Smarter dating moves and faster proposals: revisiting the phylogenetic relaxed clock model

Jordan Douglas[1*], Rong Zhang[1], Alexei J. Drummond[1,2], Remco Bouckaert[1]

1 Centre for Computational Evolution, School of Computer Science, University of Auckland, Auckland, New Zealand
2 School of Biological Sciences, University of Auckland, Auckland, New Zealand

* jordan.douglas@auckland.ac.nz

Abstract

Author summary

Introduction

Models

Preliminaries

Let $\mathcal{T}$ be a binary rooted time tree with N taxa (and $2N-2$ branches). Let L be the number of sites within the multiple sequence alignment D, and let L_{eff} be the *effective* number of sites in the alignment (ie. the number of site patterns). The posterior density of a phylogenetic model is described by

$$p(\mathcal{T}, \vec{\mathcal{R}}, \sigma, \theta | D) \propto p(D|\mathcal{T}, r(\vec{\mathcal{R}}), \theta)\ p(\mathcal{T}|\theta)\ p(\vec{\mathcal{R}}|\sigma)\ p(\sigma)\ p(\theta), \tag{1}$$

for rate standard deviation σ and other model parameters θ. $\vec{\mathcal{R}}$ is a vector of abstracted substitution rates, which is transformed into real rates by $r(\vec{\mathcal{R}})$. Three methods of representing rates as $\vec{\mathcal{R}}$ are presented in **Rate parameterisations**.

Under the *relaxed clock model*, each internal and leaf node is assigned a substitution rate $r_i = r(\mathcal{R}_i)$, which corresponds to its parent branch. There are a total of $|\vec{\mathcal{R}}| = 2N-2$ rates, which are independently distributed under the relaxed clock model prior [1].

Rate parameterisations

In Bayesian inference, the way parameters are represented in the model can affect the mixing ability of the model and the meaning of the model itself. Three methods for parameterising substitution rates are described below, and are later evaluated in **Results and Discussion**. Each parameterisation technique is associated with i) an abstraction of the rates $\vec{\mathcal{R}}$, ii) some function for transforming this parameter into real rates $r(\vec{\mathcal{R}})$, and iii) a prior density function of the abstraction $p(\vec{\mathcal{R}}|\sigma)$. The three methods are summarised in **Fig 1**.

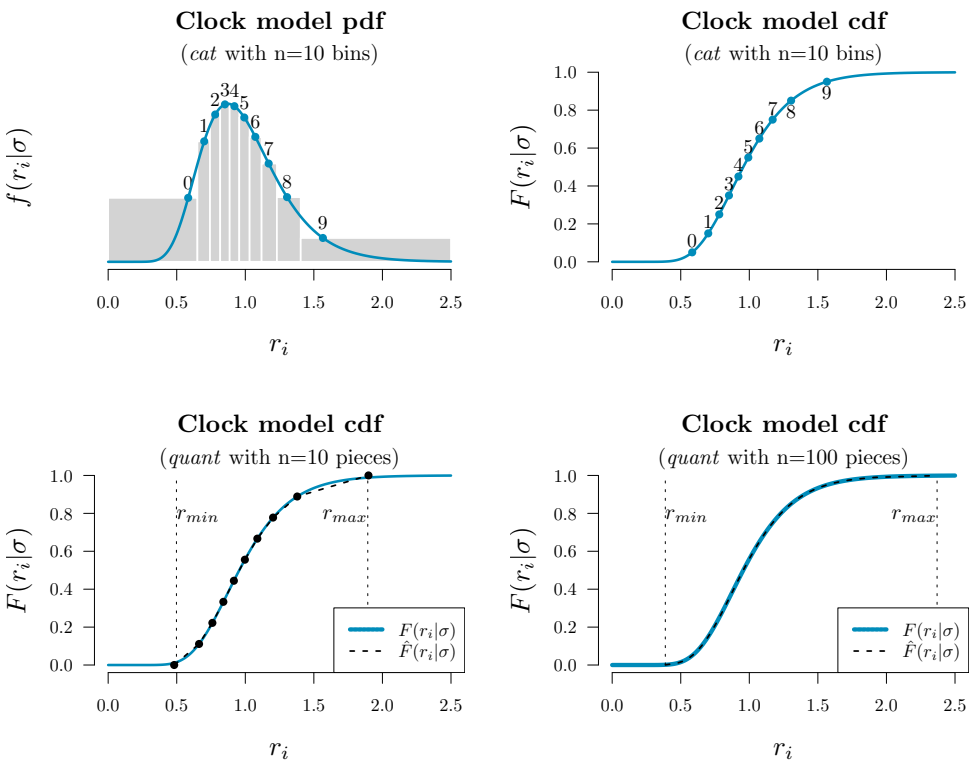

Fig 1. Methods of rate parameterisation. The *cat* and *quant* approximations are plotted on top of the true underlying rate prior distribution (*real*). In this example, rates are drawn from a LogNormal($\mu = -0.045, \sigma = 0.3$) distribution. The probability density function (pdf) and cumulative density function (cdf) of this distribution are shown.

1. Real rates

The natural (and unabstracted) parameterisation of a substitution rate is a real number $\mathcal{R}_i \in \mathbb{R}, \mathcal{R}_i > 0$ which is equal to the rate itself. Thus, under the *real* parameterisation:

$$r(\vec{\mathcal{R}}) = \vec{\mathcal{R}}\,. \tag{2}$$

Under the prior distribution $p(\vec{\mathcal{R}}\,|\sigma)$, rates are often log-normally or exponentially distributed with a mean of 1:

$$p(\mathcal{R}_i|\sigma) = \frac{1}{\mathcal{R}_i\sigma\sqrt{2\pi}} \exp\big(-\frac{(\ln \mathcal{R}_i - \mu)^2}{2\sigma^2}\big) \quad (\text{ LogNormal}(\mu, \sigma)\ ), \text{ or} \tag{3}$$

$$p(\mathcal{R}_i|\sigma) = p(\mathcal{R}_i) = e^{-\mathcal{R}_i} \quad (\text{ Exponential}(\lambda = 1)\ ) \tag{4}$$

where $\mu = -0.5\sigma^2$ is set such that the expected value of the log-normal distribution is 1.

Zhang and Drummond 2020 present a series of tree operators which propose internal/root node heights, and then recompute the rates of incident branches such that their genetic distances ($r_i \times \tau_i$) remain constant after the proposal. By maintaining genetic distances the likelihood can also be maintained. These operators account for the

correlation which exists between branch rates and branch times – a correlation which is induced by the likelihood function.

