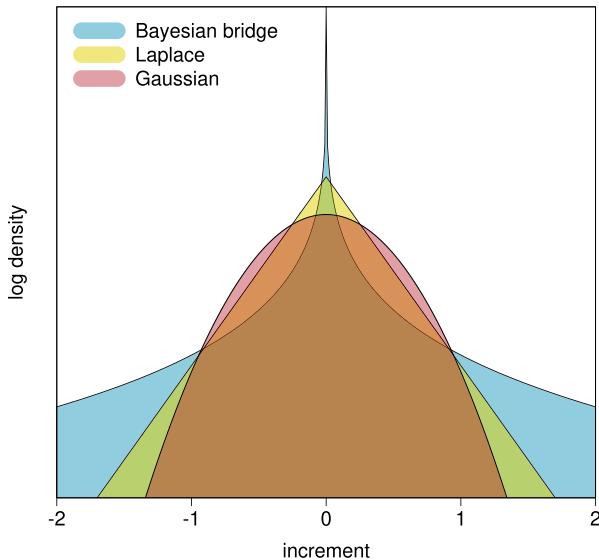


Fig. 1. The shape of various shrinkage priors on the increment of the log-rate. The Bayesian bridge prior places more mass near 0 and has heavier tails compared to other common shrinkage priors. The Bayesian bridge reflects our a priori belief that local clocks are rare, but may arbitrarily speed up or slow down the rate of molecular evolution.

(2011), Nishimura and Suchard (2023) for more details. To improve convergence speed and maintain the benefits of our heavy-tailed prior, we employ the shrunken-shoulder regularization of Nishimura and Suchard (2023) and augment our bridge to have light tails past a reasonably large point. Our scale mixture prior on an increment becomes

$$p(\phi_i | \lambda_i, \mu) = N\left(0, \left(\frac{1}{\zeta^2} + \frac{1}{\lambda_i^2 \mu^2}\right)^{-1}\right), \quad (8)$$

where slab width ζ bounds the variance of increments to ζ^2 . In the examples to follow, we set $\zeta = 2$, effecting a weakly informative, generous upper bound on clock rate changes. Specifically, a slab width of 2 asserts that there is at most 5% probability for r_i to be greater than $50 \times r_{pa(i)}$.

We are interested in learning about the posterior,

$$p(\mathbf{r}, \mu, \gamma, \boldsymbol{\theta}, F | \mathbf{S}) \propto \underbrace{p(\mathbf{S} | \mathbf{r}, \gamma, \boldsymbol{\theta}, F)}_{\text{likelihood}} \underbrace{p(\mathbf{r} | \boldsymbol{\lambda}, \mu) p(\boldsymbol{\lambda}) p(\mu) p(\gamma) p(\boldsymbol{\theta}, F)}_{\text{priors}} d\boldsymbol{\lambda}, \quad (9)$$

where $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_{2N-2}\}$, $\boldsymbol{\theta}$ represents all relevant parameters that describe the molecular substitution model and, again, F is the phylogenetic tree. We place a relatively uninformative, Gamma prior on $\mu^{-\alpha}$ with shape 1 and scale 2. Additionally, we place the continuous-time Markov chain conditional reference prior of Ferreira and Suchard (2008) on the global substitution rate γ . This prior is uninformative and yields a proper density. We detail the priors on $\boldsymbol{\theta}$ and F in each example of the sequel.

We use Markov chain Monte Carlo (MCMC) to marginalize over local scale parameters and approximate the posterior (9). Specifically, we employ a random-scan Metropolis-within-Gibbs (Levine and Casella 2006; Liu 2008) sampling approach to update the full conditional densities implicit in equation (9). Efficient sampling schemes for $\gamma, \boldsymbol{\theta}, F$ are well described by Suchard et al. (2018), while Polson et al. (2014) outline the efficient scale-mixture approach we use to sample $(\mu, \boldsymbol{\lambda})$. Here, we turn our attention to sampling

$$p(\mathbf{r} | \mu, \gamma, \boldsymbol{\theta}, F, \mathbf{S}) \propto \int p(\mathbf{S} | \mathbf{r}, \gamma, \boldsymbol{\theta}, F) p(\mathbf{r} | \boldsymbol{\lambda}, \mu) d\boldsymbol{\lambda}. \quad (10)$$

Since there are $2N - 2$ correlated branch-rate multipliers, one for each branch of the tree, univariable MCMC sampling schemes for \mathbf{r} scale poorly to large trees. To remedy this difficulty, we employ HMC to sample all \mathbf{r} simultaneously and with high acceptance probability. HMC leverages the geometry of the high-dimensional branch-rate multiplier space to propose states that are farther away than traditional proposals but stay within regions of high posterior density (HPD). HMC escapes entrapment by local extrema of the posterior by generating random momentum $\mathbf{v} = \{v_1, \dots, v_{2N-2}\}$ in each dimension where typically $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{M})$ (Neal 2011). Often mass

matrix $\mathbf{M} = \mathbf{I}_{2N-2}$ but HMC sampling may be improved by using an alternative \mathbf{M} , such as an approximation of the Hessian of the log-posterior (Zhang and Sutton 2011; Stan Development Team 2017). For further reading on HMC, see Neal (2011). While HMC samplers offer more efficient posterior exploration, they require computationally expensive gradient calculations that often diminish their usefulness. Here, we exploit and extend recent work (Ji et al. 2020) on branch-specific clock rate gradients to facilitate fast inference of \mathbf{r} under our shrinkage model.

