StarBeast3: Adaptive Parallelised Bayesian Inference of the Multispecies Coalescent

Jordan Douglas[1,*], Cinthy L. Jiménez-Silva[2], and Remco Bouckaert[1]

[1] *School of Computer Science, University of Auckland, Auckland, 1010, New Zealand*
[2] *School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand*
*jordan.douglas@auckland.ac.nz

Abstract

As genomic sequence data becomes increasingly available, inferring the phylogeny of the species as that of concatenated genomic data can be enticing. However, this approach makes for a biased estimator of branch lengths and substitution rates and an inconsistent estimator of tree topology. Bayesian multispecies coalescent methods address these issues. This is achieved by embedding a set of gene trees within a species tree and jointly inferring both under a Bayesian framework. However, this approach comes at the cost of increased computational demand. Here, we introduce StarBeast3 – a software package for efficient Bayesian inference of the multispecies coalescent model via Markov chain Monte Carlo. We gain efficiency by introducing cutting-edge proposal kernels and adaptive operators, and StarBeast3 is particularly efficient when a relaxed clock model is applied. Furthermore, gene tree inference is parallelised, allowing the software to scale with the size of the problem. We validated our software and benchmarked its performance using three real and two synthetic datasets. Our results indicate that StarBeast3 is up to one-and-a-half orders of magnitude faster than StarBeast2, and therefore more than two orders faster than *BEAST, depending on the dataset and on the parameter, and is suitable for multispecies coalescent inference on large datasets (100+ genes). StarBeast3 is open-source and is easy to set up with a friendly graphical user interface.

1 Introduction

Existing methods for testing macro-evolutionary and macro-ecological questions have not kept pace with the explosion of next generation sequence data now available (Blom et al., 2016b; Bragg et al., 2017; Stenson et al., 2017). Despite burgeoning databases of within- and between-species genomic diversity (Blom et al., 2016b; Bragg et al., 2017; Stenson et al., 2017), it is still common practice to ignore the gene tree discordance that underlies any species phylogeny inferred from multi-locus sequences and instead infer species ancestry based on concatenated sequence data taken to represent all underlying gene histories (Degnan and Rosenberg, 2009; Heled and Drummond, 2010; Ogilvie et al., 2017; Jones, 2017; Rannala and Yang, 2017). While this approach can perform well when branches are long and incomplete lineage sorting (ILS) is absent, these conditions are rarely met.

Species trees inferred from concatenated sequences are often topologically incorrect (Degnan and Rosenberg, 2009; Heled and Drummond, 2010; Ogilvie et al., 2017), provide biased estimates for branch lengths and substitution rates (Ogilvie et al., 2016; Mendes and Hahn, 2016), and underestimate uncertainty in tree topology, resulting in an unjustified degree of confidence in the wrong tree (Heled and Drummond, 2010; Ogilvie et al., 2017). Such biases are exacerbated by subsampling of incongruent genes (Edwards et al., 2016; Mendes and Hahn, 2016), and hold even for deep splits in the tree (Oliver, 2013). These are crucial concerns in themselves and, more generally, can lead to biased estimates and erroneous inferences about fundamental evolutionary and ecological processes that require accurate phylogenetic trees, such as rates of speciation and extinction (Rowe et al., 2011; Cadena et al., 2011; Pepper et al., 2013), rates of substitution in DNA sequences (Bouckaert et al., 2013) and morphological characters (Pepper et al., 2013), species ancestry and ancestral age estimation (Mitchell et al., 2014), geographical history and origins (Lemey et al., 2009; Bouckaert, 2016), and species delimitation (Grummer et al., 2013; Yang and Rannala, 2014; Leaché et al., 2014; Yang and Rannala, 2010).

The multispecies coalescent (MSC; Maddison (1997); Edwards (2009); Liu et al. (2009)) is an approach designed to minimise these potential biases by modelling macro-evolution as a distribution of gene trees embedded within a species tree (Degnan and Rosenberg, 2009; Heled and Drummond, 2010; Ogilvie et al., 2017; Jones, 2017; Rannala and Yang, 2017). In doing so, the MSC provides a more biologically realistic framework for phylogenetic inference that captures the process of ILS underlying most multi-locus phylogenies. Furthermore, by explicitly modelling both species and gene trees, the MSC can address questions that cannot be addressed under a concatenation approach – such as automatic species delimitation (Fujita et al., 2012), with important implications for biodiversity assessment and conservation (Bickford et al., 2007).

A number of software packages have implemented the MSC in various ways (see review by Liu et al. (2015)). Our work at the Centre for Computational Evolution at the University of Auckland has led the development of *BEAST (STARBeast; Heled and Drummond (2010)) and StarBeast2 (Ogilvie et al., 2017) – full Bayesian MSC frameworks for species tree estimation from multilocus sequence data – and UglyTrees for visualising these models (Douglas, 2020). By explicitly modelling the MSC and avoiding the biases associated with concatenation methods (Heled and Drummond, 2010; Ogilvie et al., 2017, 2016), an analysis using either of these software packages can significantly change the conclusions drawn from data.

However, despite some advances in computational efficiency of the full Bayesian MSC (Ogilvie et al., 2017; Jones, 2017; Rannala and Yang, 2017), these complex models remain computationally intractable for large next generation sequence datasets of 100's of sequenced loci across hundreds of individuals (i.e., 10^4–10^6 samples×loci). As a result, existing applications of the approach have tended to consider smaller datasets (Kang et al., 2014; Blom et al., 2016a) or to ignore much of the available data (Blom et al., 2016b; Bragg et al., 2017; Stenson et al., 2017), which reduces accuracy and increases uncertainty in species tree estimates (Ogilvie et al., 2017; Song et al., 2012). One approach to this problem has been the development of much simpler summary coalescent methods which utilise distributions of estimated gene tree topologies as input to rapidly process large datasets (Liu et al., 2015). These include the rooted triplet method MP-EST (Liu et al., 2010) and the quartet method ASTRAL (Mirarab et al., 2014). However, summary coalescent methods are sensitive to gene tree errors (Mirarab and Warnow, 2015; Xi et al., 2015)
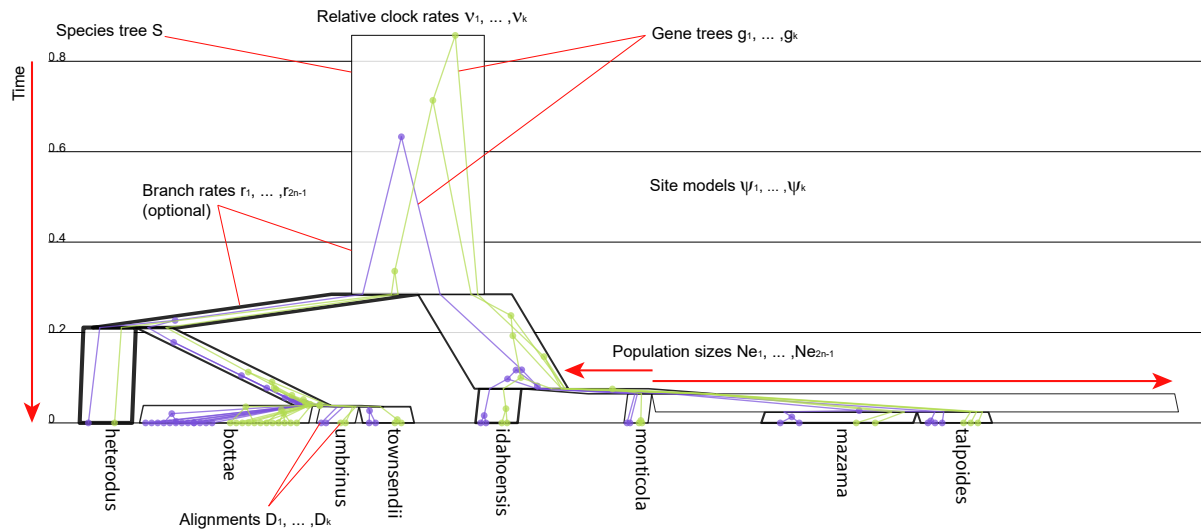
Fig. S1: A species tree S with $n = 8$ species with $k = 2$ gene trees embedded. In this depiction, node heights (time) are shown on the y-axis and effective population sizes on the x-axis. Species tree branch rates are indicated by line thickness. Tree was built from a Gopher dataset (Belfiore et al., 2008) and visualised using UglyTrees (Douglas, 2020).

and produce trees in coalescent units, and thus time and population size estimates used by downstream analyses are confounded.

Here, we aim to perform Bayesian inference on large datasets using the Markov chain Monte Carlo (MCMC) algorithm as our workhorse. As illustrated in Fig. S1, the number of parameters involved is quite large, as is the accompanying state space. We develop a set of new MCMC proposals to explore state space in a much more efficient way than previous implementations and demonstrate we can handle datasets several times faster than *BEAST and StarBeast2. The resulting software package StarBeast3 is available as an open-source BEAST 2 package (Bouckaert et al., 2019).

2 Methods

2.1 The Multispecies Coalescent

Our objective is to develop efficient methods in a Bayesian framework for analysing models where there is a phylogeny, S, such as a species or language tree, that forms a constraint on a set of k trees $\mathbf{G} = \{g_1, \ldots, g_k\}$, such as gene trees. Each taxon within $\mathbf{G}$ is assigned to a single taxon within S, from some fixed individual-to-species mapping function (Fig.S1). Species tree $S = (T_S, t_S)$ consists of a topology T_S and divergence times t_S, as does the set of gene trees $\mathbf{G} = (T_{\mathbf{G}}, t_{\mathbf{G}})$.

All trees are assumed to be binary rooted time trees, with fixed taxon node heights (typically extant; height 0). Each gene tree g_i consists of $2n_{g_i} - 1$ nodes and $2n_{g_i} - 2$ branches for taxon count n_{g_i}, while S consists of $2n_S - 1$ nodes and $2n_S - 1$ branches, including a root branch, for species count n_S. Gene tree taxa are associated with data $D = \{D_1, \ldots, D_k\}$, e.g., nucleotide sequences or cognate data. Let θ be a set of model parameters, for instance those related to the speciation or nucleotide substitution processes. Consider the posterior density function $p(S, \mathbf{G}, \theta | D)$:

$$p(S, \mathbf{G}, \theta | D) = \frac{1}{Z} \overbrace{p(S|\theta)}^{\text{species tree prior}} \times \overbrace{\prod_{i=1}^{k} p(g_i | S, \theta)}^{\text{gene tree prior}} \times$$
$$\underbrace{}_{\text{normalisation constant}}$$

$$\overbrace{p(\theta)}^{\text{parameter hyper prior}} \times \overbrace{\prod_{i=1}^{k} p(D_i | g_i, \theta)}^{\text{gene tree likelihoods}}. \tag{1}$$

The MSC model is therefore hierarchical. S can follow a range of tree prior distributions $p(S|\theta)$, such as the Yule (Yule, 1925) or birth-death models (Nee et al., 1994). Whereas, each gene tree g_i is assumed to follow the multispecies coalescent process (Degnan and Rosenberg, 2009; Heled and Drummond, 2010; Ogilvie et al., 2017; Jones, 2017; Rannala and Yang, 2017), under which each species tree branch is associated with an independently and identically distributed (effective) population size $\mathbf{N_e}$ that governs the coalescent process of $\mathbf{G}$, where $|\mathbf{N_e}| = 2n_S - 1$. Gene trees are thus assumed to be contained within S (Fig. S1).