2. Categories

The category parameterisation (*cat*) is an abstraction of the *real* parameterisation. Each branch is assigned an integer from 0 to $n-1$:

$$\vec{\mathcal{R}} \in \{0, 1, \dots, n-1\}^{2N-2}. \tag{5}$$

The domain of $\vec{\mathcal{R}}$ is uniformly distributed:

$$p(\mathcal{R}_i|\sigma) = p(\mathcal{R}_i) = \frac{1}{n}. \tag{6}$$

Let $f(x|\sigma)$ be the probability density function (pdf) and let $F(x|\sigma) = \int_0^x f(t|\sigma)\, dt$ be the cumulative distribution function (cdf) of the prior distribution used by the underlying *real* clock model. Then, in the *cat* parameterisation, $f(x|\sigma)$ is discretised into n bins and the elements of $\vec{\mathcal{R}}$ each point to one of these bins. Each bin contains uniform probability density $1/n$. The rate of each bin is equal to the median value within the bin

$$r(\mathcal{R}_i) = F^{-1}\big(\frac{\mathcal{R}_i + 0.5}{n}\big), \tag{7}$$

where F^{-1} is the inverse cumulative distribution function (i-cdf).

The key advantage of the *cat* parameterisation is the removal of a term from the posterior density (Equation 1), or more accurately the replacement of a non-trivial $p(\vec{\mathcal{R}}\,|\sigma)$ term with that of a uniform prior. Thus, one fewer term needs to be estimated per rate.

This method was suggested in the original BEAST2 relaxed clock paper [1] and has been widely used. However, the constant distance operators since introduced by Zhang and Drummond 2020 – which are incompatible with the *cat* parameterisation – yield an increase in mixing rate under *real* by up to an order of magnitude over that of *cat*.

3. Quantiles

Finally, rates can be parameterised as real numbers $0 < \mathcal{R}_i < 1$ which describe the rate's quantile with respect to some underlying clock model distribution. Under the *quant* parameterisation, each element in $\vec{\mathcal{R}}$ is uniformly distributed.

$$\vec{\mathcal{R}} \in \mathbb{R}^{2N-2}, 0 < \mathcal{R}_i < 1 \tag{8}$$

$$p(\mathcal{R}_i|\sigma) = p(\mathcal{R}_i) = 1 \tag{9}$$

Transforming these quantiles into rates invokes the i-cdf of the underlying *real* clock model distribution. Thus, while this approach has clear similarities with *cat*, the domain of rates here is continuous (as opposed to being confined to a discrete number of bins) and is therefore compatible with the class of operators described by Zhang and Drummond 2020.

A potential disadvantage of the *quant* method would be the computational requirements of continuously evaluating the i-cdf, especially for trees with large N.

Hence, rather than evaluating the exact i-cdf F^{-1}, an approximation $\hat{F}^{-1}$ will be used instead:

$$r(\mathcal{R}_i) = \hat{F}^{-1}(\mathcal{R}_i). \tag{10}$$

In this article we have extended *quant* through a linear piecewise approximation of the i-cdf. As the piecewise approximation is linear, evaluating the derivatives $\frac{\partial}{\partial \mathcal{R}_i}\hat{F}^{-1}(\mathcal{R}_i) = D\hat{F}^{-1}(\mathcal{R}_i)$ and $\frac{\partial}{\partial r_i}\hat{F}(r_i) = D\hat{F}(r_i)$ – which are required for computing Hastings ratios – is trivial. The approximation is comprised of n pieces (where n is fixed) and upper and lower rate boundaries $r_{\min}$ and $r_{\max}$. The approximation is displayed in **Fig 1** and further detailed in **S1 Appendix**.

Zhang and Drummond 2020 introduced several tree operators for the *real* parameterisation – including **Constant Distance**, **Simple Distance**, and **Small Pulley**. In this project, we extended these three operators so that they are compatible with the *quant* parameterisation. These are presented in **S1 Appendix**.

Bactrian proposal kernel

The step size of a proposal kernal $q(x'|x)$ should be such that the proposed state x' is sufficiently far from the current state x to explore vast areas of parameter space, but not so large that the proposal is rejected too often [2]. Yang et al. have challenged the widely used uniform proposal kernel in place of the Bactrian kernel [3, 4]. The Bactrian(m) distribution is defined as the sum of two Normal distributions:

$$\Sigma \sim \text{Bactrian}(m) \equiv \frac{1}{2}\text{Normal}(-m, 1-m^2) + \frac{1}{2}\text{Normal}(m, 1-m^2) \tag{11}$$

where $0 \leq m < 1$ describes the modality of the Bactrian distribution. When $m = 0$, the Bactrian distribution is equivalent to a Normal$(0, 1)$ distribution. As $m \to 1$, the distribution becomes increasingly bimodal (**Fig. 2**). Yang et al. 2013 [3] suggest that Bactrian$(m = 0.95)$ yields a proposal kernel superior to the uniform kernel, by placing minimal probability on steps which are too small or too large.

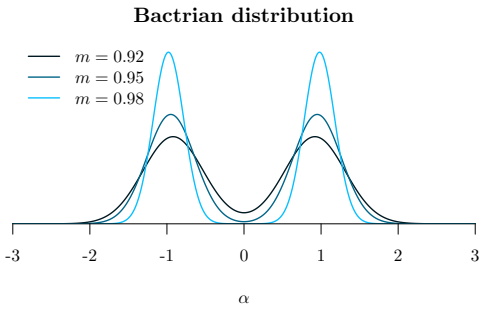

Fig 2. The Bactrian proposal kernel. Y-axis corresponds to probability density $f(\Sigma|m)$.

In this article we compare the performance of uniform and Bactrian proposal kernels in the clock model. Two Bactrian distributions are compared($m = 0.95$ and $m = 0.98$). The clock model operators which these proposal kernels apply to are described in **Table 1**.

	Operator	Proposal	Parameters
1	**Random walk operator**	$x' \leftarrow x + s\Sigma$	σ, r, q
2	**Scale operator**	$x' \leftarrow x \times e^{s\Sigma}$	σ, r
3	**Interval operator**	$y \leftarrow \frac{u-x}{x-l} \times e^{s\Sigma}$ $x' \leftarrow \frac{u+l*y}{y+1}$	q $(l = 0, u = 1)$
4	**Constant distance operators**	$x' \leftarrow x + s\Sigma$	t

Table 1. Summary of proposal kernels $q(x'|x)$ of clock model operators. In each operator, Σ is drawn from either a Bactrian(m) or Uniform distribution (distributions are normalised so that they have a mean of 0 and a variance of 1). The scale size s is tunable. The proposal kernel may apply to node heights t, clock standard deviation σ, clock rates r (*real* only), and clock rate quantiles q (*quant* only). The Scale operator acts on parameters with non-negative domains. The Interval operator proposes values which respect its domain ie. $l < x' < u$.