HMC Increment Sampler

We generate proposals in increment space, since ϕ are uncorrelated in the prior and we transform back to rate space as described by the linear transformation in equation (4). HMC sampling of the rates requires the gradient of the rate log-posterior,

$$\begin{aligned} \frac{\partial}{\partial \phi_k} \log p(\mathbf{r} | \mu, \gamma, \theta, F, \mathbf{S}) &= \int \frac{\partial}{\partial \phi_k} \underbrace{\log p(\mathbf{S} | \mathbf{r}, \gamma, \theta, F)}_{L[\rho(\mathbf{r})]} \\ &+ \frac{\partial}{\partial \phi_k} \log p(\mathbf{r} | \lambda, \mu) d\lambda, \end{aligned} \quad (11)$$

where $k \in \{1, \dots, 2N-2\}$. To compute the gradient of the log-likelihood with respect to the increments, we first find the gradient with respect to clock rates \mathbf{r} ,

$$\frac{\partial}{\partial r_j} L[\rho(\mathbf{r})] = \gamma \nabla_{\rho} L[\rho(\mathbf{r})] \mathbf{T}, \quad (12)$$

where we compute all entries in $\nabla_{\rho} L[\rho(\mathbf{r})] = (\frac{\partial}{\partial \rho_1}, \dots, \frac{\partial}{\partial \rho_{2N-2}}) L[\rho(\mathbf{r})]$ with the computational $\mathcal{O}(N)$ algorithm derived by Ji et al. (2020) and

$$\mathbf{T}_{ij} = \begin{cases} \frac{\sum_k t_k}{\sum_k r_k t_k} - r_i t_i \frac{\sum_k t_k}{(\sum_k r_k t_k)^2} & \text{if } i = j \\ -r_i t_j \frac{\sum_k t_k}{(\sum_k r_k t_k)^2} & \text{if } i \neq j. \end{cases} \quad (13)$$

We complete the gradient

$$\begin{aligned} \frac{\partial}{\partial \phi_k} L[\rho(\mathbf{r})] &= \sum_{j=1}^{2N-2} \frac{\partial}{\partial r_j} L[\rho(\mathbf{r})] \frac{dr_j}{d\phi_k} \\ \text{and } \frac{dr_j}{d\phi_k} &= \begin{cases} r_j & \text{if } i \text{ ancestral to } j \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (14)$$

where transformation $dr_j/d\phi_k$ follows directly from equation (4). To preserve the $\mathcal{O}(N)$ gradient computation, we take advantage of the tree structure explicit in equation (14) and accumulate the gradient of the log-likelihood via one postorder traversal of the tree. To begin, let i and j be both daughters of node k in F , then

$$\frac{\partial}{\partial \phi_k} L[\rho(\mathbf{r})] = \begin{cases} r_k \times \frac{\partial}{\partial r_k} L[\rho(\mathbf{r})] & \text{if } k \text{ is a tip} \\ \left(\frac{\partial}{\partial \phi_i} + \frac{\partial}{\partial \phi_j} + r_k \times \frac{\partial}{\partial r_k} \right) L[\rho(\mathbf{r})] & \text{otherwise.} \end{cases} \quad (15)$$

We next turn our attention to the gradient of the log-prior,

$$\frac{\partial}{\partial \phi_k} \log p(\phi | \lambda, \mu) = \frac{\partial}{\partial \phi_k} \left[\log p(\phi | \lambda, \mu) + \log \left| \frac{d\phi}{d\mathbf{r}} \right| \right]. \quad (16)$$

Since $p(\phi | \lambda, \mu)$ is Gaussian, the first term gradient unwinds,

$$\begin{aligned} \frac{\partial}{\partial \phi_k} \log p(\phi | \lambda, \mu) &= \frac{\partial}{\partial \phi_k} \sum_{j=1}^{2N-2} \log p(\phi_j | \lambda_j, \mu) \\ &= \frac{\partial}{\partial \phi_k} \sum_{j=1}^{2N-2} \log N \left(0, \left(\frac{1}{\xi^2} + \frac{1}{\lambda_j^2 \mu^2} \right)^{-1} \right) \\ &= -\phi_k \left(\frac{1}{\xi^2} + \frac{1}{\lambda_j^2 \mu^2} \right). \end{aligned} \quad (17)$$

Numerical solutions to the second term in equation (11) involve the change-of-variable Jacobian $(\partial/\partial \phi_k) \log |d\phi/d\mathbf{r}|$ and appear to necessitate an $\mathcal{O}(N^2)$ sparse determinant computation. To facilitate faster computation of the transform, we index nodes of the tree such that $i < j_i$ is not ancestral to j . Under this indexing, $d\phi/d\mathbf{r}$ is an upper triangular matrix with $1/r_i$ along its diagonal, see [supplementary Fig. S1, Supplementary Material](#) online for an example. The gradient of the log determinant,

$$\begin{aligned} \frac{\partial}{\partial \phi_k} \log \left| \frac{d\phi}{d\mathbf{r}} \right| &= \frac{\partial}{\partial \phi_k} \log \prod_{j=1}^{2N-2} \frac{1}{r_j} = -\frac{\partial}{\partial \phi_k} \sum_{j=1}^{2N-2} \log r_j \\ &= \underbrace{\sum_j \mathbf{1}_{[r_j \text{ depends on } \phi_k]}}_{d_k}, \end{aligned} \quad (18)$$

where $\mathbf{1}$ is the indicator function and d_k is the number of descendants of node k . Altogether,

$$\frac{\partial}{\partial \phi_k} \log p(\mathbf{r} | \lambda, \mu) = -\phi_k \left(\frac{1}{\xi^2} + \frac{1}{\lambda_k^2 \mu^2} \right) - d_k, \quad (19)$$

and we accumulate d_k in one recursive postorder tree traversal by observing $d_k = d_i + d_j + 1$, where, again, i and j are both daughters of k .