Site evolution is assumed to follow a continuous-time Markov process (Felsenstein, 1981) under some substitution model and clock model:

$$p(D_i | g_i, \theta) = p(D_i | g_i, \psi_i, \mathbf{r}, \mu). \tag{2}$$

ψ_i can adopt a range of molecular substitution models, such as the HKY nucleotide evolution model (Hasegawa et al., 1985) or the WAG amino acid evolution model (Whelan and Goldman, 2001). Tree g_i has relative molecular substitution rate $\nu_i \in \psi_i$. Each branch in S is associated with a substitution rate $\mathbf{r}$ which governs the rate of site evolution of $\mathbf{G}$ along the respective branch, where $|\mathbf{r}| = 2n_S - 1$ (Fig. S1). Branch rates $\mathbf{r}$ are assumed to be independently and identically distributed under a log-normal distribution with standard deviation σ (i.e., the multispecies coalescent relaxed clock model (Ogilvie et al., 2017; Drummond et al., 2006)). Lastly, the clock rate μ can be estimated when accompanied by time-calibration data, such as ancient fossil records (Heled and Drummond, 2012; Sauquet et al., 2011; Ballesteros and Sharma, 2019), or left fixed when no such data is available. Overall, the total substitution rate of any given branch in g_i is the product of ν_i, μ, and a subset of the elements in $\mathbf{r}$ (weighted by their coverage of the gene tree branch; Ogilvie et al. (2017)).

In this article, we develop tools that allow the MSC to be applied to large datasets using complex models of evolution. Although we focus on MSC models, we anticipate that in the future other models of the form expressed in Eq. (1) will be developed, e.g., models that allow some lateral gene transfer and therefore

allow some gene tree branches to cross species boundaries in the species tree. We design a number of MCMC operators which generate proposals that explore the state space more efficiently – using a Gibbs sampler for population sizes, a combination of Bactrian (Yang and Rodríguez, 2013; Thawornwattana et al., 2018) and adaptable variance multivariate normal (Baele et al., 2017) proposal kernels, a parallel operator for sampling gene trees and substitution model parameters, and an MCMC operator which selects other operators based on their exploration efficiency (Douglas et al., 2021b). Moreover, in the special case of the multispecies relaxed clock model (Ogilvie et al., 2017), we introduce methods for operating on the species tree, the gene trees, and the clock model simultaneously (Zhang and Drummond, 2020; Douglas et al., 2021b).

2.2 Effective Population Size Gibbs Operator

The StarBeast2 (Ogilvie et al., 2017) and DISSECT (Jones et al., 2015) packages have the capability of integrating out effective population sizes $\mathbf{N_e}$ when using an inverse gamma distributed prior on $\mathbf{N_e}$, based on a technique introduced by Liu et al. (2008) and detailed out by Jones (2017). This approach greatly reduces the state space. However, consequently the posterior Eq. 1 can no longer be broken down in a product over components over individual gene trees:

$$\prod_{i=1}^{k} p(g_i|S,\theta) = \int_{\mathbf{N_e}} \prod_{i=1}^{k} p(g_i|S,\mathbf{N_e}) \, d\mathbf{N_e} \tag{3}$$

Thus, the technique is not suitable for gene-tree operator parallelisation, and therefore we estimate $\mathbf{N_e}$ instead.

Suppose that N_{e_b}, for species-tree branch b, follows an inverse gamma prior distribution Inv-$\Gamma(\alpha_N, \mu_N)$, where the shape α_N is fixed at 2 and therefore the scale μ_N is the expected value (because $\mathbb{E}(N_{e_b}) = \frac{\beta}{\alpha-1}$). Following the results by Jones (2017), the posterior of N_{e_b} follows an inverse gamma Inv-$\Gamma(\alpha_N', \mu_N')$, such that $\alpha_N' = \alpha_N + a$ and $\mu_N' = \mu_N + c$ where a is the total number of coalescent events of all gene trees in branch b and $c = \sum_j \frac{1}{p_j} \sum_i c_{jbi} \binom{n_{jb}-i}{2}$. Here p_j is the ploidy of gene g_j, c_{jbi} the size of the ith coalescent interval for gene g_j in branch b, and n_{jb} the number of lineages of gene tree g_j at the tip-side of branch b (so that $n_{jb} - i$ is the number of lineages at the start of the ith coalescent interval for g_j).

Instead of integrating out $\mathbf{N_e}$, our `GibbsPopulation` operator samples from the posterior. All $2n_S - 1$ elements in $\mathbf{N_e}$ are proposed simultaneously. As demonstrated later, this turns out to be more efficient than standard $\mathbf{N_e}$ random walk operators, with the added advantage of sampling effective population sizes – which may be a parameter of interest – as well as the ability to parallelise gene tree proposals. This technique is readily applicable for periodically sampling and logging $\mathbf{N_e}$ to implementations that do integrate this term out.

2.3 Bactrian Operators for Trees

The step size of a proposal kernel 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 far that the proposal is rejected too often (Roberts et al., 1997). The Bactrian distribution (Yang and Rodríguez, 2013; Thawornwattana et al., 2018) has minimal probability mass around the center, and a higher concentration flanking the center, akin to the humps of a Bactrian camel (Fig. S2; left). This distribution is a preferred alternative to standard uniform- or normal-distributed random walk kernels, as it places minimal probability on step sizes that are too large or too small, and has successfully improved phylogenetic inference in previous studies (Yang and Rodríguez, 2013; Zhang and Drummond, 2020; Douglas et al., 2021b).

In this article, we apply Bactrian proposals to trees. The standard set of tree node height proposals in BEAST 2 consists of a `Scale` operator which embarks all nodes in the tree on a random walk (in log-space), a `RootScale` operator which does so for only the root of a tree, an `UpDown` operator which
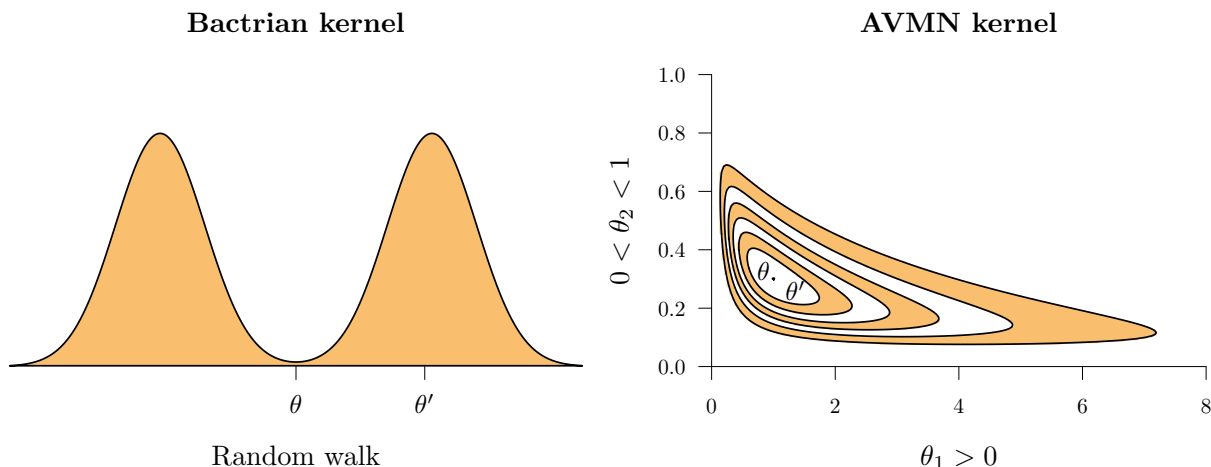
Fig. S2: Depiction of random walks $\theta \to \theta'$ under varying proposal kernels. Left: The random walk occurs from the origin between the two modes, where the vertical axis shows the probability density function of the kernel $p(\theta'|\theta)$ (Yang and Rodríguez, 2013). Right: A two dimensional random walk on inversely correlated parameters $\theta = (\theta_1, \theta_2)$ with different domains (Baele et al., 2017). Contours describe the joint probability density function $p(\theta_1, \theta_2)$ under a transformed multivariate normal distribution learned during MCMC.

changes species/gene node heights and various continuous parameters simultaneously (Drummond et al., 2002) a `SubtreeSlide` operator which slides a node up or down branches (Hohna et al., 2008), and constant distance operators when a relaxed clock model is applied (Zhang and Drummond, 2020). Each operator would normally draw a random variable from a uniform distribution, but here we instead use a Bactrian distribution and apply appropriate transformations. We also introduce the `Interval` operator, which transforms parameters with lower- and upper-bounds (such as tree node heights) by applying a Bactrian random walk in their real-space transformations.

2.4 AVMN Operator

An adaptive variance multivariate normal (AVMN) operator (Baele et al., 2017) provides proposals for a set of real-space parameters by learning the posterior throughout the run of the MCMC algorithm and approximating it as a multivariate normal distribution to capture correlations between parameters (Fig. S2; right). The space spanned by such a set continuous parameters may need to be transformed to satisfy the real-space assumption, by applying a log-transformation to parameters with positive domains (such as substitution rates), or a log-constrained sum transformation to multivariate parameters with unit sums (such as nucleotide frequencies), for instance. AVMN has been demonstrated to be more efficient in estimating phylogenetic parameters than standard random walk or scale operators (Baele et al., 2017; Bouckaert, 2020; Douglas et al., 2021b).

Consider a single gene tree g_i and its substitution model ψ_i, consisting of substitution rates and nucleotide frequencies for instance. Performing a single proposal for any single parameter would require a full recalculation of the tree likelihood $p(D_i|g_i, \psi_i, \mathbf{r}, \mu)$ (see peeling algorithm by Felsenstein (1981)). Therefore, proposing all site model parameters ψ_i simultaneously can reduce the number of likelihood calculations required and thus lower the computational runtime.

2.5 Parallel Gene Tree Operator

During MCMC, operators are typically sampled proportionally to fixed weights (or proposal probabilities), to ensure the chain is ergodic. Here, we present an alternative method, where a single gene tree g_i and

its substitution model ψ_i is selected, and N_p operators are sequentially sampled and applied to g_i and ψ_i, before returning to the full parameter space. This is equivalent to running a small MCMC chain of N_p steps – applying only gene tree and substitution model operators on g_i and ψ_i – and then accepting the resulting g_i' and ψ_i' afterwards with probability 1, as if it were a single Gibbs sampling operation (Geman and Geman, 1984).