Narrow Exchange Rate

The **Narrow Exchange** operator [5], widely used in BEAST [6] and BEAST2 [7], is similar to NNI, and works as follows (**Fig. 3**):

Step 1. Sample an internal/root node E from tree $\mathcal{T}$, where E has grandchildren.

Step 2. Identify the child of E with the greater height. Denote this child as D and its sibling as C (ie. $t_D > t_C$).

Step 3. Randomly identify the two children of D as A and B.

Step 4. Relocate the $B - D$ branch onto the $C - E$ branch, so that B and C become siblings and their parent is D. All node heights are unchanged.

Lakner et al. 2008 [8] found that tree operators which perterb topology (such as NNI and SPR) consistently perform better than those which also change branch lengths (such as LOCAL [9] and Continuous Change [10]). If Narrow Exchange was adapted to the relaxed clock model by ensuring that genetic distances remain constant after the proposal, its performance may be improved even further. This may in turn permit proposing a new node height t_D and therefore changing branch (time) lengths.

Here, we present the **Narrow Exchange Rate** (NER) operator. Let r_A, r_B, r_C, and r_D be the clock rates of nodes A, B, C, and D, respectively. In addition to the modest topological change applied by Narrow Exchange, NER also proposes new clock rates $r_A{}'$, $r_B{}'$, $r_C{}'$, and $r_D{}'$. While NER does not alter t_D (ie. $t_D{}' \leftarrow t_D$), we also consider NERw - a special case of the NER operator which embarks t_D on a random walk:

$$t_D{}' \leftarrow t_D + s\Sigma \tag{12}$$

for random walk step size $s\Sigma$ where s is a tunable scaler parameter and Σ is drawn from a uniform or **Bactrian proposal kernel**. NER (and NERw) are compatible with both the *real* and *quant* parameterisations. Analagous to the Constant Distance operator, new rates are proposed such that genetic distances between nodes A, B, C, and E are maintained. Thus, there are $\binom{4}{2} = 6$ pairwise distance constraints.

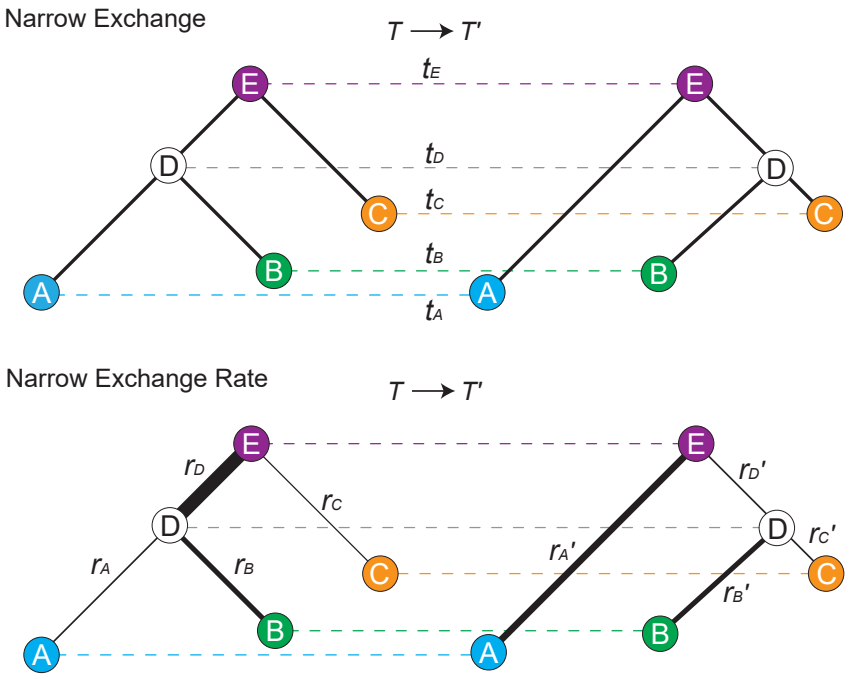

Fig 3. Depiction of Narrow Exchange and Narrow Exchange Rate operators. Proposals are denoted by $\mathcal{T} \to \mathcal{T}'$. The vertical axis corresponds to node height t. In the bottom figure, branch rates r are indicated by line thickness. In this example, the $\mathcal{D}_{AE}$ and $\mathcal{D}_{CE}$ constraints are satisfied.

$$\begin{aligned}
\mathcal{D}_{AB}: \quad & r_A(t_D - t_A) + r_B(t_D - t_B) = \\
& r_A'(t_E - t_A) + r_D'(t_E - t_D') + r_B'(t_D' - t_B) && (13)\\
\mathcal{D}_{AC}: \quad & r_A(t_D - t_A) + r_D(t_E - t_D) + r_C(t_E - t_C) = \\
& r_A'(t_E - t_A) + r_D'(t_E - t_D') + r_C'(t_D' - t_C) && (14)\\
\mathcal{D}_{AE}: \quad & r_A(t_D - t_A) + r_D(t_E - t_D) = \\
& r_A'(t_E - t_A) && (15)\\
\mathcal{D}_{BC}: \quad & r_B(t_D - t_B) + r_D(t_E - t_D) + r_C(t_E - t_D) = \\
& r_B'(t_D' - t_B) + r_C'(t_D' - t_C) && (16)\\
\mathcal{D}_{BE}: \quad & r_B(t_D - t_B) + r_D(t_E - t_D) = \\
& r_B'(t_D' - t_B) + r_D'(t_E - t_D') && (17)\\
\mathcal{D}_{CE}: \quad & r_C(t_E - t_C) = \\
& r_C'(t_D' - t_C) + r_D'(t_E - t_D') && (18)
\end{aligned}$$

Further constraints are imposed by the model itself:

$$r_i > 0 \text{ and } r_i' > 0 \text{ for } i \in \{A, B, C, D\} \quad (19)$$

$$\max\{t_B, t_C\} < t_D' < t_E. \quad (20)$$

Unfortunately, it is not possible to solve all six $\mathcal{D}_{ij}$ constraints without permitting non-positive rates or illegal trees. Therefore rather than conserving all six pairwise

distances, NER conserves a *subset* of distances. It is not immediately clear which distances should be conserved.