To further improve the proposals of our HMC sampler, we precondition the mass matrix \mathbf{M} to be the current-state absolute value of the Hessian of the log-prior,

$$\left| \frac{\partial^2}{\partial \phi_i \partial \phi_j} \log p(\mathbf{r} | \lambda, \mu) \right| = \begin{cases} \left(\frac{1}{\xi^2} + \frac{1}{\lambda_i^2 \mu^2} \right) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

This diagonal matrix weights momentum draws by prior increment precision. Intuitively, equation (20) improves HMC sampling by rescaling increment proposals by the variance of ϕ , allowing larger steps to be taken in dimensions with larger variance. See Neal (2011) for further discussion on mass matrix transformations.

Results

Local Clocks in 3 Nuclear Genes of Rodents and Other Mammals

To verify the ability of our model to recover clock rate variation, we turn to a well-studied example of adaptive radiation in mammals and rodents. Huchon et al. (2002) and Douzery et al. (2003) examine the adaptive radiation of 21 rodents compared to 19 other placental mammals and two marsupial outgroups using the first two codon positions for 3 nuclear genes: ADRA2B, IRBP, and vWF (2,422 alignment sites). Douzery et al. (2003) establish the presence of clock variability within this set of taxa and report their best fitting model contains 5 “local clocks” and about 13 rate changes. Note that Douzery et al. (2003) do not require a single “local clock” to be comprised of connected components on the tree. Drummond and Suchard (2010) use the RLC model to re-examine this claim and estimate the existence of between 6 and 12 local clocks. Here, we employ our shrinkage-clock model to jointly infer the mammalian phylogeny as well as the number and location of local clocks. Because this is an ultrametric example, $\gamma = 1$. We follow the specifications of Drummond and Suchard (2010) and Douzery et al. (2003)

and use a general time reversible (GTR) substitution model with a 4 category discrete- Γ site rate model. We run 10 separate Markov chains of our shrinkage clock with 10 different starting trees for 30M states and build a maximum clade credibility (MCC) tree from the combined results (Fig. 2). We further run 100 RLC chains with 100 different starting trees for 30M states and build an MCC tree for comparison. Under the combinatorial parameter space of the RLC, we observe suboptimal mixing and that some chains converge to different modes, hence our choice for combining 100 independent chains; the 10 independent chains for the shrinkage clock simply errs on the side of caution since each independent shrinkage-clock chain converges to the same topology. Incidentally, the shrinkage-clock MCC topology differs from the RLC MCC in two places. First, *Bradypus* attaches to one of two neighbor internal branches deep in the tree. Second, *Anomalurus* is more closely related to the *Dipus* than *Castor* under the shrinkage clock. This second difference highlights the well-known difficulty in *Anomalurus* placement (Horner et al. 2007). Indeed, in an analysis of 6 nuclear genes, Bonga-Kanfi et al. (2009) group Anomaluridae closer to Dipodidae, but support for this grouping is slight and the authors remark that all 3 possible relationships among *Anomalurus*, *Dipus*, and *Castor* have been suggested in other published studies. The posterior probabilities of *Bradypus* and *Anomalurus* parent branches under the shrinkage clock are almost equal (0.64 and 0.49, respectively).

We estimate the existence of 4 local clocks (and thus about 4 rate changes) where we define a local clock on branch i if the posterior odds $\phi_i > 0$ is greater than 10 or less than $\frac{1}{10}$. The posterior odds here is equivalent to a

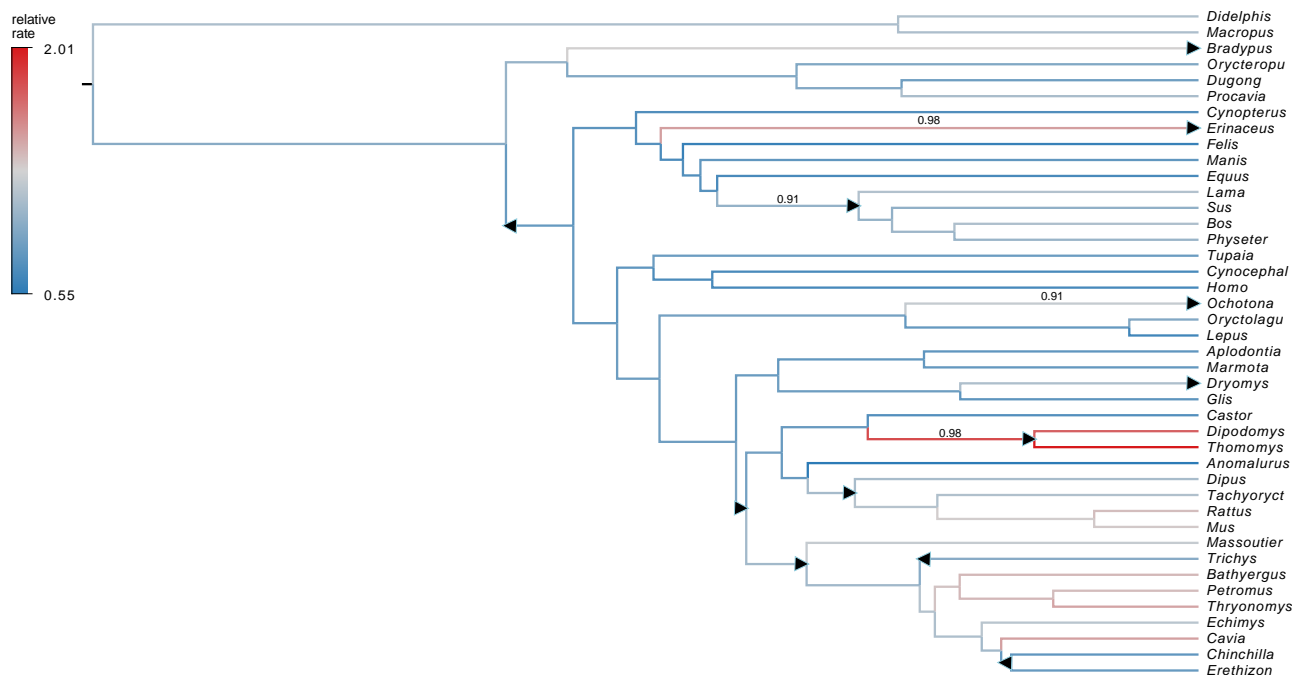


Fig. 2. MCC tree under shrinkage clock of mammalian and rodent radiation where branches are colored by posterior mean relative clock rates r . If branch i starts a new clock, it is labeled with the posterior probability $\phi_i > 0$. For comparison, local clocks of the RLC model are depicted as triangles. Triangles pointing left designate clock rate slowdown, while triangles pointing right indicate a rate increase. Two local clocks of the RLC are excluded due to topological differences between the RLC and shrinkage-clock MCC trees.