Observe that because only g_i and its associated parameters change, part of Eq (1) can be rewritten as:

$$\overbrace{\prod_{i=1}^{k} p(D_i|g_i, \psi_i, \mathbf{r})}^{\text{gene tree likelihoods}} \times \overbrace{\prod_{i=1}^{k} p(g_i|S, \mathbf{N_e})}^{\text{gene tree priors}} = \overbrace{\prod_{i=1}^{k} p(D_i|g_i, \psi_i, \mathbf{r}) \times p(g_i|S, \mathbf{N_e})}^{\text{gene tree posteriors}} . \tag{4}$$

Thus, the posterior distribution can be decomposed into the product of contributions of individual gene trees and their substitution models. Assuming that substitution model parameters ψ_i are distinct for each gene tree g_i, an N_p-step MCMC chain could be run for each of g_i and g_j for $(i \neq j)$ in parallel, and the resulting g_i' and g_j' each accepted with probability 1, as if two Gibbs operators were sequentially applied. Because the posterior density for g_i is proportional to $p(D_i|g_i, \theta)\ p(g_i|S, \theta)$ and that of g_j proportional to $p(D_j|g_j, \theta)\ p(g_j|S, \theta)$, then provided that any shared parameters (such as $\mathbf{r}$, S, and $\mathbf{N_e}$) are not being operated on, these two N_p-step MCMC chains can run in parallel.

Where there are N_t threads available, the k gene trees are split into N_t groups (assuming $k \geq N_t$). The N_t sets of N_p-step MCMC chains are run in parallel and the resulting gene trees $\mathbf{g}$ are accepted into the main MCMC chain. Here, we introduce a parallel operator `ParallelGeneTreeOperator(G,■)`. This operator partitions gene trees into N_t threads and operates on their topologies, node heights, and substitution models. Tree node height proposals employ the Bactrian kernel where applicable (Fig. S2), and substitution model proposals invoke the AVMN kernel (Fig. S2). The chain length N_p of each thread is learned during MCMC (Fig. S3).

Since each small MCMC chain for a thread can be considered a single Gibbs proposal, for N_t threads in principle N_t steps should be added to the main chain. If the operator is selected just before logging a state, in principle some threads may need to be disregarded before logging in order to maintain exactly equal intervals in the trace log. Due to the low frequency at which the operator is selected, and the logging intervals being orders of magnitude larger than the number of threads, this does not appear to be a problem in practice.

2.6 Species Tree Relaxed Clock Model Operators

The constant distance operator family exploits the negative correlations between divergence times and branch substitution rates by proposing both terms simultaneously (Zhang and Drummond, 2020). This technique has yielded a parameter convergence rate one to two orders of magnitude faster, particularly for large datasets that come with peaked posterior distributions (Douglas et al., 2021b). Under the multispecies coalescent relaxed clock model used by StarBeast2, the branch rate of gene tree branch b is the length-weighted branch rate $\mathbf{r}$ of all species tree branches that contain b (Ogilvie et al., 2017). Moreover, effective population sizes $\mathbf{N_e}$ are positively correlated with divergence times, so this correlation could also be readily exploited.

Extending the work by Zhang and Drummond (2020), we introduce the `ConstantDistanceMSC` operator. This operator proposes a node height t_X for species tree internal node X, the three branch rates $\mathbf{r}$ and population sizes $\mathbf{N_e}$ incident to X, and heights for all gene tree non-leaf nodes that are contained within these three incident branches (Fig. S4). t_X is embarked on a Bactrian random walk (Yang and Rodríguez, 2013) to give t_X', then $\mathbf{r}$ and the node heights in $\mathbf{G}$ are proposed such that all genetic distances are conserved following the change in t_X, and $\mathbf{N_e}$ is proposed such that the positive correlation between itself and the branch lengths incident to X is respected (see Algorithm S1).

Previously, we introduced the narrow exchange rate (`NER`) operator (Douglas et al., 2021b). This operator combined the simple `NarrowExchange` operator (i.e., a proposal which swaps a subtree with its uncle
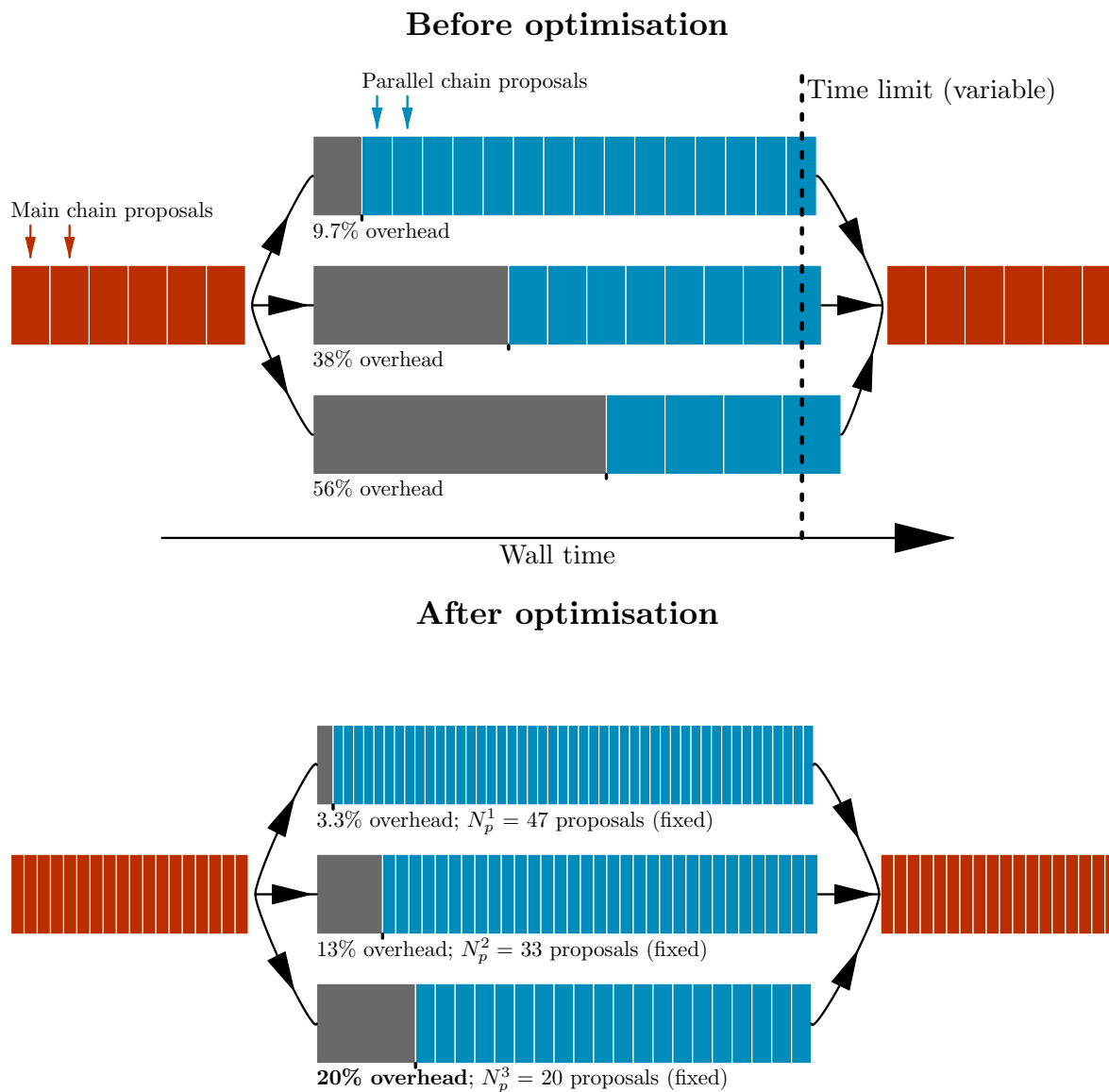
Before optimisation

Parallel chain proposals

Time limit (variable)

Main chain proposals

9.7% overhead

38% overhead

56% overhead

Wall time

After optimisation

3.3% overhead; $N_p^1 = 47$ proposals (fixed)

13% overhead; $N_p^2 = 33$ proposals (fixed)

20% overhead; $N_p^3 = 20$ proposals (fixed)

Fig. S3: Optimisation of gene tree parallel operator chain lengths. Top: The time limit of each parallel MCMC chain is randomised on each call so that the overhead (intercept) and time-per-proposal (slope) can be learned as a linear regression model. Bottom: The linear regression model is applied, and parallel MCMC chain lengths are set such that the slowest thread attains the user-specified target overhead (i.e., the bottom thread has attained 20% overhead in the example above).
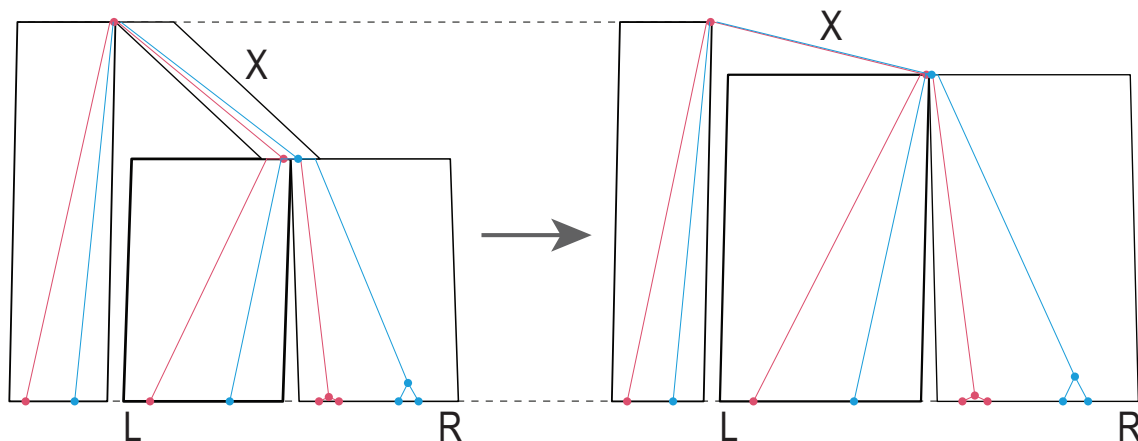
Fig. S4: An example of an `ConstantDistanceMSC` proposal, with species nodes X, and its two children L and R indicated. The height of X is increased, and **r** (indicated by species node line thickness) and $\mathbf{N_e}$ (indicated by species node width) are increased or decreased accordingly. Figure generated by UglyTrees (Douglas, 2020).

subtree; Drummond et al. (2002)) with the `ConstantDistance` operator (Zhang and Drummond, 2020), by applying a small topological change to the tree and then recomputing branch substitution rates such that evolutionary distances are preserved. We demonstrated that this operator assisted the traversal of tree topology space on longer alignments compared with shorter ones.

Here, we combine this work with the `CoordinatedExchange` operator implemented by Ogilvie et al. (2017) – based on work by Jones (2017) and Rannala and Yang (2017) – and introduce the coordinated narrow exchange rate (`CNER`) operator. This operator exchanges a species tree node with its uncle node, adjusts gene tree topologies **g** to preserve compatibility with S, and proposes three nearby branch rates **r** to preserve genetic distances (Algorithm S2).

