Recent advances in Bayesian phylogenetics offer substantial computational savings to accommodate increased genomic sampling that challenges traditional inference methods. In this review, we begin with a brief summary of the Bayesian phylogenetic framework, and then conceptualize a variety of methods to improve posterior approximations via Markov chain Monte Carlo (MCMC) sampling. Specifically, we discuss methods to improve the speed of likelihood calculations, reduce MCMC burn-in, and generate better MCMC proposals. We apply several of these techniques to study the evolution of HIV virulence along a 1536-tip phylogeny and estimate the internal node heights of a 1000-tip SARS-CoV-2 phylogenetic tree in order to illustrate the speed-up of such analyses using current state-of-the-art approaches. We conclude our review with a discussion of promising alternatives to MCMC that approximate the phylogenetic posterior.

This article is part of a discussion meeting issue 'Genomic population structures of microbial pathogens'.

### 1. Introduction

Pathogen genomic data have become an invaluable resource to study microbial evolution and infectious disease transmission. The widespread adoption of whole-genome sequencing technologies ensures an ever-increasing scale at which genomic data have become available. EnteroBase for example offers over 300,000 assembled genomes for several bacterial genera and facilitates the genotyping and characterisation of novel strains against all natural populations [1]. Rapidly expanding genome sequencing efforts are being undertaken in response to every new viral epidemic. Over 1500 Ebolavirus genomes were generated during the 2013–2016 West-African Ebola epidemic, representing over 5% of the known cases [2]. At the time, this was considered as a new scale in viral genomic monitoring, and the adoption of portable genome sequencing also offered the perspective of providing it in close to real time [3]. Global sequencing efforts during the COVID-19 pandemic have again taken genomic epidemiology to new heights, with over 4,000,000 SARS-CoV-2 genomes available in GISAID at the time of writing [4].

The potential contribution of genomic epidemiology to the public health response during outbreaks and epidemics hinges on the ability to provide insights in short turnaround times. This has motivated the development of analytic and visualization tools such as Nextstrain for real-time tracking of pathogen evolution and spread [5]. However, many of the phyldynamic models that rely on phylogenies to unravel pathogen transmission dynamics have traditionally been implemented in Bayesian inference frameworks. A Bayesian approach accommodates phylogenetic uncertainty in the phyldynamic estimates and

© 2022 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.
offers extensive modelling flexibility, but represents a computationally challenging endeavour as it requires averaging over all plausible evolutionary histories. This becomes particularly problematic when confronting data encompassing thousands of genomes. The COVID-19 genomic response has therefore demonstrated the pressing need for more scalable phylodynamic inference methodologies.

In this review, we present recent developments in Bayesian phylogenetics that begin to answer this call for scalable phylodynamic methods. In addition to offering a unified and coherent framework to quantify uncertainty in phylogenetic parameters, the Bayesian approach allows one to learn the joint distribution of all phylogenetic parameters simultaneously. This is particularly important when combining the phylogenetic model with molecular clock models, tree generative models and trait evolutionary models, which is the focus of the BEAST Bayesian inference framework [6,7]. Here we briefly review this Bayesian phylogenetic framework to guide further exposition. For an introductory review of Bayesian phylogenetics, see Nascimento et al. [8].

Within the Bayesian phylogenetic setting, one chooses a data generative model to describe the evolution of genetic sequence data, phenotypic traits, the geographic displacement of sampled taxa and more. This data generative model is referred to as the likelihood. Subsequently, one places a prior distribution on parameters of the likelihood that describe a plausible range of values these parameters may take. Next, one estimates the posterior distribution of parameters of interest, typically via Markov chain Monte Carlo (MCMC) integration, the main workhorse of Bayesian phylogenetic inference. To begin MCMC sampling, initialize the Markov chain by setting a possible state for each unknown model parameter. MCMC sampling proceeds by iterating over two steps. First, propose a new state where one or more parameters are updated and compare the posterior density of the new state with the old. Next, accept the new state if the corresponding posterior is higher but occasionally accept the new state if the posterior is lower. See Van Ravenzwaaij et al. [9] for an introduction to MCMC. For the purpose of this review, we classify the current limitations of inference as follows:

1. every step in MCMC requires evaluating the posterior ratio which in turn requires a potentially computationally demanding likelihood calculation;
2. tree-space can be enormous, requiring many steps before the MCMC algorithm is able to sample from the posterior (the burn-in period);
3. standard MCMC proposals only allow one to make small steps, limiting the efficiency by which samples from the posterior can be obtained.