Bayes factor since an increment is equally likely to be positive or negative under the prior. While some regard a Bayes factor greater than 10 as suggestive of “strong evidence” against an alternative hypothesis (Kass and Raftery 1995), we acknowledge that this is an arbitrary cutoff and further investigation might find use for alternative cutoffs. In the approach, we do not make assumptions about the magnitude of local clocks on a tree and instead use posterior probability of increment sign to define a clock. To keep track of branch i across nonfixed topologies, we use the set of descendant tips of node i to define the branch.

To illustrate the benefits of using a heavy-tailed prior on the increments, we further compare the performance of our Bayesian bridge prior to the more usual Laplace prior for shrinkage (see Fig. 1). We again fit our shrinkage clock as described above but remove the slab and fix $\alpha = 1$, thus placing a Laplace prior on each increment. We find the posterior mean of increment variance under both the Laplace and Bridge priors is 0.057 with 95% HPD intervals (0.036, 0.089) and (0.035, 0.093), respectively. Despite having very similar variance, we find the posterior mean

of the absolute maximum increment is 0.84 (0.57, 1.22) and 1.01 (0.70, 1.54) under the Laplace and Bridge priors, respectively. This provides evidence for induced smoothing of the clock rates under the exponential tails of a Laplace prior, that on average shrinks the largest increment by approximately 20%.

Simulation Study

We compare the scalability and accuracy of our shrinkage clock to the RLC under a simulated example. We generate 1,000, 2,000, and 3,000 character nucleotide sequences from fixed 40, 80, and 160 tip trees 5 times each. In each simulation, there are 4 distinct lineages (A, B, C, D) of taxa, each with an equal number of tips. In all cases, time to most recent common ancestor for each lineage is 40 years and tree height is 80 years. Lineages B, C, and D evolve with a relative clock rate of 1.0, while the most recent common ancestor of lineage A starts a new clock with relative rate 2.0.

We compare the accumulation of effective sample size (ESS) per unit time of branch-specific clock rates under