2.7 Adaptive Operator Weighing

Previously, we developed the `AdaptableOperatorSampler(x)` operator (Douglas et al., 2021b). This operator learns the weights (or proposal probabilities) behind a set of sub-operators during MCMC, by rewarding operators which bring about large changes to parameter x in short computational runtime, with respect to some distance function: Euclidean distance when x is real, and RNNI distance (Collienne and Gavryushkin, 2021) when x is tree topology. This approach can account for the scenario when an operator's performance is conditional on the dataset. When a dataset contains very little signal with respect to a certain parameter x and its prior distribution, then resampling that parameter from its prior distribution using the `SampleFromPrior(x)` operator may be more efficient than embarking x on a random walk, for instance (Douglas et al., 2021b). In contrast, datasets with more signal are likely to prefer smarter operators which account for correlations in the posterior distribution, such as the constant distance or `NER` operators (Zhang and Drummond, 2020; Douglas et al., 2021b).

Here, we have applied the `AdaptableOperatorSampler` to seven areas of parameter space: the species and gene tree node heights (t_s and $t_\mathbf{G}$), the relaxed clock model rates **r** and standard deviation σ, the mean effective population size μ_N, the species tree birth rate λ (assuming a Yule speciation model (Yule, 1925)), and the species tree topology T_S. These operator schemes are explicated in **Tables S1** and **S2**.

Operator	Weight	Reference
Species tree		
`NodeReheight`$(T_S, t_S, T_\mathbf{G}, t_\mathbf{G})$	30	Ogilvie et al. (2017)
`CoordinatedUniform`$(t_S, t_\mathbf{G})$	30	Ogilvie et al. (2017); Jones (2017)
`CoordinatedExponential`$(t_S, t_\mathbf{G})$	15	Ogilvie et al. (2017); Jones (2017)
`SubtreeSlide`$(T_S, t_S)^\dagger$	15	Hohna et al. (2008)
`WilsonBalding`(T_S)	15	Drummond et al. (2002)
`WideExchange`(T_S)	15	Drummond et al. (2002)
`AdaptableOperatorSampler`(T_S)	15	
`NarrowExchange`(T_S)		Drummond et al. (2002)
`CoordinatedExchange`$(T_S, T_\mathbf{G})$		Ogilvie et al. (2017)
`NER`$(T_S, \mathbf{r})$		Douglas et al. (2021b)
`CNER`$(T_S, \mathbf{r}, T_\mathbf{G})$		*Species Tree Relaxed Clock Model Operators*
`Uniform`$(t_S)^\dagger$	3	
`RootScale`$(t_S)^\dagger$	3	
`Interval`$(t_S)^\dagger$	3	*Bactrian Operators for Trees*
`AdaptableOperatorSampler`(t_S)	100	
`Uniform`$(t_S)^\dagger$		
`TreeScale`$(t_S)^\dagger$		
`Interval`$(t_S)^\dagger$		*Bactrian Operators for Trees*
`ConstantDistanceMSC`$(t_S, t_\mathbf{G}, \mathbf{r}, \mathbf{N_e})^\dagger$		*Species Tree Relaxed Clock Model Operators*
`CoordinatedUniform`$(t_S, t_\mathbf{G})$		Ogilvie et al. (2017); Jones (2017)
`CoordinatedExponential`$(t_S, t_\mathbf{G})$		Ogilvie et al. (2017); Jones (2017)
`UpDown`$([t_S, t_\mathbf{G}, \mathbf{N_e}, \mu_N], [\degree, \lambda, \mu])^\dagger$		Drummond et al. (2002)
Gene trees / site models		
`ParallelMCMC`$(T_\mathbf{G}, t_\mathbf{G}, \psi)$	3.42	**Table S2**
Tree hyperparameters		
`GibbsPopulation`$(\mathbf{N_e})$	50	*Effective Population Size Gibbs Operator*
`AdaptableOperatorSampler`(μ_N)	5	
`Scale`$(\mu_N)^\dagger$		
`UpDown`$([t_S, t_\mathbf{G}, \mathbf{N_e}, \mu_N], [\degree, \lambda, \mu])^\dagger$		Bouckaert et al. (2019)
`SampleFromPrior`(μ_N)		Douglas et al. (2021b)
`AdaptableOperatorSampler`(λ)	5	
`Scale`$(\lambda)^\dagger$		
`UpDown`$([t_S, t_\mathbf{G}, \mathbf{N_e}, \mu_N], [\degree, \lambda, \mu])^\dagger$		Bouckaert et al. (2019)
`SampleFromPrior`(λ)		Douglas et al. (2021b)
Relaxed clock model		
`AdaptableOperatorSampler`$(\mathbf{r})$	30	
`Scale`$(\mathbf{r})^\dagger$		
`ConstantDistanceMSC`$(t_S, t_\mathbf{G}, \mathbf{r}, \mathbf{N_e})^\dagger$		*Species Tree Relaxed Clock Model Operators*
`SampleFromPrior`$(\mathbf{r})$		Douglas et al. (2021b)
`AdaptableOperatorSampler`(σ)	5	
`Scale`$(\sigma)^\dagger$		
`SampleFromPrior`(σ)		Douglas et al. (2021b)

Table S1: StarBeast3 operator scheme, assuming a Yule tree prior on the species tree with birth rate λ (Yule, 1925). The `ParallelMCMC` operator weight was set such that it consists of 1% of the all proposals. Further operator details can be found in Drummond and Bouckaert (2015). †Bactrian kernel applied to random walk (Yang and Rodríguez, 2013).

Operator	Weight	Reference
`ParallelMCMC(`$T_\mathbf{G}, t_\mathbf{G}, \psi$`)`		*Parallel Gene Tree Operator*
Gene trees		
$\forall_{i \in 1,\dots,k}$ `WilsonBalding(`T_{g_i}`)`	15	Drummond et al. (2002)
$\forall_{i \in 1,\dots,k}$ `WideExchange(`T_{g_i}`)`	15	Drummond et al. (2002)
$\forall_{i \in 1,\dots,k}$ `NarrowExchange(`T_{g_i}`)`	15	Drummond et al. (2002)
$\forall_{i \in 1,\dots,k}$ `SubtreeSlide(`T_{g_i}, t_{g_i}`)`[†]	10	Hohna et al. (2008)
$\forall_{i \in 1,\dots,k}$ `Uniform(`t_{g_i}`)`	30	
$\forall_{i \in 1,\dots,k}$ `RootScale(`t_{g_i}`)`[†]	10	
$\forall_{i \in 1,\dots,k}$ `Interval(`t_{g_i}`)`[†]	10	*Bactrian Operators for Trees*
$\forall_{i \in 1,\dots,k}$ `AdaptableOperatorSampler(`t_{g_i}`)`	100	
$\quad$ `TreeScale(`t_{g_i}`)`[†]		
$\quad$ `Uniform(`t_{g_i}`)`[†]		
$\quad$ `SubtreeSlide(`T_{g_i}, t_{g_i}`)`[†]		Hohna et al. (2008)
$\quad$ `EpochOperator(`t_{g_i}`)`[†]		Bouckaert (2021)
Site models		
$\forall_{i \in 1,\dots,k}$ `AVMN(`ψ_i, t_{g_i}`)`	5	Baele et al. (2017)
$\forall_{i \in 1,\dots,k}$ `Scale(`κ_i`)`[†]	0.5	
$\forall_{i \in 1,\dots,k}$ `Scale(`ν_i`)`[†]	0.5	
$\forall_{i \in 1,\dots,k}$ `DeltaExchange(`f_i`)`	0.5	

Table S2: StarBeast3 parallel operator scheme for gene trees and their associated site models (assumed to be an HKY model with transition-transversion ratio κ and nucleotide frequencies f). Each operator is applicable to a single gene tree g_i or its site model ψ_i. `AVMN(`ψ_i, t_{g_i}`)` generated proposals for the site model and complete set of tree node heights simultaneously. Operator weights are normalised into proposal probabilities within a single MCMC chain called by `ParallelMCMC`. Further operator details can be found in Drummond and Bouckaert (2015). [†]Bactrian kernel applied to random walk (Yang and Rodríguez, 2013).

3 Results

In this section we prove StarBeast3 is correctly implemented through a well-calibrated simulation study. Further, we demonstrate that StarBeast3 is efficient at doing Bayesian inference on large datasets compared with StarBeast2. We did not compare to *BEAST directly, since it does not provide relaxed clock models on species trees, but note that Ogilvie et al. (2017) benchmarked StarBeast2 against *BEAST for strict clocks and found StarBeast2 to be an order faster than *BEAST, so any gain over StarBeast2 will be more so over *BEAST.

3.1 The Implementation Is Correct

In order to validate the correctness of StarBeast3, we performed two well-calibrated simulation studies. These were achieved by simulating nucleotide alignments (of two varying sizes) using parameters directly sampled from the prior distribution, and then recovering the posterior estimates of these parameters by doing Bayesian inference on the simulated alignments using StarBeast3. For each study, the 95%-coverage of each parameter was approximately 95% (meaning that the true parameter estimate was within the 95% highest posterior density interval approximately 95% of the time). Therefore, these experiments provide confidence in StarBeast3's correctness, and are presented in Fig. S5 and Section 4 of *Supplementary Material*.