In order to maximize the amount of information that can be obtained about the posterior per unit of time, scalable machinery largely focuses on directly improving MCMC by speeding up likelihood calculation, reducing burn-in, or developing more efficient proposals as well as indirectly improving MCMC by constructing models amenable to scalable MCMC techniques.

The first challenge, slow likelihood computation, is a burden shared by various phylogenetic inference approaches and this has motivated the development of libraries for optimized likelihood evaluation [10,11]. The BEAST inference framework relies on the BEAGLE high-performance general-purpose library (https://github.com/beagle-dev/beagle-lib) and we refer interested readers to Ayres et al. [11] and Baele et al. [12] for further details. From a practical perspective, it is important to note that BEAGLE implements parallelism in the likelihood calculation offering significant speed-ups on modern graphics processing units (GPUs) and multicore central processing units (CPUs). In the remainder of this paper, we turn our attention to inference machinery that reduces burn-in and generates more efficient proposals, and finally conclude with a discussion of novel methods alternative to MCMC for posterior estimation.

2. Reducing burn-in via ‘online’ inference

Bayesian phylogenetic inference via MCMC often requires days or weeks of runtime to obtain reasonable estimates of parameters of interest. The addition of new observations extends runtime further by requiring one to restart their MCMC chain as new data become available. Additionally, as the number of taxa in an analysis grows, so does the dimension of parameter space one must explore. For example, the number of possible evolutionary trees grows exponentially with the number of taxa under study. As the size of parameter space grows, it takes longer for the chain to reach regions of high posterior density. This period of searching for the posterior is called the ‘burn-in’ period of the chain. The need to update analyses as new data accumulate is a scenario that now typically arises with real-time genomic responses to viral outbreaks. Sequential analyses associated with growing parameter spaces would therefore benefit from initiating the chain from an informed starting point. Gill et al. [13] developed a principled distance-based procedure to add new taxa to an existing tree and install previous MCMC chain samples into a new chain thus allowing one to incorporate new data in an ‘online’ fashion. The approach of Gill et al. [13] is publicly available in the BEAST software package. Lemey et al. [14] employ this online framework to study geographic spread of SARS-CoV-2 in near real time, adding viral genomes as they became available.

We illustrate the online procedure of Gill et al. [13] by reconstructing the evolutionary history of 588 full SARS-CoV-2 genomes (29 409 nt) collected from 30 January to 30 September 2020. We follow Lemey et al. [14] and use an HKY85 model with among-site rate heterogeneity modelled using a four-category discretised $\Gamma$ distribution, and a coalescent tree prior. In figure 1, we augment the most recent MCMC sample of our 588-tip tree to include 132 new SARS-CoV-2 sequences sampled between 1 October and 15 October 2020. This procedure takes seconds (less than 1 min) to complete. Notice that all added sequences that belong to the B.1.177 variant, which rapidly spread from Spain to many other European countries over the summer [15], correctly form a monophyly before a single MCMC step is taken with the new 720-tipped tree. As a result, the new, online chain initializes with a log joint density evaluation of $-79 768$. In comparison, across five runs from five different naive starting trees with 720 tips, it took an average of 6.4 h to sample a joint density evaluation at least as favourable.
3. Constructing new chains via better Markov chain Monte Carlo proposals

(a) Adaptive Markov chain Monte Carlo

As developments in sequencing technology increase both the number and length of genetic sequences available to researchers, it is standard and often appropriate practice to partition long genetic sequences into smaller components (perhaps into individual genes) and assume each partition has its own, independent evolutionary model [16]. This partitioning of the data results in a more flexible and realistic statistical model but dramatically expands the number of parameters to be inferred, as it associates each individual partition with its own set of parameters.

Classical MCMC update procedures, such as univariable Metropolis–Hastings (uMH) update individual parameters sequentially (rather than simultaneously). Univariable samplers are poorly suited for this high-dimensional scenario, as the number of potentially costly likelihood evaluations required to update all parameters scales with the total number of parameters. Similarly, simultaneously updating all parameters (or large blocks of parameters) independently in a single MCMC step typically has extremely low acceptance probability since many parameters have strong and complex posterior relationships that such independent proposals ignore.