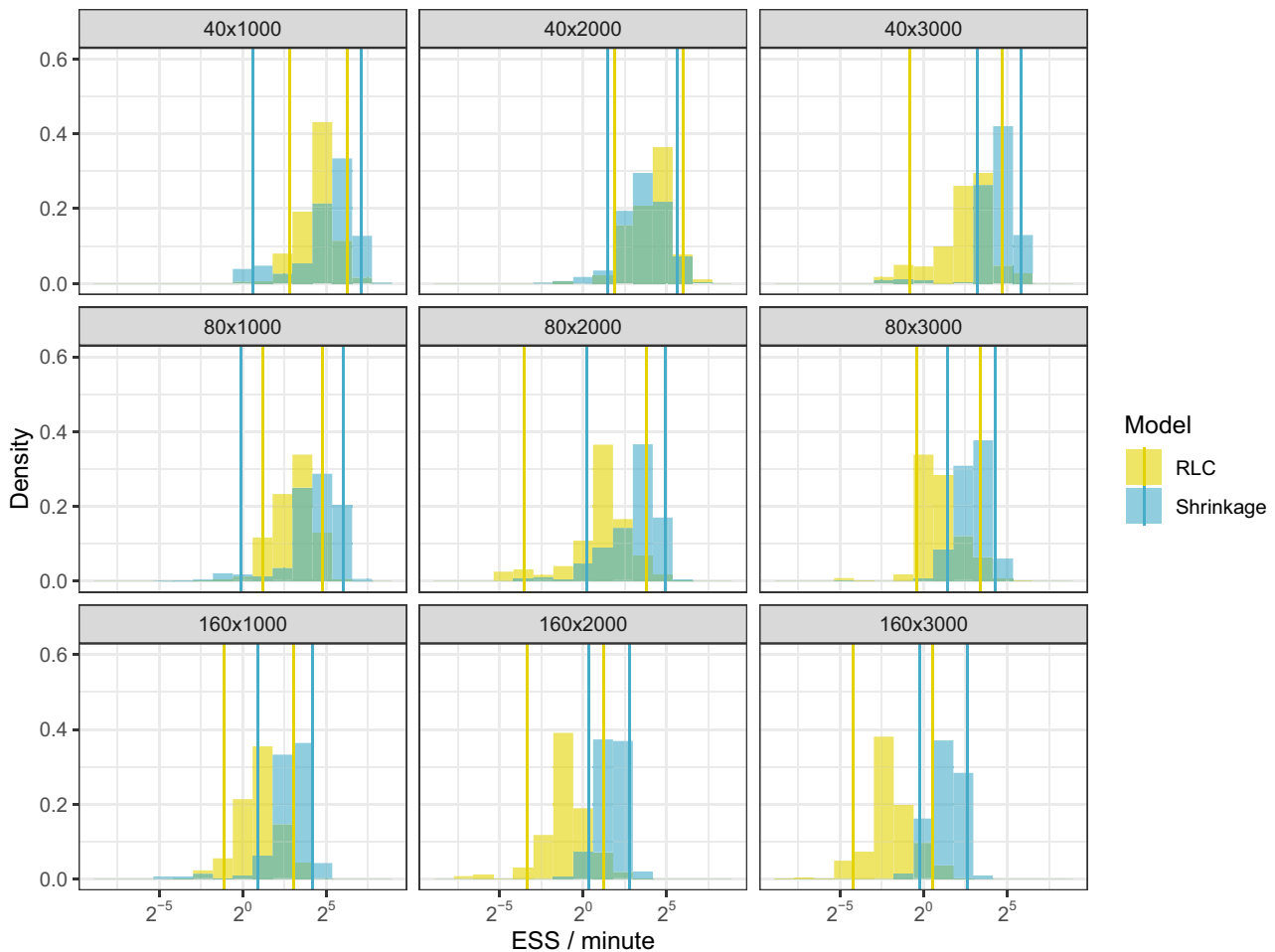


Fig. 3. Effective sample size (ESS) of branch-specific clock rates per minute of BEAST runtime under the shrinkage clock and RLC during a full joint phylogenetic analysis. Each facet title shows data dimensions (tips \times sequence length). Vertical lines show 0.05 and 0.95 quantiles for shrinkage and RLC ESS/minute.

Table 1. Ratio (Shrinkage:RLC) of median branch-specific clock rates ESS per minute of BEAST runtime during a full joint phylogenetic analysis

Sequence length	Number of tips		
	40	80	160
1,000	1.85	2.43	3.95
2,000	0.53	4.54	6.73
3,000	3.07	5.12	11.14

both our Bayesian bridge shrinkage clock and the RLC while simultaneously inferring the phylogeny. ESS approximates the number of independent samples from a Markov chain and we use this metric to evaluate how well each inference procedure explores clock rate space. We report the results across all 45 simulated datasets in Fig. 3. We summarize median ESS/minute speed-up in Table 1 and find the shrinkage-clock speed-up scales as both the number of tips and length of sequence data increases. In all but one 40-tip example, the shrinkage clock outperforms RLC. This reveals that on certain trees with a modest number of tips, the RLC is competitive. As the tips double to 80 or quadruple to 160, the set of all possible local clocks under the RLC grows exponentially with the number of taxa (24 to 72 orders of magnitude), challenging inference under the RLC. On average, across all 45 simulations, the shrinkage clock accumulates ESS/minute 4.3× as fast.

In all examples herein, we precondition the mass matrix.

We report here that in 40 tip examples, preconditioning the mass matrix improves ESS/minute 15× for the worst-explored clock rate under the shrinkage clock.

Additionally, the true relative clock rate for the “A” clade is 2.0 and we report posterior mean estimates, together with 95% HPD intervals under both our shrinkage clock and the RLC in Table 2. In all cases, the true relative clock rate is captured within the 95% HPD interval.

We further report, in [Supplementary Material](#) online, posterior distributions of root age ([supplementary Fig. S2, Supplementary Material](#) online) and clade age ([supplementary Fig. S3, Supplementary Material](#) online) under each model, as well as posterior mean estimates of branch lengths ([supplementary Fig. S4, Supplementary Material](#) online). We compare these estimates to the true values we simulate from. In these simulated examples, the shrinkage clock typically exhibits higher accuracy and lower variance in parameter estimates. See the [Supplementary Material](#) online for more detail.

Influenza A Virus

We further demonstrate the scalability and utility of our shrinkage-clock model by examining the evolution of two major influenza A virus (IVA) surface glycoprotein subtypes: hemagglutinin (HA) H7, and neuraminidase (NA) N7. Both HA and NA protein mutations impact IVA’s epitope and allow IVA to escape adaptive immune responses ([Wilson and Cox 1990](#); [McAuley et al. 2019](#)). [Worobey et al. \(2014\)](#) find divergence time estimation is sensitive to molecular clock model specification. To

Table 2. Posterior mean and 95% HPD interval of the relative clock rate of clade “A” averaged across 5 replications for each simulation

Sequence length	# of tips	Shrinkage	RLC
1,000	40	1.71 (0.95, 2.43)	1.88 (1.00, 2.41)
2,000	40	1.71 (0.96, 2.27)	1.89 (1.40, 2.37)
3,000	40	2.15 (1.70, 2.65)	2.40 (1.22, 4.28)
1,000	80	1.89 (1.12, 2.66)	2.15 (1.56, 2.94)
2,000	80	2.05 (1.48, 2.59)	1.67 (0.99, 2.64)
3,000	80	1.98 (1.59, 2.44)	1.34 (0.99, 2.01)
1,000	160	2.05 (1.47, 2.74)	1.32 (0.98, 2.36)
2,000	160	1.96 (1.56, 2.37)	1.62 (1.00, 2.22)
3,000	160	2.04 (1.74, 2.38)	1.57 (0.99, 2.22)

The true value is 2. Here, both the shrinkage clock and RLC run for approximately equal amounts of wall time after converging to the posterior.

consistently estimate divergence times, [Worobey et al. \(2014\)](#) allow the clock rates of various glycoprotein subtypes to vary only between viral hosts and find H7 and N7 each evolve slower in equine hosts than avian hosts. We reexamine this claim with our more general shrinkage-clock model that does not assume the existence of host-dependent clock rates. Specifically, we re-analyze 146 complete gene (1,716 nt) sequences of H7 and 92 complete gene (1,416 nt) sequences of N7. In each case, we follow the model specifications of [Worobey et al. \(2014\)](#) and employ a GTR substitution model with 4 category discrete- Γ site rate model. We depart from their example, however, in our use of tree prior. We employ a Bayesian skygrid prior ([Gill et al. 2013](#)) with 50 population size bins and a cutoff of 200 years instead of using the skyride prior ([Minin et al. 2008](#)).