3.2 Performance Benchmarking

We evaluated the performance of StarBeast3 for its ability to achieve multispecies coalescent parameter convergence in a Bayesian framework, compared with that of StarBeast2. This was measured by computing effective sample sizes (ESS) generated per hour during MCMC across multiple replicates of three real and two simulated datasets (**Table S3**). The ESS of any parameter should be over 200 in order to estimate its posterior distribution (Tracer; Rambaut et al. (2018)). To allow both software packages to perform at their best, effective population sizes were integrated out by StarBeast2, but were estimated by StarBeast3. This section provides a general comparison of StarBeast3 and StarBeast2, however the performances of individual operators can be found in Sections 5 and 6 of *Supplementary Material*.
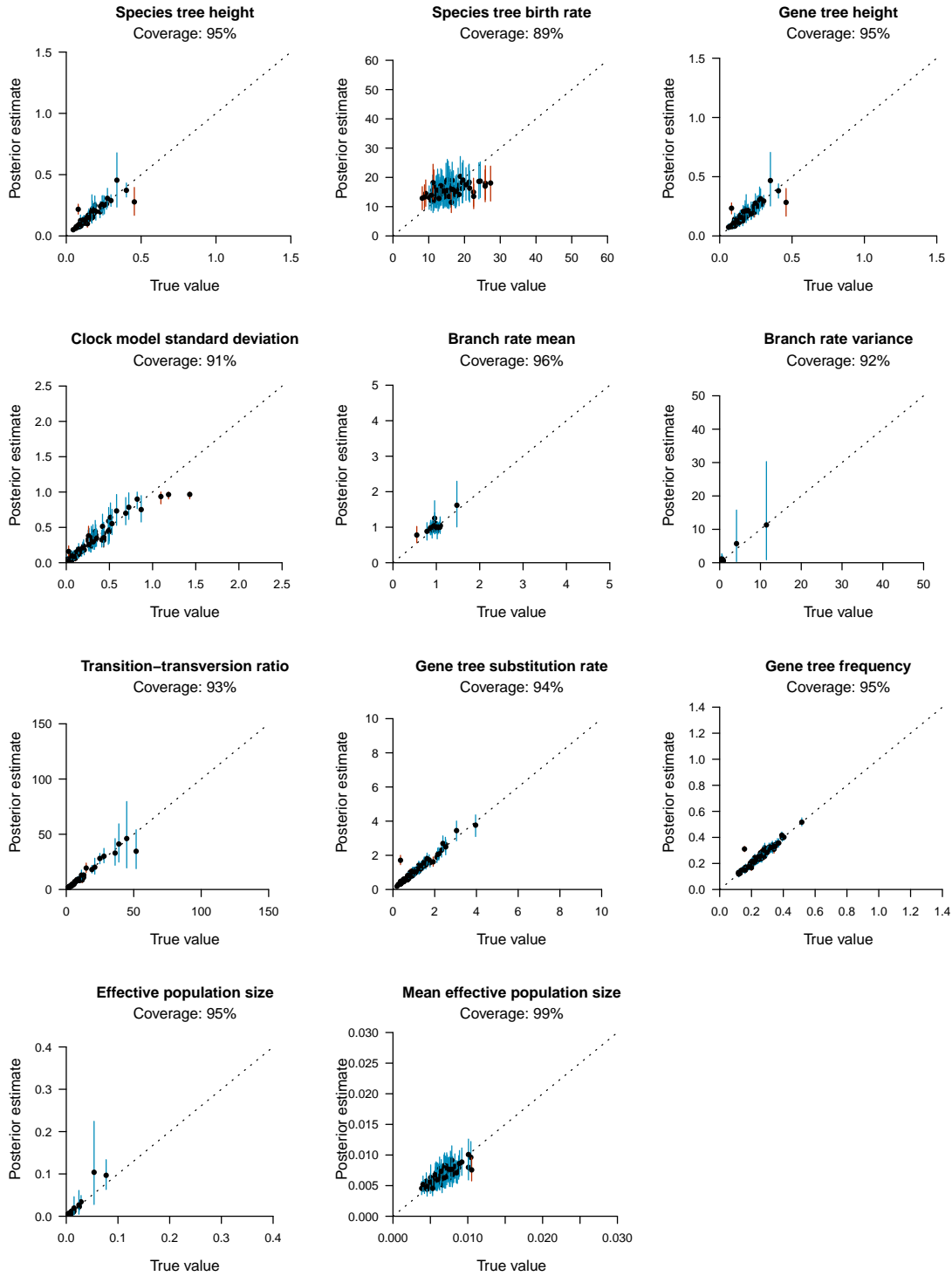
Fig. S5: Well-calibrated simulation study analysing $n_S = 16$ species, $n_G = 48$ taxa, and $k = 50$ genes. 100 simulations were performed to recover the coverage between "true" simulated values and their estimates under the posterior distribution. 95% highest posterior density (HPD) intervals of parameters are represented by vertical lines. Each line represents a single simulation, and is coloured blue when the true value was contained within the 95% interval, or red otherwise. The top of each plot shows the coverage of each parameter (i.e., the number of MCMC simulations for which the "true" parameter value was contained within the 95% HPD).

Dataset	Number of species n_S	Number of taxa $n_{\mathbf{G}}$	Number of gene trees k
Frog (Barrow et al., 2014)	21	88	26
Skink (Bryson Jr et al., 2017)	10	59	50
Spider (Hamilton et al., 2016)	36	83	50
Simulated(12)	4	12	100
Simulated(48)	16	48	100

Table S3: Benchmark datasets. 50 gene trees were subsampled from the Skink and Spider datasets. The Simulated datasets were directly sampled from the model specification used during Bayesian inference (described in Section 3 of *Supplementary Material*).

The ESS/hr was evaluated in five distinct areas of parameter space. First, we considered generic summaries of convergence: the ESS/hr of the posterior density $p(\theta|D)$, the likelihood $p(D|\theta)$, and the prior density $p(\theta)$. Second, species tree S convergence was evaluated in terms of its height h_S, its length l_S, and hyperparameters λ – the Yule model birth rate (Yule, 1925) – and μ_N – the mean effective population size. In the case of StarBeast3, where effective population sizes are estimated, we also measured the mean ESS/hr associated with species tree leaf nodes of $\mathbf{N_e}$. Third, gene tree convergences were evaluated by their heights $h_{\mathbf{G}}$, their lengths $l_{\mathbf{G}}$, and the RNNI distances (Collienne and Gavryushkin, 2021) to their UPGMA D_{UPGMA} (Sokal, 1958) and neighbour-joining trees D_{NJ} (Saitou and Nei, 1987). As there are multiple gene trees, we only considered the mean ESS/hr of each term. Fourth, substitution model convergence (HKY substitution model; Hasegawa et al. (1985)) was measured from the transition-transversion ratio κ, nucleotide frequencies f, and gene tree substitution rates ν, where the ESS/hr of each term was averaged across all k substitution models. Lastly, relaxed clock model convergence was evaluated by considering the mixing of branch rate mean $\mathbb{E}(\mathbf{r})$ and variance $\mathrm{var}(\mathbf{r})$, as well as the relaxed clock standard deviation parameter σ.

These results showed that, depending on the dataset, the "slowest" parameter generally converged considerably faster for StarBeast3 than it did for StarBeast2 (see the **min** term in Fig. S6–S10). On the smallest dataset considered (Frog), StarBeast2 and 3 performed comparably well overall (and no significant difference in **min**). However, StarBeast3 performed better on all of the other datasets, with the "slowest" parameter converging between 3.9 and 23× as fast, and the posterior density $p(\theta|D)$ converging between 2.5 and 25× as fast, usually at a statistically significant level.

Notably, relaxed clock model parameters converged up to 52× as fast under StarBeast3. This was credited to the use of a real-space branch rate parameterisation (as opposed to the discrete branch rate categorisation employed by StarBeast2) as well as constant distance operators, which adjust branch rates and divergence times simultaneously (Zhang and Drummond, 2020; Douglas et al., 2021b). The disparity between StarBeast3 and StarBeast2 increased with the number of genes k, with one-and-a-half orders of magnitude improvement observed for the larger datasets, but only a 2-fold improvement for the $k = 26$ Frog data (Barrow et al., 2014), consistent with previous experiments (Douglas et al., 2021b).

Substitution model parameters ψ generally converged faster for StarBeast2 than they did for StarBeast3. Note, however, that this is by design. The total operator weight assigned to ψ parameters were 50% smaller in StarBeast3, in order to ensure balanced convergence across all areas of parameter space. In all datasets considered, substitution models converged significantly faster than any other area of parameter space, despite receiving relatively little operator weight, and therefore computational resources that were being spent on the substitution model were better off spent in "slower" areas of parameter space, such as gene tree node heights.

The `AdapableOperatorSampler` operators (**Table S1**) confirmed the value in the `NER` and `ConstantDistanceMSC` operators for operating on their respective areas of parameter space. The `ConstantDistanceMSC` operator almost always outperformed other operators at proposing species node heights t_S (Table S4). The exception to this was the Skink dataset, for which the `UpDown` operator was superior at proposing branch lengths, and the Frog dataset, for which `ConstantDistanceMSC`, `CoordinatedExponential`, and `UpDown` were all on a par. In general, very little operator weight was rewarded to the `Uniform`, `Interval`, `TreeScale` `CoordinatedUniform`, and `CoordinatedExponential` operators for their abilities to propose species node heights. Similarly, among `NarrowExchange` variants evaluated by `AdaptableOperatorSampler`(T_S), the

`NER` operator was marginally favoured by all datasets (Table S5). This was due to the operator making larger or more frequent topological changes to the species tree, in faster computational runtime, especially compared with `CoordinatedNarrowExchange` and `CNER`. Overall, this experiment reinforced the value of learning operator weights on a problem-by-problem basis. A full breakdown of the remaining four adaptive operators can be found in Section 6 of *Supplementary Material*.

Dataset	Uniform	Interval	ConstantDistanceMSC	TreeScale	CUnif	CExp	UpDown
Frog	0.06	0.0078	0.34*	9.8e-05	0.04	0.22	0.33
Simulated(12)	1.1e-05	0.00013	0.99*	2.6e-05	0.00043	8e-04	0.0061
Simulated(48)	3.9e-05	0.00049	0.98*	5.8e-06	0.0011	3e-04	0.019
Skink	0.008	0.0087	0.34	4.8e-05	0.013	0.04	0.59*
Spider	0.0019	0.0034	0.84*	1.5e-05	0.0063	0.0025	0.15

Table S4: Learned weights of the sub-operators of `AdaptableOperatorSampler($t_S$)`, averaged across 5 replicates (2 sf). The operator which attained the highest proposal probability is indicated by a *.

Dataset	NE	NER	CNE	CNER
Frog	0.29 (0.29ms)	0.29 (0.29ms)*	0.21 (0.4ms)	0.21 (0.39ms)
Simulated(12)	0.34 (0.091ms)	0.35 (0.094ms)*	0.15 (0.19ms)	0.15 (0.19ms)
Simulated(48)	0.25 (0.79ms)	0.3 (0.64ms)*	0.22 (1ms)	0.23 (0.83ms)
Skink	0.31 (0.3ms)	0.31 (0.3ms)*	0.19 (0.5ms)	0.19 (0.48ms)
Spider	0.23 (1.6ms)	0.3 (1.2ms)*	0.2 (1.8ms)	0.27 (1.3ms)

Table S5: Learned proposal probabilities (and operator runtimes) for the suboperators of `AdaptableOperatorSampler($T_S$)`, averaged across 5 replicates (2sf). Note that the timer starts at the beginning of the proposal and ends when the proposal has accepted or rejected. Notation: NE – narrow exchange; NER – narrow exchange rates; CNE – coordinated narrow exchange; CNER – coordinated narrow exchange rates. The operator which was rewarded the highest proposal probability for each dataset is indicated by a *.