Automated generation of operators and constraint satisfaction

The total space of NER operators is comprised of all possible subsets of distance constraints (ie. $\{\}, \{\mathcal{D}_{AB}\}, \{\mathcal{D}_{AC}\}, \ldots, \{\mathcal{D}_{AB}, \mathcal{D}_{AC}, \mathcal{D}_{AE}, \mathcal{D}_{BC}, \mathcal{D}_{BE}, \mathcal{D}_{CE}\}$) which are solvable. The simplest NER – the null operator denoted by NER$\{\}$ – does not satisfy any distance constraints. This is equivalent to Narrow Exchange.

As it is unclear which NER variants would perform the best, we developed an automated pipeline for generating and testing these operators.

1. Solution finding. Using standard analytical linear-system solving libraries in MATLAB, the $2^6 = 64$ subsets of distance constraints are solved. 54 out of the 64 subsets were found to be solvable, and the unsolvables were discarded.

2. Solving Jacobian determinants. The determinant of the Jacobian matrix J is required for computng the Hastings ratio of the proposal. J is defined as

$$J = \begin{bmatrix} \frac{\partial r_A{}'}{\partial r_A} & \frac{\partial r_A{}'}{\partial r_B} & \frac{\partial r_A{}'}{\partial r_C} & \frac{\partial r_A{}'}{\partial r_D} \\ \frac{\partial r_B{}'}{\partial r_A} & \frac{\partial r_B{}'}{\partial r_B} & \frac{\partial r_B{}'}{\partial r_C} & \frac{\partial r_B{}'}{\partial r_D} \\ \frac{\partial r_C{}'}{\partial r_A} & \frac{\partial r_C{}'}{\partial r_B} & \frac{\partial r_C{}'}{\partial r_C} & \frac{\partial r_C{}'}{\partial r_D} \\ \frac{\partial r_D{}'}{\partial r_A} & \frac{\partial r_D{}'}{\partial r_B} & \frac{\partial r_D{}'}{\partial r_C} & \frac{\partial r_D{}'}{\partial r_D} \end{bmatrix}. \tag{21}$$

Computing the determinant $|J|$ invokes standard analytical differentiation and linear algebra libraries of MATLAB. 6 of the 54 operators were found to have $|J| = 0$, corresponding to irreversible proposals, and were discarded.

3. Automated generation of BEAST2 operators. Java class files are generated using string processing. Each class corresponds to a single operator, extends the class of a meta-NER-operator, and is comprised of the solutions found in **1** and the Jacobian determinant found in **2**. $|J|$ is further augmented if the *quant* parameterisation is employed.

The 48 operators generated by this pipeline are evaluated and compared in **Results and Discussion**. Each operator is considered with and without a random walk on t_D and thus there are 96 total settings.

Clock model averaging

Results and Discussion

Assessment criteria and datasets

To avoid a cross-product explosion, the three targets for clock model improvement (**Rate parameterisations**, **Bactrian proposal kernel**, and **Narrow Exchange Rate**) are evaluated sequentially, in the order presented in this paper. The setting(s) which are considered to be the best in each step are then incorporated into the following step. This protocol and its outcomes are summarised in **Fig. 4**.

Methodologies are assessed according to the following criteria.

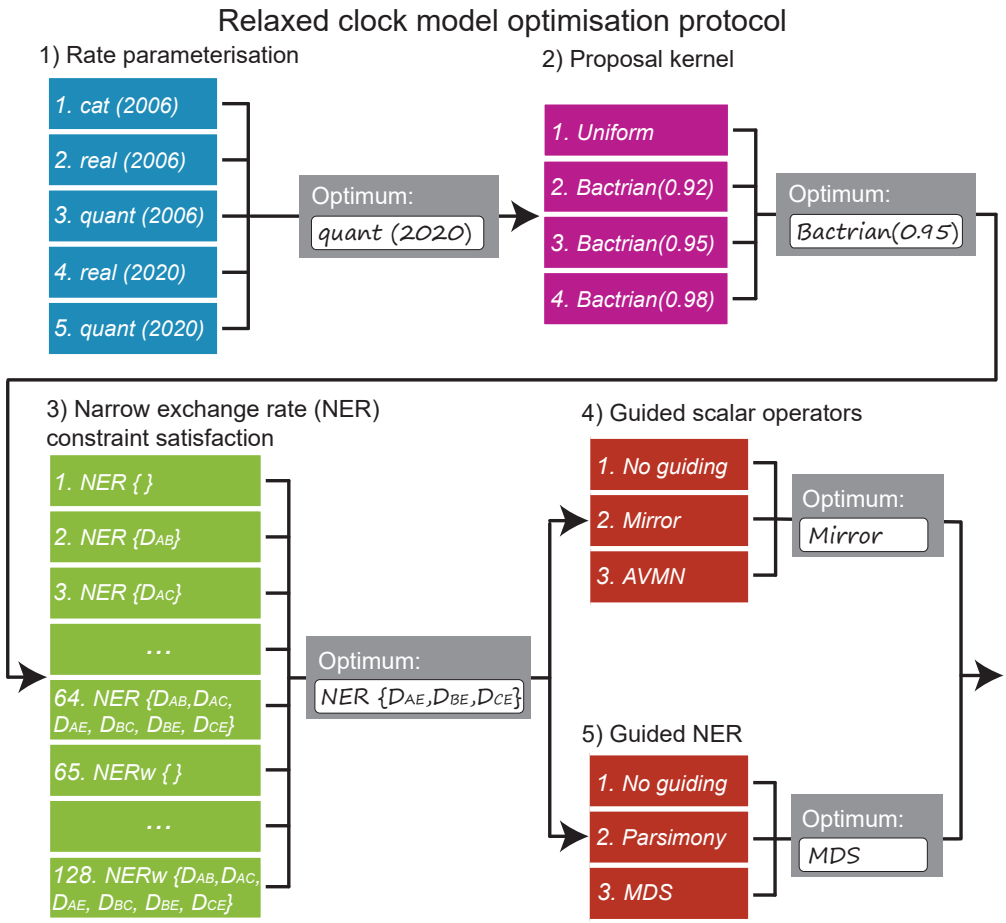

Fig 4. Protocol for optimising methodology settings. The three areas (detailed in **Models**) are optimised sequentially, where the best setting from each step is used when optimising the following step.