We find no sharp local clocks exist with Bayes factor > 10 or $< \frac{1}{10}$ on the NA N7 tree under our shrinkage-clock model but do see evidence for rate heterogeneity. Incidentally, the most likely clock occurs on the branch that begins the Eastern avian clade of the NA N7 tree ([Fig. 4](#)). The second most probable clock is found in a subclade of the equine lineage. On the other hand, we estimate the existence of 7 local clocks on the HA H7 tree and report these in [Fig. 5](#). Overall, the mean posterior clock rate for NA N7 is lower than HA H7. We report the posterior mean and 95% HPD intervals of γ are 2.7 $\{2.1-3.3\} \times 10^{-3}$ and 3.3 $\{2.8-3.9\} \times 10^{-3}$ for NA N7 and HA H7, respectively. Additionally, under our shrinkage clock, the posterior mean root dates and 95% HPD intervals of the N7 and H7 trees in absolute time are 1798 (1733–1855) and 1853 (1808–1897), respectively. Furthermore, the median ESS/minute is .50 and .11 for the N7 and H7 examples respectively. While the size of this example is comparable to our simulation study, additional modeling considerations slow the speed of inference. In this example, the RLC fails to converge to the posterior when run for an equivalent amount of time.

Discussion

We develop a robust autocorrelated heritable clock rate model that scales to large trees, avoids excessively shrinking clock rates and learns the number and location of local

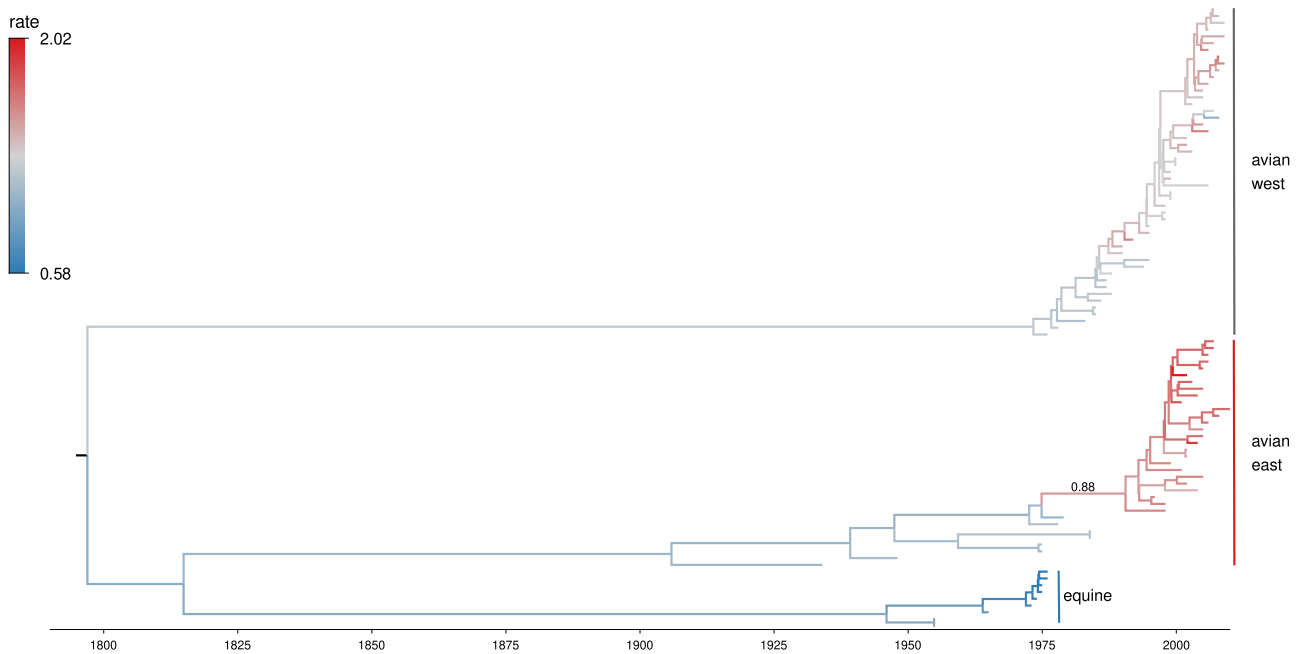


FIG. 4. MCC tree for influenza A's NA subtype N7. Branches are colored by posterior clock rates. The most probable local clock is reported and labeled with posterior probability that $\phi_i > 0$. The second most probable clock starts a subclade of the equine lineage and has a Bayes factor $\phi_i > 0$ of 0.351.