Lastly, we evaluated the effect of threading on StarBeast3, by comparing its performance under 1, 2, 4, 8, and 16 threads allotted to the `ParallelMCMC` gene tree operator. There was a positive-but-modest correlation between the number of threads and the overall rate of convergence among the terms considered, with an overall log-linear slope coefficient of 0.19. This can be interpreted as follows: across the range of threads and datasets considered, doubling the number of threads was associated with an increase in mixing by 14%. Multithreading provided the strongest boost for the Skink and Spider datasets, and made little difference to the simulated dataset (48 taxa).
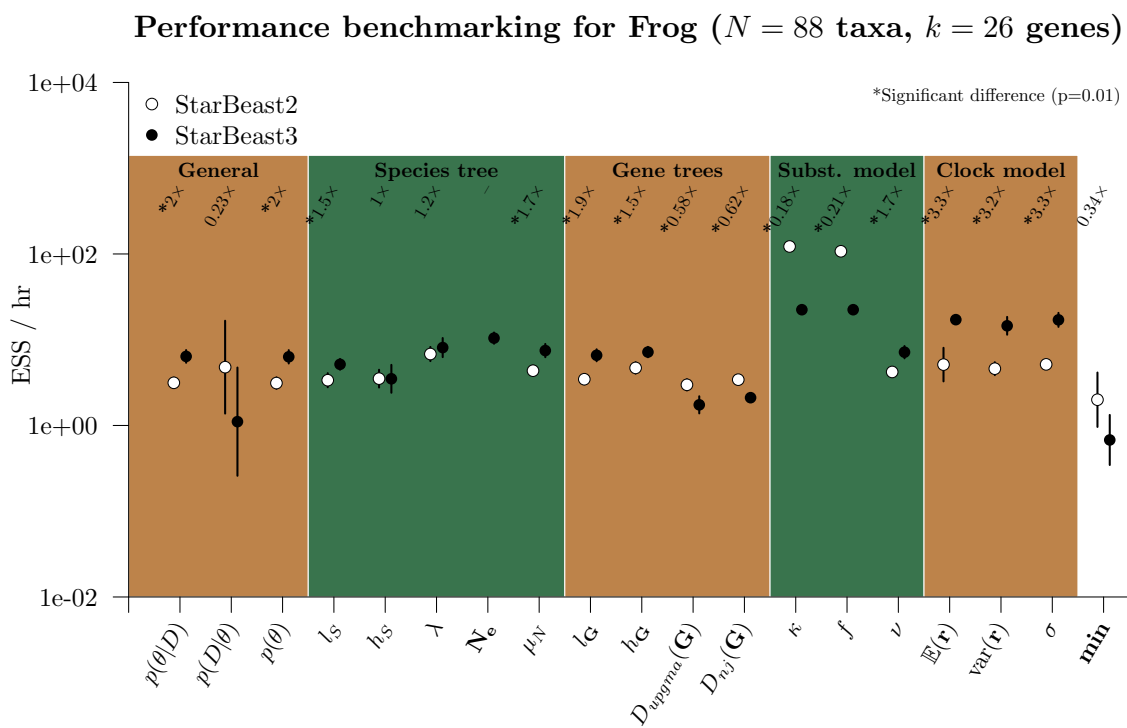
Fig. S6: Performance benchmarking the Frog dataset (Barrow et al., 2014). Each point is the geometric-mean ESS/hr across 5 replicates, for either StarBeast2, or StarBeast3 with 16 threads. The geometric-mean relative performance of StarBeast3, compared with StarBeast2, is indicated above each term, and a * is present if the difference across 5 replicates is significant according to a Student t-test. Note that the y-axis is in log-space.

Performance benchmarking for Skink ($N = 59$, taxa $k = 50$ genes)
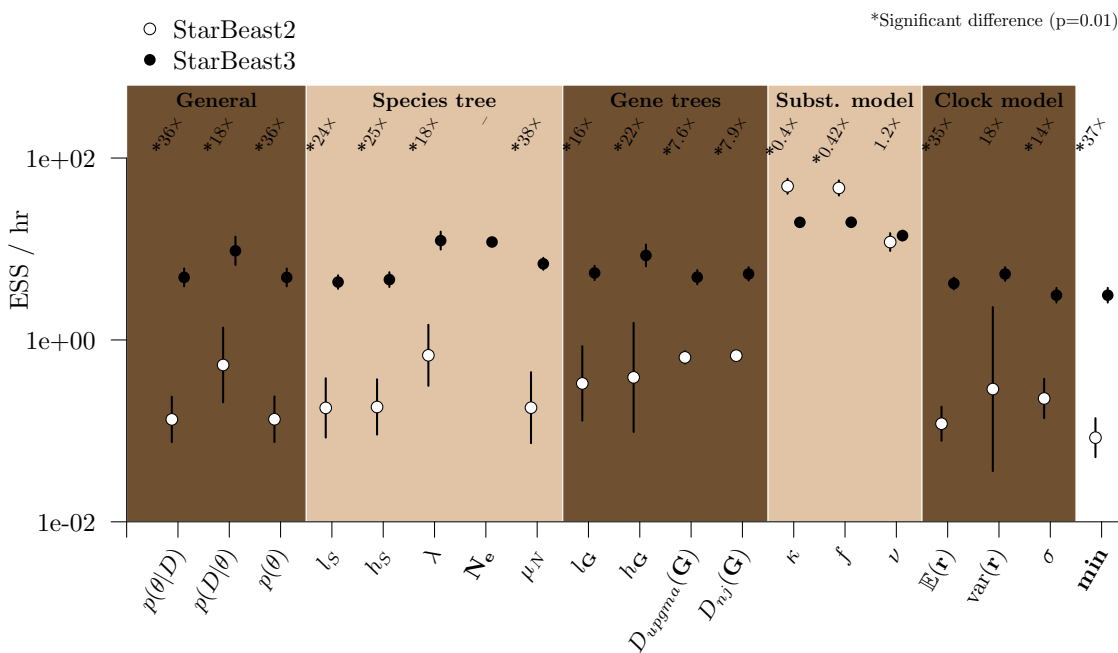
Fig. S7: Performance benchmarking the Skink dataset (Bryson Jr et al., 2017). See Fig. S6 caption for figure notation.

Performance benchmarking for Spider ($N = 83$ **taxa,** $k = 50$ **genes**)
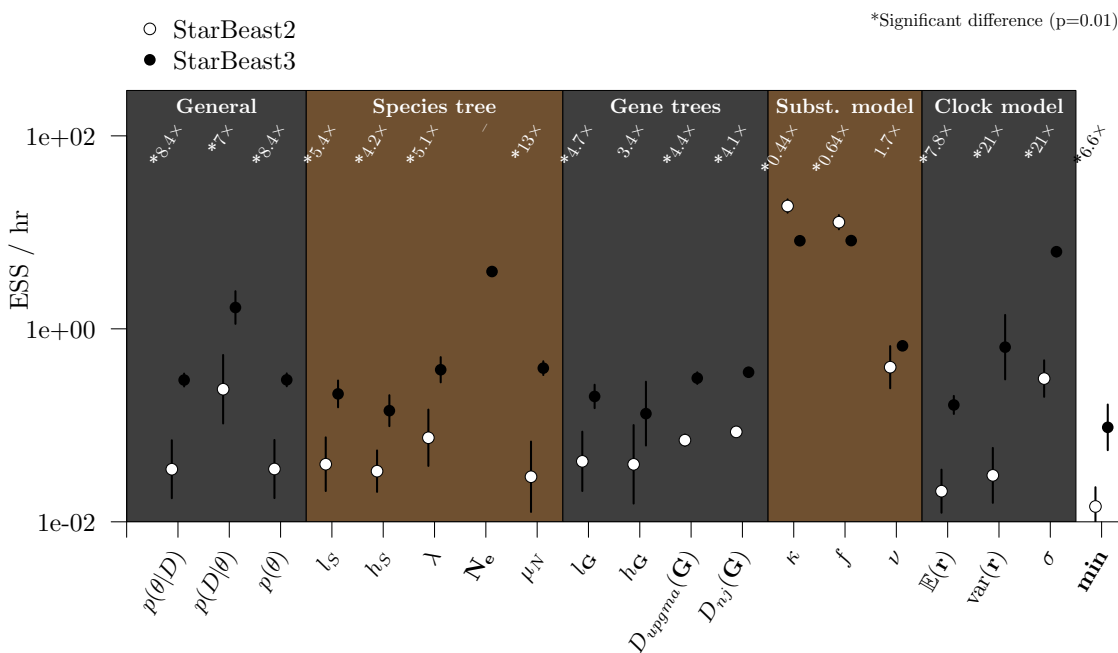
Fig. S8: Performance benchmarking the Spider dataset (Hamilton et al., 2016). See Fig. S6 caption for figure notation.
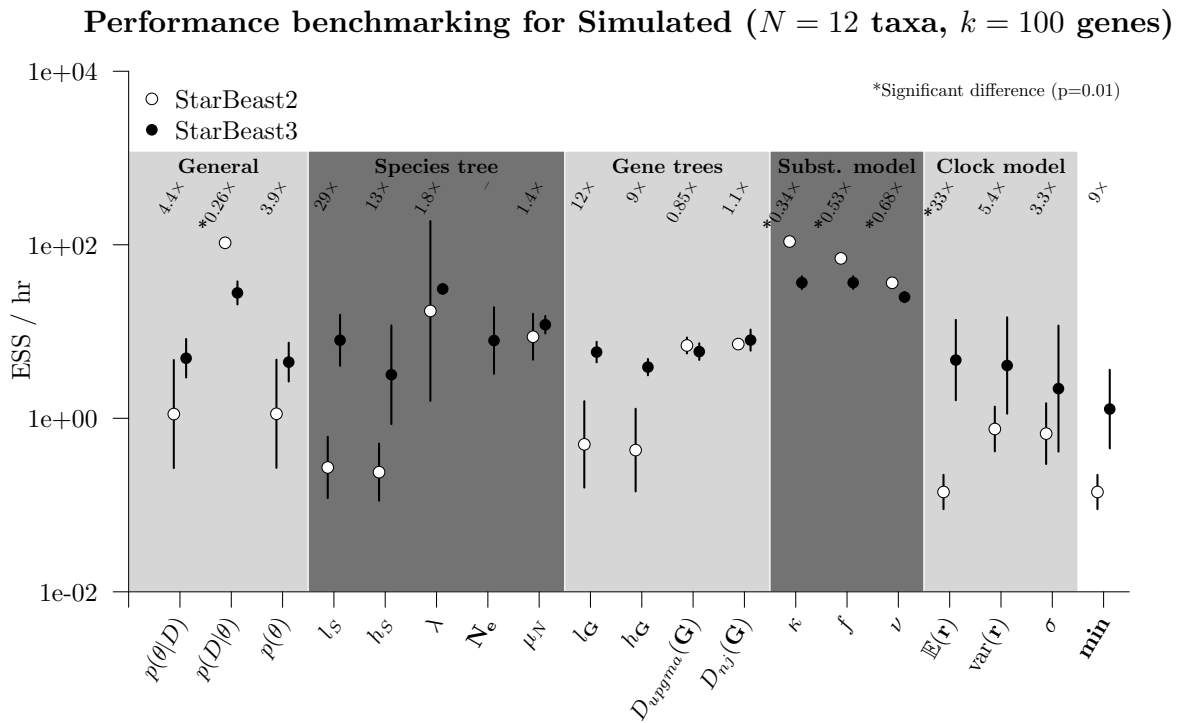
Performance benchmarking for Simulated ($N = 12$ taxa, $k = 100$ genes)

Fig. S9: Performance benchmarking the 4 species Simulated dataset. See Fig. S6 caption for figure notation.