1. Validation. This is assessed by measuring the coverage of all estimated parameters in a well-calibrated simulation study, using 100 simulated datasets (with $N = 100$ taxa and $L = 5000$ nucleotide alignments). These are presented in **S2 Appendix**.

2. Time to convergence. Two independent MCMC chains are run and the time is measured until: a) the absolute difference in clade posterior probability between the two chains is less than 0.05 for all clades, b) the Rubin-Gelman statistic $\hat{R}$ [11] of every estimated parameter is less than 1.05, and c) the effective sample size [12] of every estimated parameter is greater than 100 in each chain.

3. Mixing of parameters. Key parameters are evaluated for the number of effective samples generated per hour (ESS/hr).

For the latter two criteria, methodologies are benchmarked using the empirical datasets compiled and partitioned [13] by Lanfear [14]. Two datasets (the turtle data by Crawford et al. 2012 [15]; and the songbird data by Moyle et al. 2016 [16]) are used during the optimisation stage, where each setting is run 15× and the average statistics across the $15 \times 2 = 30$ chains are reported. All 25 datasets presented in **Table 2** are benchmarked together 1× each at the end of the protocol. Datasets and BEAST2 .xml templates are provided in the GitHub repository accompanying this article.

Methodologies are benchmarked using the Intel(R) Xeon(R) Gold 6138 CPU (2.00 GHz), with two processing units per pair of MCMC chains.

	N	P	L (kb)	L_{eff} (kb)	Ref.
1	6	1/4/8/16	0.4/2.1/10.2/13.7	0.1/0.5/1/1.1	Richart 2015 [17]
2	**10**	1/4/8/**16**	0.4/1.5/3.1/**6.7**	0.1/0.3/0.5/**0.9**	**Crawford 2012** [15]
3	11	1/4/8/16	0.5/2.2/4.9/10.1	0.1/0.4/0.7/1.3	Leache 2015 [18]
4	18	1/4/8/16	0.4/1.8/2.9/6.7	0.1/0.3/0.4/0.9	Meiklejohn 2016 [19]
5	27	1/4/8/16	0.8/1.4/2/4.3	0.3/0.6/1/1.7	Faircloth 2013 [20]
6	33	1/4/8/16	0.4/1.7/2.6/5.1	0.1/0.5/0.7/1.5	McCormack 2013 [21]
7	38	1/4	0.4/2.1	0.2/0.6	Bergsten 2013 [22]
8	38	1/4/8/16	1/2.4/8.1/14.9	0.7/1.8/5.6/10.1	Ran 2018 [23]
9	41	1/3	0.4/1.7	0.3/1.1	Brown 2012 [24]
10	44	1/3	0.6/1.9	0.2/0.8	Cognato 2001 [25]
11	44	1/4/7	0.8/2.9/5.9	0.3/0.8/1.8	Dornburg 2012 [26]
12	51	1/4/6	0.6/3.5/5.4	0.1/0.9/1.8	Sauquet 2011 [27]
13	61	1/4/8/16	0.9/4.2/7.5/15	0.7/2.8/4.6/9.6	Broughton 2013 [28]
14	69	1/2	0.7/0.8	0.1/0.1	Devitt 2013 [29]
15	70	1/3	0.7/2.2	0.3/0.9	Kawahara 2013 [30]
16	78	1/4/8/16	0.4/1.9/4.1/7.2	0.3/1.7/3.6/6.2	Cannon 2016 [31]
17	79	1/4/8/16	0.1/1.3/0.8/4.5	0/0.1/0.1/0.8	Oaks 2011 [32]
18	94	1/4/8/11	0.1/1.8/3/3.7	0.1/0.8/1.3/1.7	Rightmyer 2013 [33]
19	**106**	1/**4**/8/16	0.6/**3.1**/5.1/11.6	0.3/**1.2**/2/4.5	**Moyle 2016** [16]
20	110	1/4/8/16	0.5/2.1/3.8/7.1	0.3/1.4/2.5/4.3	Fong 2012 [34]
21	152	1/4/5	1/3.5/3.6	0.6/1.6/1.6	Day 2013 [35]
22	187	1/4/8/16	0.3/0.9/1.4/3.5	0.2/0.7/1.1/2.9	Branstetter 2017 [36]
23	197	1/4/8/14	0.9/3.1/6.1/11.6	0.3/1.6/3.5/6.3	Horn 2014 [37]
24	235	1/4/8/16	2.3/6.4/8.5/25.7	0.9/2.3/3.2/12.2	Reddy 2017 [38]
25	237	1/4/5	0.8/2.4/3.1	0.2/1.2/1.5	Murray 2013 [39]

Table 2. Empirical datasets used during benchmarking, sorted in increasing order of taxa count N. Number of partitions P, total alignment length L, and number of patterns L_{eff} are also specified. Where a dataset is sampled more than once, the dimensions of its multiple samples are separated by '/'. The datasets in bold (Crawford [15] and Moyle [16]) are used during the optimisation protocol.

Comparison of rate parameterisations

We compared the three rate parameterisations described in **Rate parameterisations**. All three settings use the standard BEAST2 clock model operators from Drummond et al. 2006 [1]. *real* and *quant* additionally use the constant-distance tree operators described by Zhang and Drummond 2020. To determine whether the difference in performance between *real*/*quant* versus *cat* is because of the constant-distance tree operators or the parameterisation itself, we also included benchmarked two additional settings: *real 2006* and *quant 2006*, which do not used the constant-distance operators.

These five settings are validated in **S2 Appendix**.

Fig. 5 shows that the *real 2006* performs considerably worse than any of the other settings. This is due to the poor sampling of the prior under this setting (ie. low ESS of $p(\theta)$). The failure of *real 2006* thus highlights the appeal of 1) the *cat* or *quant* parameterisations, both of which have trivial contributions to the prior density (ie. uniform priors), and 2) the smarter operators used by *real* (Zhang and Drummond 2020). Due to its computational burden, *real 2006* was not benchmarked for all of the datasets in **Table 2**.

Our results show that the *quant* parameterisation yields the best performance with respect to effective samples per hour. *quant* outperforms *quant 2006*, suggesting that the constant distance operators are effective. Furthermore, *quant* outperforms *real* especially at sampling from the posterior and prior distributions (ie. high ESS/hr for $p(\theta|D)$ and $p(\theta)$). This is most likely because of the uniform prior distribution of rate quantiles.