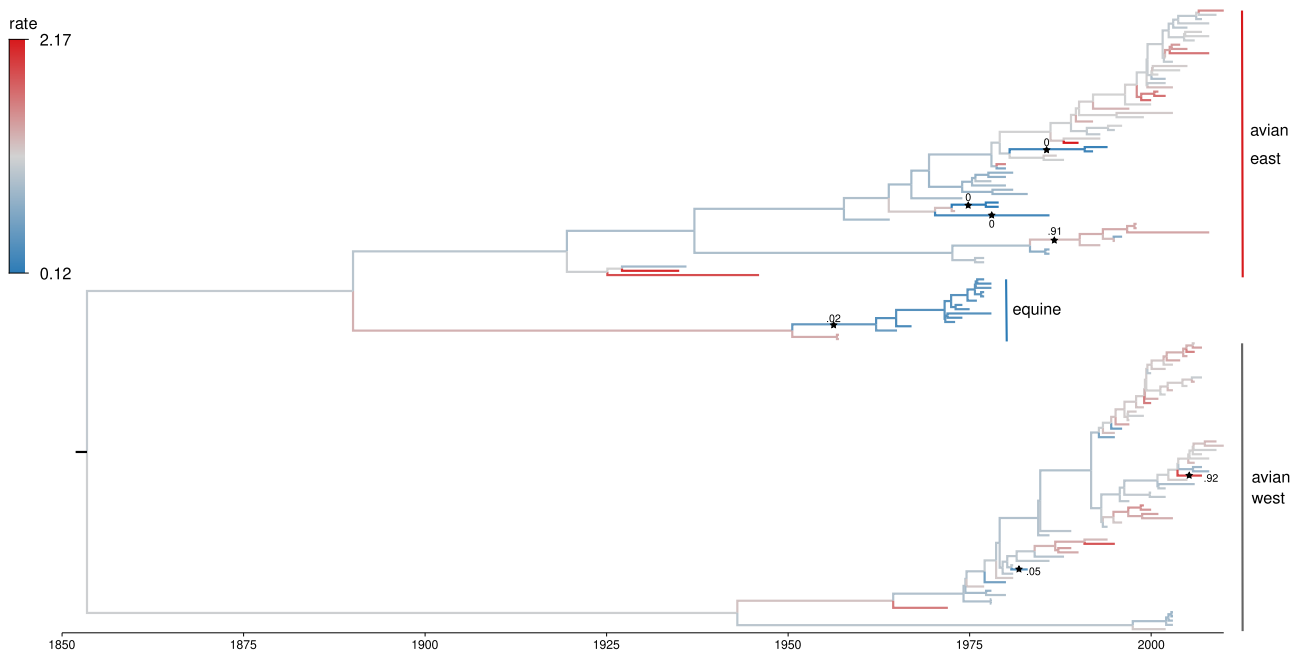


FIG. 5. MCC tree for influenza A's HA subtype H7. Branches are colored by posterior clock rates. Local clocks are labeled with a star and the posterior probability $\phi_i > 0$.

clocks on a tree. Crucially, we model the incremental difference between log clock rates on the tree as drawing from a Bayesian bridge prior that shrinks most changes to approximately 0 unless the data warrant otherwise. To facilitate scalability, we employ HMC to generate proposals in the independent increment space and derive recursive postorder algorithms to compute the gradient and

its requisite transforms. Our recursive algorithms achieve $\mathcal{O}(N)$ computational speed, signifying that they will continue to work well as N grows large. We further improve the speed of our HMC sampler by preconditioning the mass matrix with the Hessian of the log-prior.

In our examination of the adaptive radiation of rodents and other mammals (see the section “Local Clocks in 3

Nuclear Genes of Rodents and Other Mammals”), our shrinkage clock recovers the location of 4 local clocks estimated under the RLC. Douzery et al. (2003) also recover these 4 rate changes in addition to several more. We note that our shrinkage clock prefers branches with similar clock rates to be connected. However, we do not classify some rate heterogeneity as new local clocks due to our particular choice of classification threshold. We choose a Bayes factor of 10 to classify clocks, but shrinkage-clock users may wish to adjust this threshold to increase or decrease clock rate sensitivity. Comparing the statistical properties of different clock-classification schemes remains an important avenue for future work.

We apply our shrinkage-clock model to reexamine the number of local clocks present in IVA surface glycoproteins NA N7 and HA H7 across equine and avian hosts. We confirm the equine slowdown reported by Worobey et al. (2014) but interestingly find that the N7 tree shows marked rate variation (Fig. 4) between western and eastern hemisphere avian influenza lineages, however, this rate variation is not supported by the Bayes factor cutoff. Root date estimates vary from Worobey et al. (2014) but this may be in part due to the different tree priors. Additionally, we find 7 local clocks under our shrinkage clock across the H7 tree. Since 3 of these clocks belong to edges of tip nodes, this may reflect incomplete sampling or sequencing error. Despite inferring 6 more clocks than the host-specific model, the posterior mean estimate of root date under our shrinkage clock is within 5 years of previous estimates (Worobey et al. 2014).

Our shrinkage clock accumulates ESS/minute of clock rates 4.3× faster on average than the RLC across a variety of simulated examples. If ESS is used as a stopping criterion for phylogenetic reconstruction, we expect this may save shrinkage clock users 76% of BEAST runtime on average compared to the RLC. As the bridge exponent α approaches 0, the bridge prior density is more peaked near zero resulting in sharper increment shrinkage and thus better distinguishable local clocks. Nishimura and Suchard (2023) examine multiple α and report closer to optimal coverage for smaller α but with increasing computational cost due to mixing. For this reason, shrinkage-clock users may find it useful to adjust α depending on desired clock rate coverage or to tackle even larger tree studies.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgements

This work was supported through National Institutes of Health grants R01AI153044, R01AI162611, and U19AI135995, European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement no. 725422-ReservoirDOCS) and the Wellcome Trust project 206298/Z/17/Z (ARTICNetwork). P.L. acknowledges

support by the Research Foundation—Flanders (“Fonds voor Wetenschappelijk Onderzoek—Vlaanderen,” G066215N, G0D5117N, and G0B9317N). G.B. acknowledges support by the DURABLE project which receives funding under the EU4Health program, by the Research Foundation—Flanders (“Fonds voor Wetenschappelijk Onderzoek—Vlaanderen,” G0E1420N and G098321N) and by the Internal Funds KU Leuven (grant no. C14/18/094).

Data Availability

We provide installation instructions for the development branch of BEAST in which we implement our model and make all BEAST XML files used in this work publicly available at https://github.com/suchard-group/shrinking_clocks.