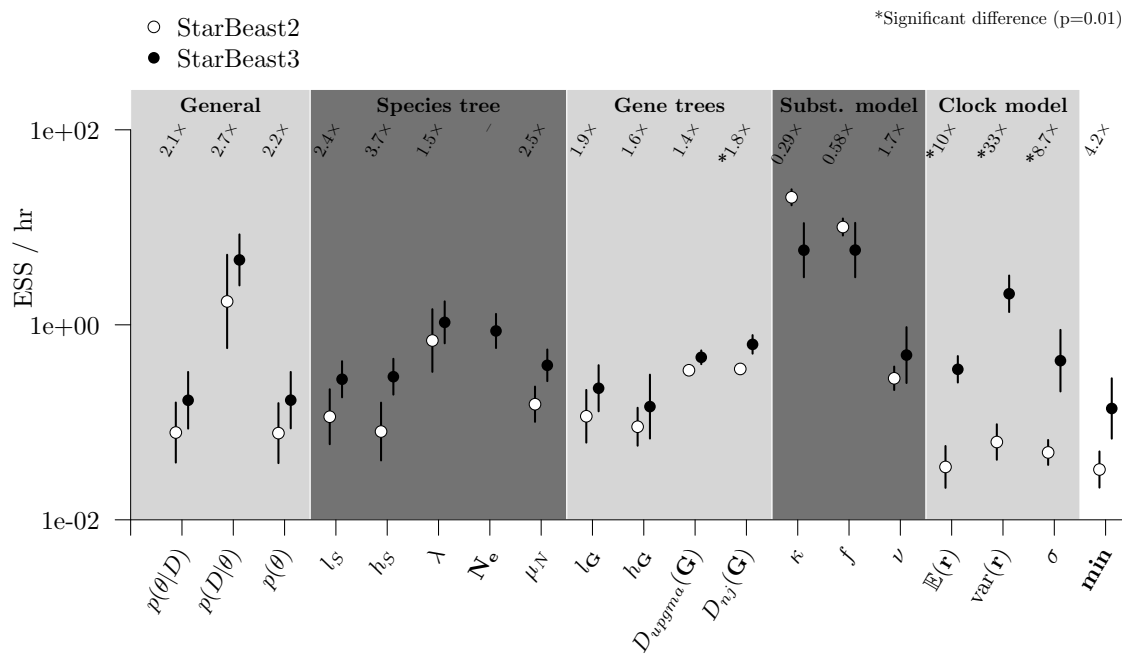**Performance benchmarking for Simulated ($N = 48$ taxa, $k = 100$ genes)**

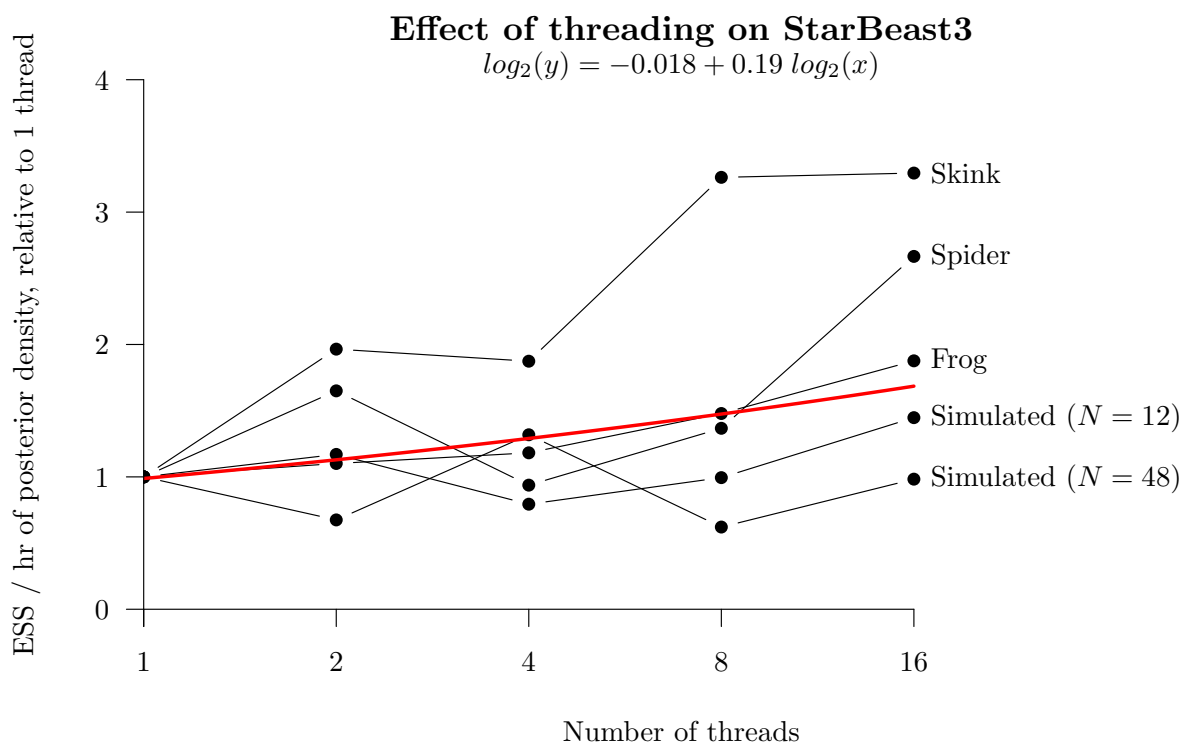**Fig. S10:** Performance benchmarking the 16 species Simulated dataset. See Fig. S6 caption for figure notation.

Effect of threading on StarBeast3

$$log_2(y) = -0.018 + 0.19\, log_2(x)$$

ESS / hr of posterior density, relative to 1 thread

Number of threads

Skink

Spider

Frog

Simulated ($N = 12$)

Simulated ($N = 48$)

Fig. S11: Effect of threading on StarBeast3 performance. Each point represents the ESS/hr of the posterior density $P(\theta|D)$ (averaged across 5 replicates), for the indicated thread count and dataset. These terms are normalised to enable comparison across datasets, by dividing it by that of 1 thread. A linear model was fit to the ESS/hr and number of threads, each in log_2 space, and is reported at the top of the plot. The positive coefficient of the slope indicates that performance increased with the number of threads, across the range of threads considered. Parallel MCMC chain lengths were optimised using the adaptive scheme presented in Fig. S3.

4   Discussion

4.1   The Next Generation of Bayesian MCMC Operators

In recent years, Bayesian MCMC proposals have advanced significantly beyond that of the unidimensional random walk. The use of adaptive algorithms and advanced proposal kernels have become increasingly prevalent in recent years (Haario et al., 2001; Vihola, 2012; Benson et al., 2018; Yang and Rodríguez, 2013). In phylogenetic inference in particular, tree proposals have been guided by conditional clade probabilities and parsimony scores (Höhna and Drummond, 2012; Zhang et al., 2020), and mirror kernels learn target distributions which act as "mirror images" (Thawornwattana et al., 2018), for instance.

Here, we introduced a range of recently developed MCMC operators to the multispecies coalescent, including Bactrian proposal kernels (Yang and Rodríguez, 2013), which have been successfully applied to bird phylogeny (Maliet et al., 2019), and tree "flex" operators (BICEPS; Bouckaert (2021)), which have been applied to COVID-19 genomic data (Douglas et al., 2021a). We also invoked a series of more meticulous operators which account for known correlations, such as the AVMN kernel (Baele et al., 2017), constant distance operators (Zhang and Drummond, 2020), and the narrow exchange rate operator (Douglas et al., 2021b), as well as adaptive operators that improve over the course of MCMC, such as the adaptable operator sampler (Douglas et al., 2021b), parallel gene tree operators, and the AVMN kernel (Baele et al., 2017). Indeed, these operators have yielded a software package which outperforms StarBeast2 by up to one-and-a-half orders of magnitude, depending on the dataset and the parameter.

In order for Bayesian inference to keep up with the large volumes of genomic data, the development of efficient, meticulous, and adaptive MCMC operators is essential.

4.2   Efficient Parallelised Bayesian Inference of the Multispecies Coalescent

As genomic data becomes increasingly available, concatenating genomic sequences and inferring the phylogeny of the species as that of the genes can become enticing. However, this approach makes for an inconsistent estimator of topology when divergence times are small (Pamilo and Nei, 1988), and a biased estimator of species divergence times and substitution rates when incomplete lineage sorting is present (Arbogast et al., 2002; Ogilvie et al., 2016; Mendes and Hahn, 2016). Bayesian multispecies coalescent methods address these issues, but at the drawback of their demanding computational runtimes.

Therefore, as multithreading technologies become increasingly ubiquitous, the appeal in parallelising multispecies inference becomes clear. StarBeast3 exploits the assumption of conditional independence between gene trees, by doing Bayesian inference on gene trees in parallel, and therefore it scales with the size of the problem. StarBeast3 can process large datasets (100+ genes) and achieve convergence several times faster than its predecessors.

4.3   A Balanced Traversal Through Parameter Space

All areas of parameter space should be explored approximately evenly during MCMC. If one area of parameter space is being explored more rapidly than another, then computational resources allotted to the former should be diverted to the latter. This is best exemplified by the phylogenetic substitution model which, despite requiring relatively little attention to converge, still requires full recalculation of the tree likelihood upon every proposal (Felsenstein, 1981). Conversely, tree topologies often converge rather poorly and can require significant attention to be rescued from local optima. By fine tuning our MCMC operator proposal probabilities, we have achieved a balanced traversal through all areas of the multispecies coalescent parameter space. Although some parameters converge slower for StarBeast3 than they do for StarBeast2 (such as those in the substitution model), the slowest parameters converge significantly faster for the former; up to $30\times$ as fast (see the **min** term in Fig. S6 – S10).

For StarBeast3, we employed adaptable operators which are able to learn the proposal probabilities of other operators based on their ability to explore a single area of parameter space (Douglas et al., 2021b).

However, there would be great benefit in an adaptable operator scheme which learns and applies a balanced exploration across different areas of parameter space on a problem-by-problem basis.

5   Conclusion

Here we introduce StarBeast3 – a software package for performing efficient Bayesian inference on genomic data under the multispecies coalescent model. We verified StarBeast3's correctness and we benchmarked its performance against StarBeast2, which is an order of magnitude faster than its still popular predecessor *BEAST. We showed that StarBeast3 is significantly faster than StarBeast2. Notably, relaxed clock parameters converged up to one-and-a-half orders of magnitude faster, but most importantly even the "slowest" parameters converged up to $22\times$ faster. Our adaptive operator scheme allows proposal probabilities to be learned on a problem-by-problem basis, making StarBeast3 suitable for a range of datasets. By estimating effective population sizes (instead of analytically integrating the term out), we were able to parallelise gene tree proposals, and demonstrated that doubling the number of allotted threads was associated with an increase in performance by around 14%. StarBeast3 is highly effective at performing fast Bayesian inference on large datasets.