Overall, *quant* yields a median ESS/hr approximately 150 % faster with respect to sampling the posterior probability, and approximately 30 % faster with respect to sampling branch rates r and clock standard deviation σ, compared with *real*.

Comparison of Bactrian and uniform proposal kernels on the clock model

Comparison of NER variants

We compared the Narrrow Exchange Rate (NER) operators described in **Narrow Exchange Rate**. This protocol selects the best among 48 NER (no random walk) and 48 NERw (Bactrian(0.95) random walk) operators, and has two phases. First, the best of the 96 is selected by comparing operator acceptance rates on simulated data. Second, the selected operator is benchmarked with respect to convergence time and sampling rate on real data (**Table 2**). All the analyses in this section invoke the *quant* parameterisation and Bactrian(0.95) proposal kernels on clock model parameters.

Initial screening by acceptance rate

We selected the best operator variant by performing MCMC on 100 simulated datasets, where each MCMC employed all 96 NER/NERw variants. The acceptance rate of each operator is compared to that of the null operator NER{} (ie. Narrow Exchange).

Fig. 6 shows that NER variants which satisfy the genetic distances between nodes B and A (ie. $\mathcal{D}_{AB}$) or between B and C (ie. $\mathcal{D}_{BC}$) consistently perform worse than the standard Narrow Exchange operator, where B is the node being interchanged from the A branch to the C branch (**Fig. 3**). This is an intuitive result. If the posterior distribution is relatively flat, and the data presents high uncertainty in the positioning of B, with respect to A and C, then the topological rearrangement performed by Narrow Exchange will be favoured. However, this uncertainty in the *topology* is likely coupled with uncertainty in the *distance* between B and A or between B and C. Thus, in this case, respecting the $\mathcal{D}_{AB}$ and $\mathcal{D}_{BC}$ constraints (by proposing branch rates) makes too many unnecessary changes to the state and the operator performs worse.

Fig. 6 also reveals a cluster of NER variants which – under the conditions of the simulation – performed better than the null operator NER{} around 10% of the time and performed worse around 40% of the time. One such operator is NER$\{\mathcal{D}_{AE}, \mathcal{D}_{BE}, \mathcal{D}_{CE}\}$. This variant conserves the genetic distance between all child nodes A, B, and C, and the grandparent node E. This is performed by proposing rates for r_A, r_B, and r_C while obeying the distance constraints imposed by the operator. Exploring this operator further, we can see that NER$\{\mathcal{D}_{AE}, \mathcal{D}_{BE}, \mathcal{D}_{CE}\}$ is at its best

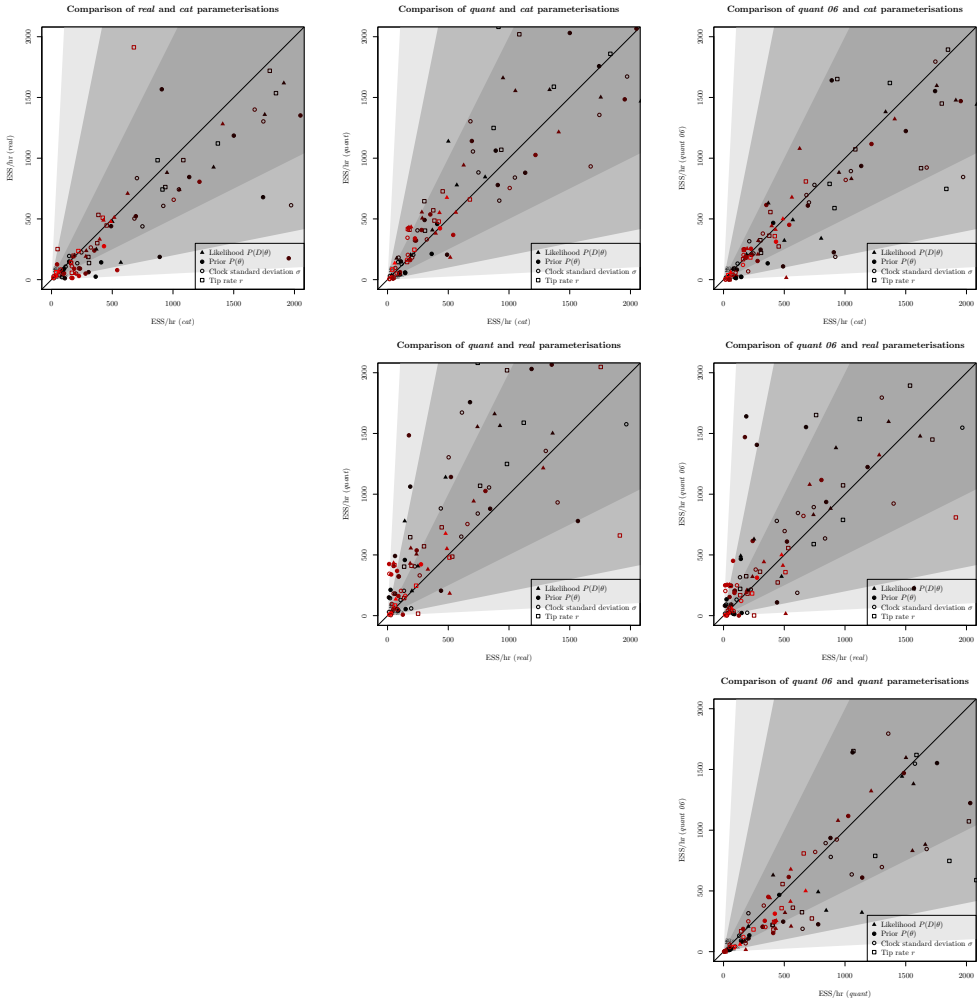

Fig 5. Rate parameterisation evaluation. Comparison of ESS/hr (averaged across two independent MCMC chains) with respect to relevant terms – L: likelihood, p: prior density, r: clock rate averaged across all leaves, σ: clock standard deviation. Each point is from one partition sample of the empirical data in **Table 2**.

when there is a large variance in branch rate ie. when clock standard deviation σ is high ($\sigma \gtrsim 0.3$), corresponding to data which is not clock-like. On the other hand, NER{} is much preferred when the operator's acceptance rate is high ($\gtrsim 0.15$), which corresponds to poorer signal in the data. Overall, NER$\{\mathcal{D}_{AE}, \mathcal{D}_{BE}, \mathcal{D}_{CE}\}$ outperforms the standard Narrow Exchange operator provided that the data is not clock-like and contains strong signal.