References

- Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW, Huchon D. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evol Biol*. 2009;9(1):1–12. <https://doi.org/10.1186/1471-2148-9-71>.
- Bletsa M, Suchard MA, Ji X, Gryseels S, Vrancken B, Baele G, Worobey M, Lemey P. Divergence dating using mixed effects clock modelling: an application to HIV-1. *Virus Evol*. 2019;5(2):vez036. <https://doi.org/10.1093/ve/vez036>.
- Davidson I, Fusaro A, Heidari A, Monne I, Cattoli G. Molecular evolution of H9N2 avian influenza viruses in Israel. *Virus Genes*. 2014;48(3):457–463. <https://doi.org/10.1007/s11262-014-1037-0>.
- Douzery EJ, Delsuc F, Stanhope MJ, Huchon D. Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *J Mol Evol*. 2003;57(1):S201–S213. <https://doi.org/10.1007/s00239-003-0028-x>.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):e88. <https://doi.org/10.1371/journal.pbio.0040088>.
- Drummond AJ, Suchard MA. Bayesian random local clocks, or one rate to rule them all. *BMC Biol*. 2010;8(1):114. <https://doi.org/10.1186/1741-7007-8-114>.
- Ferreira MA, Suchard MA. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can J Stat*. 2008;36(3):355–368. <https://doi.org/10.1002/cjs.v36.3>.
- Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol*. 2013;30(3):713–724. <https://doi.org/10.1093/molbev/mss265>.
- Ho SY, Duchêne S. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol*. 2014;23(24):5947–5965. <https://doi.org/10.1111/mec.2014.23.issue-24>.
- Hofert M. Sampling exponentially tilted stable distributions. *ACM Trans Model Comput Simul (TOMACS)*. 2011;22(1):1–11. <https://doi.org/10.1145/2043635.2043638>.
- Horner DS, Lefkimmatis K, Reyes A, Gissi C, Saccone C, Pesole G. Phylogenetic analyses of complete mitochondrial genome sequences suggest a basal divergence of the enigmatic rodent anomalurus. *BMC Evol Biol*. 2007;7(1):1–12. <https://doi.org/10.1186/1471-2148-7-16>.
- Huchon D, Madsen O, Sibbald MJ, Ament K, Stanhope MJ, Catzeflis F, De Jong WW, Douzery EJ. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Mol Biol Evol*. 2002;19(7):1053–1065. <https://doi.org/10.1093/oxfordjournals.molbev.a004164>.
- Ji X, Zhang Z, Holbrook A, Nishimura A, Baele G, Rambaut A, Lemey P, Suchard MA. Gradients do grow on trees: a linear-time O(N)-dimensional gradient for statistical phylogenetics. *Mol*

- Biol Evol.* 2020;**37**(10):3047–3060. <https://doi.org/10.1093/molbev/msaa130>.
- Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc.* 1995;**90**(430):773–795. <https://doi.org/10.1080/01621459.1995.10476572>.
- Lepage T, Bryant D, Philippe H, Lartillot N. A general comparison of relaxed molecular clock models. *Mol Biol Evol.* 2007;**24**(12):2669–2680. <https://doi.org/10.1093/molbev/msm193>.
- Levine RA, Casella G. Optimizing random scan Gibbs samplers. *J Multivar Anal.* 2006;**97**(10):2071–2100. <https://doi.org/10.1016/j.jmva.2006.05.008>.
- Liu JS. *Monte Carlo strategies in scientific computing*. New York (NY): Springer Science & Business Media; 2008.
- McAuley JL, Gilbertson BP, Trifkovic S, Brown LE, McKimm-Breschkin JL. Influenza virus neuraminidase structure and functions. *Front Microbiol.* 2019;**10**:39. <https://doi.org/10.3389/fmicb.2019.00039>.
- Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol.* 2008;**25**(7):1459–1471. <https://doi.org/10.1093/molbev/msn090>.
- Neal RM. MCMC using Hamiltonian dynamics. In: Brooks S, Gelman A, Jones GL, Meng X-L, editors. *Handbook of Markov chain Monte Carlo*. Vol. 2. New York (NY): CRC Press; 2011. p. 113–162.
- Nishimura A, Suchard MA. Shrinkage with shrunken shoulders: Gibbs sampling shrinkage model posteriors with guaranteed convergence rates. *Bayesian Anal.* 2023;**18**(2):367–390. <https://doi.org/10.1214/22-BA1308>.
- Polson NG, Scott JG, Windle J. The Bayesian bridge. *J R Stat Soc B (Stat Methodol)*. 2014;**76**(4):713–733. <https://doi.org/10.1111/rssb.12042>.
- Rannala B, Yang Z. Inferring speciation times under an episodic molecular clock. *Syst Biol.* 2007;**56**(3):453–466. <https://doi.org/10.1080/10635150701420643>.
- Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 2002;**19**(1):101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>.
- Smith SA, Beaulieu JM, Donoghue MJ. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc Natl Acad Sci U S A.* 2010;**107**(13):5897–5902. <https://doi.org/10.1073/pnas.1001225107>.
- Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. Placental mammal diversification and the cretaceous–tertiary boundary. *Proc Natl Acad Sci U S A.* 2003;**100**(3):1056–1061. <https://doi.org/10.1073/pnas.0334222100>.
- Stan Development Team. Stan modeling language users guide and reference manual, version 2.17.0; 2017.
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 2018;**4**(1):vey016. <https://doi.org/10.1093/ve/vey016>.
- Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 1998;**15**(12):1647–1657. <https://doi.org/10.1093/oxfordjournals.molbev.a025892>.
- West M. On scale mixtures of normal distributions. *Biometrika* 1987;**74**(3):646–648. <https://doi.org/10.1093/biomet/74.3.646>.
- Wilson IA, Cox NJ. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu Rev Immunol.* 1990;**8**(1):737–787. <https://doi.org/10.1146/immunol.1990.8.issue-1>.
- Worobey M, Han G-Z, Rambaut A. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* 2014;**508**(7495):254–257. <https://doi.org/10.1038/nature13016>.
- Yang Z. *Molecular evolution: a statistical approach*. New York (NY): Oxford University Press; 2014.
- Yoder AD, Yang Z. Estimation of primate speciation dates using local molecular clocks. *Mol Biol Evol.* 2000;**17**(7):1081–1090. <https://doi.org/10.1093/oxfordjournals.molbev.a026389>.
- Zhang Y, Sutton C. Quasi-Newton methods for Markov chain Monte Carlo. *Adv Neural Inf Process Syst.* 2011;**24**:2393–2401.