6   Software Availability

StarBeast3 is available as an open-source BEAST 2 package with an easy-to-use graphical user interface. Instructions for downloading and running StarBeast3 can be found at `https://github.com/rbouckaert/starbeast3`.

7   Acknowledgements

References

Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annual review of Ecology and Systematics 33:707–740.

Baele, G., P. Lemey, A. Rambaut, and M. A. Suchard. 2017. Adaptive mcmc in bayesian phylogenetics: an application to analyzing partitioned data in beast. Bioinformatics 33:1798–1805.

Ballesteros, J. A. and P. P. Sharma. 2019. A critical appraisal of the placement of xiphosura (chelicerata) with account of known sources of phylogenetic error. Systematic Biology 68:896–917.

Barrow, L. N., H. F. Ralicki, S. A. Emme, and E. M. Lemmon. 2014. Species tree estimation of north american chorus frogs (hylidae: Pseudacris) with parallel tagged amplicon sequencing. Molecular phylogenetics and evolution 75:78–90.

Belfiore, N. M., L. Liu, and C. Moritz. 2008. Multilocus phylogenetics of a rapid radiation in the genus thomomys (rodentia: Geomyidae). Systematic Biology 57:294–310.

Benson, A., N. Friel, et al. 2018. Adaptive mcmc for multiple changepoint analysis with applications to large datasets. Electronic Journal of Statistics 12:3365–3396.

Bickford, D. et al. 2007. Cryptic species as a window on diversity and conservation. Trends in ecology & evolution 22:148–155.

Blom, M., J. G. Bragg, S. Potter, and C. Moritz. 2016a. Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian lizards. Sys Bio 66:352–366.

Blom, M., P. Horner, and C. Moritz. 2016b. Convergence across a continent: adaptive diversification in a recent radiation of Australian lizards. Proc. R. Soc. B 283.

Bouckaert, R. 2016. Phylogeography by diffusion on a sphere: whole world phylogeography. PeerJ 4:e2406.

Bouckaert, R. 2021. An efficient coalescent epoch model for bayesian phylogenetic inference. bioRxiv .

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, et al. 2019. Beast 2.5: An advanced software platform for bayesian evolutionary analysis. PLoS computational biology 15:e1006650.

Bouckaert, R. et al. 2013. Evolutionary rates and HBV: issues of rate estimation with Bayesian molecular methods. Antivir Ther 18:497–503.

Bouckaert, R. R. 2020. Obama: Obama for bayesian amino-acid model averaging. PeerJ 8:e9460.

Bragg, J., others, and M. C. 2017. Resources for phylogenomic analyses of Australian terrestrial vertebrates. Molecular ecology resources 17:869–876.

Bryson Jr, R. W., C. W. Linkem, C. J. Pavón-Vázquez, A. Nieto-Montes de Oca, J. Klicka, and J. E. McCormack. 2017. A phylogenomic perspective on the biogeography of skinks in the plestiodon brevirostris group inferred from target enrichment of ultraconserved elements. Journal of Biogeography 44:2033–2044.

Cadena, C. D., others, C. Moritz, et al. 2011. Latitude, elevational climatic zonation and speciation in New World vertebrates. Proc R Soc Lond [Biol] Page rspb20110720.

Collienne, L. and A. Gavryushkin. 2021. Computing nearest neighbour interchange distances between ranked phylogenetic trees. Journal of Mathematical Biology 82:1–19.

Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in ecology & evolution 24:332–340.

Douglas, J. 2020. Uglytrees: a browser-based multispecies coalescent tree visualiser. Bioinformatics .

Douglas, J., J. L. Geoghegan, J. Hadfield, R. Bouckaert, M. Storey, X. Ren, J. de Ligt, N. French, and D. Welch. 2021a. Real-time genomics for tracking severe acute respiratory syndrome coronavirus 2 border incursions after virus elimination, new zealand. Emerging Infectious Diseases 27:2361.

Douglas, J., R. Zhang, and R. Bouckaert. 2021b. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. PLoS computational biology 17:e1008322.

Drummond, A. J. and R. R. Bouckaert. 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press, Cambridge.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLOS Biol 4:e88.

Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161:1307–1320.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Edwards, S. V., Z. Xi, A. Janke, et al. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol Phyl Evol 94:447–462.

Felsenstein, J. 1981. Evolutionary trees from dna sequences: a maximum likelihood approach. Journal of molecular evolution 17:368–376.

Fujita, M. K., others, and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in ecology & evolution 27:480–488.

Geman, S. and D. Geman. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence Pages 721–741.

Grummer, J. A. et al. 2013. Species delimitation using Bayes factors: simulations and application to the Sceloporus scalaris species group. Sys Bio 63:119–133.

Haario, H., E. Saksman, J. Tamminen, et al. 2001. An adaptive metropolis algorithm. Bernoulli 7:223–242.

Hamilton, C. A., A. R. Lemmon, E. M. Lemmon, and J. E. Bond. 2016. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. BMC Evolutionary Biology 16:1–20.

Hasegawa, M., H. Kishino, and T.-a. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of molecular evolution 22:160–174.

Heled, J. and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27:570–580.

Heled, J. and A. J. Drummond. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Systematic biology 61:138–149.

Hohna, S., M. Defoin-Platel, and A. J. Drummond. 2008. Clock-constrained tree proposal operators in bayesian phylogenetic inference. Pages 1–7 in 2008 8th IEEE International Conference on BioInformatics and BioEngineering IEEE.

Höhna, S. and A. J. Drummond. 2012. Guided tree topology proposals for bayesian phylogenetic inference. Systematic biology 61:1–11.

Jones, G. 2017. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. Journal of mathematical biology 74:447–467.

Jones, G., Z. Aydin, and B. Oxelman. 2015. Dissect: an assignment-free bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 31:991–998.

Kang, Y. J., S. K. Kim, M. Y. Kim, P. Lestari, K. H. Kim, B.-K. Ha, T. H. Jun, W. J. Hwang, T. Lee, J. Lee, et al. 2014. Genome sequence of mungbean and insights into evolution within Vigna species. Nature communications 5:ncomms6443.

Leaché, A. D., M. K. Fujita, V. N. Minin, and R. R. Bouckaert. 2014. Species delimitation using genome-wide SNP data. Sys Bio 63:534–542.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. PLOS Comput Biol 5.

Liu, L., D. K. Pearl, R. T. Brumfield, and S. V. Edwards. 2008. Estimating species trees using multiple-allele dna sequence data. Evolution: International Journal of Organic Evolution 62:2080–2091.

Liu, L., S. Wu, and L. Yu. 2015. Coalescent methods for estimating species trees from phylogenomic data. Journal of Systematics and Evolution 53:380–390.

Liu, L., L. Yu, and S. V. Edwards. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology 10:302.

Liu, L., L. Yu, L. Kubatko, D. K. Pearl, and S. V. Edwards. 2009. Coalescent methods for estimating phylogenetic trees. Mol Phylogenet Evol 53:320–328.

Maddison, W. P. 1997. Gene trees in species trees. Systematic biology 46:523–536.

Maliet, O., F. Hartig, and H. Morlon. 2019. A model with many small shifts for estimating species-specific diversification rates. Nature ecology & evolution 3:1086–1092.

Mendes, F. K. and M. W. Hahn. 2016. Gene tree discordance causes apparent substitution rate variation. Systematic biology 65:711–721.

Mirarab, S., R. Reaz, M. S. Bayzid, T. Zimmermann, et al. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

Mirarab, S. and T. Warnow. 2015. ASTRAL-ii: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.

Mitchell, K. J., B. Llamas, J. Soubrier, et al. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science 344:898–900.

Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 344:305–311.

Ogilvie, H., R. Bouckaert, and A. Drummond. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Mol Biol Evol Page msx126.

Ogilvie, H., others, and A. J. Drummond. 2016. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Sys Bio 65:381–396.

Oliver, J. C. 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. Evolution 67:1823–1830.

Pamilo, P. and M. Nei. 1988. Relationships between gene trees and species trees. Molecular biology and evolution 5:568–583.

Pepper, M., C. Moritz, et al. 2013. Speciation on the rocks: integrated systematics of the heteronotia spelea species complex (Gekkota; Reptilia). PLOS One 8:e78110.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in bayesian phylogenetics using tracer 1.7. Systematic biology 67:901.

Rannala, B. and Z. Yang. 2017. Efficient Bayesian species tree inference under the multispecies coalescent. Sys Bio Page syw119.

Roberts, G. O., A. Gelman, W. R. Gilks, et al. 1997. Weak convergence and optimal scaling of random walk metropolis algorithms. The annals of applied probability 7:110–120.

Rowe, K. C., K. P. Aplin, P. R. Baverstock, and C. Moritz. 2011. Recent and rapid speciation with limited morphological disparity in the genus Rattus. Sys Bio 60:188–203.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4:406–425.

Sauquet, H., S. Y. W. Ho, M. A. Gandolfo, G. J. Jordan, P. Wilf, D. J. Cantrill, M. J. Bayly, L. Bromham, G. K. Brown, R. J. Carpenter, D. M. Lee, D. J. Murphy, J. M. K. Sniderman, and F. Udovicic. 2011. Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of nothofagus (fagales). Systematic Biology 61:289–313.

Sokal, R. R. 1958. A statistical method for evaluating systematic relationships. Univ. Kansas, Sci. Bull. 38:1409–1438.

Song, S. et al. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. PNAS 109:14942–14947.

Stenson, P. D. et al. 2017. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Human genetics 136:665–677.

Thawornwattana, Y., D. Dalquen, Z. Yang, et al. 2018. Designing simple and efficient markov chain monte carlo proposal kernels. Bayesian Analysis 13:1037–1063.

Vihola, M. 2012. Robust adaptive metropolis algorithm with coerced acceptance rate. Statistics and Computing 22:997–1008.

Whelan, S. and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular biology and evolution 18:691–699.

Xi, Z., L. Liu, and C. C. Davis. 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol Phyl Evol 92:63–71.

Yang, Z. and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. PNAS 107:9264–9269.

Yang, Z. and B. Rannala. 2014. Unguided species delimitation using DNA sequence data from multiple loci. Mol Biol Evol 31:3125–3135.

Yang, Z. and C. E. Rodríguez. 2013. Searching for efficient markov chain monte carlo proposal kernels. Proceedings of the National Academy of Sciences 110:19307–19312.

Yule, G. U. 1925. Ii. a mathematical theory of evolution, based on the conclusions of dr. jc willis, fr s. Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character 213:21–87.

Zhang, C., J. P. Huelsenbeck, and F. Ronquist. 2020. Using parsimony-guided tree proposals to accelerate convergence in bayesian phylogenetic inference. Systematic Biology .

Zhang, R. and A. Drummond. 2020. Improving the performance of bayesian phylogenetic inference under relaxed clock models. BMC Evolutionary Biology 20:1–28.