Finally, **Fig. 6** shows that by applying a (Bactrian) random walk to t_D – the height of internal node D – the acceptance rate of NER plummets dramatically. This effect is most dominant for the NER variants which satisfy distance constraints (ie. the operators which are not NER{}). This result is unfortunate however not unexpected, and is consistent with Lakner et al. 2008 [8], who observed that tree operators perform best when they change either topology, or branch lengths, but not both.

Although there are several operators tying for first place, we selected the NER$\{\mathcal{D}_{AE}, \mathcal{D}_{BE}, \mathcal{D}_{CE}\}$ operator to proceed to the next round of optimisation.

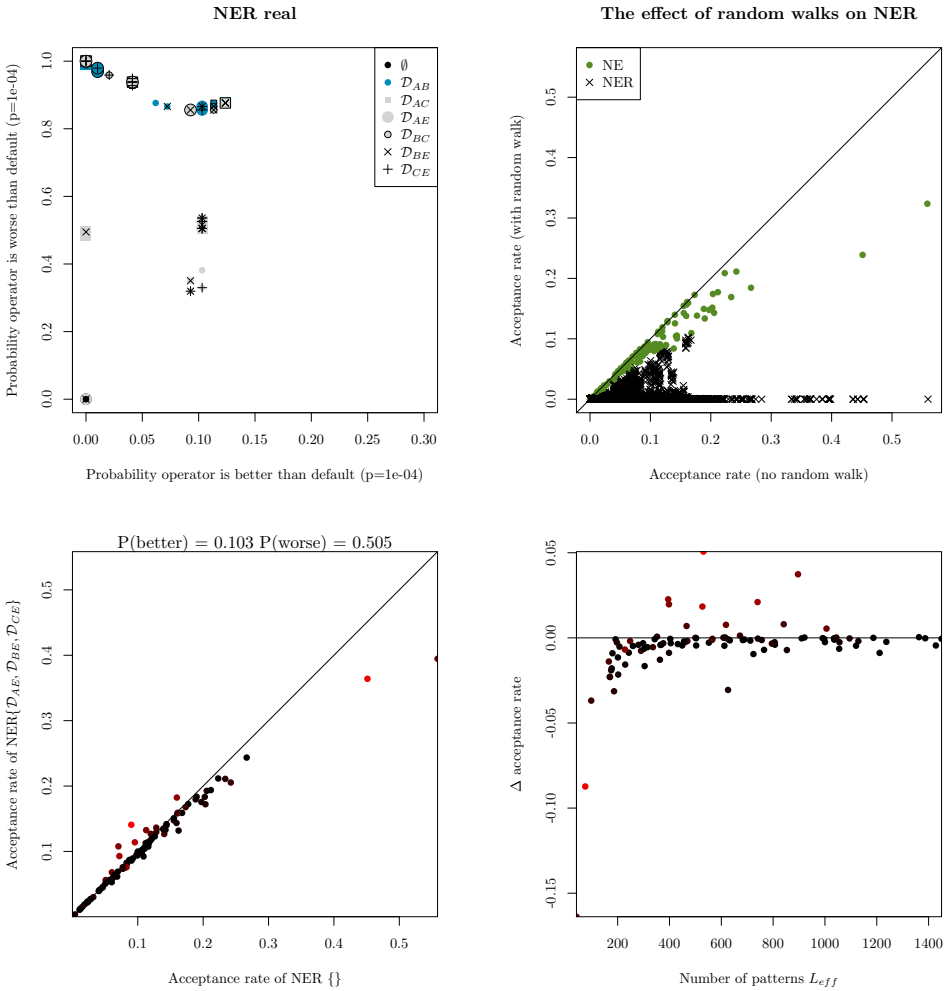

Fig 6. Screening of NER and NERw variants by acceptance rate. Left: comparison of NER variants with the null operator NER{} (ie. Narrow Exchange). Each of the 48 operators are represented by a single point, uniquely encoded by the point stylings. The number of times each operator is proposed and accepted is compared with that of NER{}, and a Z-test is performed to determine whether there is a significant difference in the two acceptance rates. This process is repeated for each of 100 simulated datasets. The axes of each plot are the proportion of these 100 simulations for which there is significant evidence that the operator is better than NER{} (x-axis) or worse than NER{} (y-axis). Right: comparison of NER with NERw. Each point is one NER/NERw variant from a single simulation. This plot shows that the random walk hinders the acceptance rates of all NER variants, including NER{}.

Advanced screening by congergence time

Conclusion

Supporting information

S1 Appendix. Rate quantiles. The linear piecewise approximation used in the *quant* parameterisation is described. Tree operators presented by Zhang and Drummond 2020 are extended to the *quant* parameterisation.

S2 Appendix. Well-calibated simulation studies. Methodologies are validated using well-calibated simulation studies.

References

1. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS biology. 2006;4(5):e88.

2. Roberts GO, Gelman A, Gilks WR, et al. Weak convergence and optimal scaling of random walk Metropolis algorithms. The annals of applied probability. 1997;7(1):110–120.

3. Yang Z, Rodríguez CE. Searching for efficient Markov chain Monte Carlo proposal kernels. Proceedings of the National Academy of Sciences. 2013;110(48):19307–19312.

4. Thawornwattana Y, Dalquen D, Yang Z, et al. Designing simple and efficient Markov chain Monte Carlo proposal kernels. Bayesian Analysis. 2018;13(4):1037–1063.

5. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 2002;161(3):1307–1320.

6. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012;29(8):1969–1973.

7. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS computational biology. 2019;15(4):e1006650.

8. Lakner C, Van Der Mark P, Huelsenbeck JP, Larget B, Ronquist F. Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. Systematic biology. 2008;57(1):86–103.

9. Simon D, Larget B. Bayesian analysis in molecular biology and evolution (BAMBE) http://www.mathcs.duq.edu/larget/bambe.html. Pittsburgh, Pennsylvania. 1998;.

10. Jow H, Hudelot C, Rattray M, Higgs P. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Molecular Biology and Evolution. 2002;19(9):1591–1601.

11. Gelman A, Rubin DB, et al. Inference from iterative simulation using multiple sequences. Statistical science. 1992;7(4):457–472.

12. Rambaut A, Drummond A. Tracer 1.6. University of Edinburgh, Edinburgh. UK. Technical report; 2013.

13. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular biology and evolution. 2016;34(3):772–773.

14. Lanfear R. BenchmarkAlignments https://github.com/roblanf/BenchmarkAlignments. GitHub. 2019;.

15. Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters. 2012;8(5):783–786. doi:10.1098/rsbl.2012.0331.

16. Moyle RG, Oliveros CH, Andersen MJ, Hosner PA, Benz BW, Manthey JD, et al. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. Nature Communications. 2016;7(1). doi:10.1038/ncomms12709.

17. Richart CH, Hayashi CY, Hedin M. Phylogenomic analyses resolve an ancient trichotomy at the base of Ischyropsalidoidea (Arachnida, Opiliones) despite high levels of gene tree conflict and unequal minority resolution frequencies. Molecular Phylogenetics and Evolution. 2016;95:171–182. doi:10.1016/j.ympev.2015.11.010.

18. Leaché AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD, Linkem CW. Phylogenomics of Phrynosomatid Lizards: Conflicting Signals from Sequence Capture versus Restriction Site Associated DNA Sequencing. Genome Biology and Evolution. 2015;7(3):706–719. doi:10.1093/gbe/evv026.

19. Meiklejohn KA, Faircloth BC, Glenn TC, Kimball RT, Braun EL. Analysis of a Rapid Evolutionary Radiation Using Ultraconserved Elements: Evidence for a Bias in Some Multispecies Coalescent Methods. Systematic Biology. 2016;65(4):612–627. doi:10.1093/sysbio/syw014.

20. Faircloth BC, Sorenson L, Santini F, Alfaro ME. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). PLoS ONE. 2013;8(6):e65923. doi:10.1371/journal.pone.0065923.

21. McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT. A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment and High-Throughput Sequencing. PLoS ONE. 2013;8(1):e54848. doi:10.1371/journal.pone.0054848.

22. Bergsten J, Nilsson AN, Ronquist F. Bayesian Tests of Topology Hypotheses with an Example from Diving Beetles. Systematic Biology. 2013;62(5):660–673. doi:10.1093/sysbio/syt029.

23. Ran JH, Shen TT, Wang MM, Wang XQ. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. Proceedings of the Royal Society B: Biological Sciences. 2018;285(1881):20181012. doi:10.1098/rspb.2018.1012.

24. Brown RM, Siler CD, Das I, Min Y. Testing the phylogenetic affinities of Southeast Asia's rarest geckos: Flap-legged geckos (Luperosaurus), Flying geckos (Ptychozoon) and their relationship to the pan-Asian genus Gekko. Molecular Phylogenetics and Evolution. 2012;63(3):915–921. doi:10.1016/j.ympev.2012.02.019.

25. Cognato AI, Vogler AP. Exploring Data Interaction and Nucleotide Alignment in a Multiple Gene Analysis of Ips (Coleoptera: Scolytinae). Systematic Biology. 2001;50(6):758–780. doi:10.1080/106351501753462803.

26. Dornburg A, Moore JA, Webster R, Warren DL, Brandley MC, Iglesias TL, et al. Molecular phylogenetics of squirrelfishes and soldierfishes (Teleostei: Beryciformes: Holocentridae): Reconciling more than 100 years of taxonomic confusion. Molecular Phylogenetics and Evolution. 2012;65(2):727–738. doi:10.1016/j.ympev.2012.07.020.

27. Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, et al. Testing the Impact of Calibration on Molecular Divergence Times Using a Fossil-Rich Group: The Case of Nothofagus (Fagales). Systematic Biology. 2011;61(2):289–313. doi:10.1093/sysbio/syr116.

28. Broughton RE, Betancur-R R, Li C, Arratia G, Ortí G. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Currents. 2013;doi:10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e.

29. Devitt TJ, Devitt SEC, Hollingsworth BD, McGuire JA, Moritz C. Montane refugia predict population genetic structure in the Large-blotched Ensatina salamander. Molecular Ecology. 2013;22(6):1650–1665. doi:10.1111/mec.12196.

30. Kawahara AY, Rubinoff D. Convergent evolution of morphology and habitat use in the explosive Hawaiian fancy case caterpillar radiation. Journal of Evolutionary Biology. 2013;26(8):1763–1773. doi:10.1111/jeb.12176.

31. Cannon JT, Vellutini BC, Smith J, Ronquist F, Jondelius U, Hejnol A. Xenacoelomorpha is the sister group to Nephrozoa. Nature. 2016;530(7588):89–93. doi:10.1038/nature16520.

32. Oaks JR. A TIME-CALIBRATED SPECIES TREE OF CROCODYLIA REVEALS A RECENT RADIATION OF THE TRUE CROCODILES. Evolution. 2011;65(11):3285–3297. doi:10.1111/j.1558-5646.2011.01373.x.

33. RIGHTMYER MG, GRISWOLD T, BRADY SG. Phylogeny and systematics of the bee genus Osmia (H ymenoptera: M egachilidae) with emphasis on N orth A merican Melanosmia : subgenera, synonymies and nesting biology revisited. Systematic Entomology. 2013;38(3):561–576. doi:10.1111/syen.12013.

34. Fong JJ, Brown JM, Fujita MK, Boussau B. A Phylogenomic Approach to Vertebrate Phylogeny Supports a Turtle-Archosaur Affinity and a Possible Paraphyletic Lissamphibia. PLoS ONE. 2012;7(11):e48990. doi:10.1371/journal.pone.0048990.

35. Day JJ, Peart CR, Brown KJ, Friel JP, Bills R, Moritz T. Continental Diversification of an African Catfish Radiation (Mochokidae: Synodontis). Systematic Biology. 2013;62(3):351–365. doi:10.1093/sysbio/syt001.

36. Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, et al. Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. Current Biology. 2017;27(7):1019–1025. doi:10.1016/j.cub.2017.03.027.

37. Horn JW, Xi Z, Riina R, Peirson JA, Yang Y, Dorsey BL, et al. Evolutionary bursts inEuphorbia(Euphorbiaceae) are linked with photosynthetic pathway. Evolution. 2014;68(12):3485–3504. doi:10.1111/evo.12534.

38. Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, et al. Why Do Phylogenomic Data Sets Yield Conflicting Trees? Data Type Influences the Avian Tree of Life more than Taxon Sampling. Systematic Biology. 2017;66(5):857–879. doi:10.1093/sysbio/syx041.

39. Murray EA, Carmichael AE, Heraty JM. Ancient host shifts followed by host conservatism in a group of ant parasitoids. Proceedings of the Royal Society B: Biological Sciences. 2013;280(1759):20130495. doi:10.1098/rspb.2013.0495.