Adaptive MCMC (e.g. [17]) addresses this issue by updating all parameters (or large blocks of parameters) simultaneously in a way that accounts for their posterior correlations. Specifically, adaptive MCMC uses the history of the Markov chain itself to approximate posterior covariance between relevant parameters and uses that covariance to propose correlated, rather than independent, parameter updates. Owing to this feature, the adaptive kernel is not Markovian but the resulting chain is still valid [18,19]. As the chain progresses forward, the adaptive MCMC procedure can continuously re-evaluate the posterior covariance and tune relevant hyper-parameters to maximize sampling efficiency. Baele et al. [20] adapt this procedure to Bayesian phylogenetics in the BEAST framework and demonstrate an order of magnitude increase in sampling efficiency for partitioned datasets. While the adaptive procedure is not limited to partition parameters, Baele et al. [20] exploit the synergistic interaction between adaptive MCMC and parallel computation in partitioned models. In multi-processor computing environments, each model partition is typically assigned to a single processor. When performing uMH to update a parameter in a partitioned model, only the processor assigned to that partition recomputes the likelihood while the others remain idle. However, when all parameters are updated simultaneously, all processors are simultaneously engaged. As such, while a single proposal in adaptive MCMC requires likelihood calculations over numerous partitions, these calculations require considerably less time than the same number of calculations performed sequentially.

(b) Hamiltonian Monte Carlo

Like adaptive MCMC, Hamiltonian Monte Carlo (HMC; [21]) can propose updates to many parameters simultaneously with high acceptance probabilities. The key difference is that adaptive MCMC learns about the properties of the posterior empirically (i.e. it relies only on the past history of the Markov chain), while HMC uses the analytic properties of the posterior distribution.

Conceptually, HMC treats the state of model parameters as the position of a particle in a landscape informed by the posterior. At each step in the Markov chain this ‘particle’ (the parameters) is kicked in a random direction and allowed to traverse this posterior-informed landscape according to Hamiltonian dynamics (imagine a hockey puck moving around an uneven, frictionless surface). The position of this particle after a pre-defined amount of time is the parameter proposal.

If the trajectory through the posterior landscape is calculated exactly, then the parameter proposal is accepted with 100% probability. However, this continuous, nonlinear trajectory cannot typically be calculated analytically and must be approximated by a series of discrete, linear steps through the parameter space [22]. In general, taking many small steps results in a better approximation (and therefore higher acceptance probability) than fewer, larger steps. However, each of these discrete steps requires evaluating the derivative or gradient of the log-posterior with respect to the parameters of interest, which is typically computationally rate-limiting. As such, efficiently exploring the posterior via HMC requires balancing the computational cost of the gradient calculations.

Figure 1. The online addition of 132 SARS-CoV-2 sequences to a 588-tip time-measured tree drawn from the posterior. Appended branches are blue while original branches are black. We omit timescale since this augmented tree is not sampled from the posterior.
with the size of the discrete steps. Fortunately, methods such as the no-U-turn sampler [23] automatically tune these hyperparameters and have found success in Bayesian phylogenetic applications, such as for inference of high-dimensional trait correlation [24].

We showcase the computational efficiency of HMC against the standard uMH in figure 2. Figure 2a displays the trajectory of two clock rate parameters $\gamma_1$ and $\gamma_2$ over their joint posterior using both HMC and uMH for equivalent amounts of BEAST runtime. For this demonstration, we simulate 12 000 nt of sequence data on a fixed tree with three tips and two different local clocks, $\gamma_1 = 0.003$ and $\gamma_2 = 0.006$. We choose a three-tip tree since it is the smallest example with identifiable clock structure. We estimate the branch-specific clock rates under a relaxed clock model using the HMC machinery developed and described by Ji et al. [25]. In figure 2b, we demonstrate the relative efficiency of HMC over uMH in inferring branch-specific rates of phenotypic evolution on a large, fixed, 1536-tip HIV-1 tree. We report effective sample size (ESS) per second of BEAST runtime as a measure of efficiency since the ESS of a parameter is the number of effectively independent posterior samples in a given chain. We use HMC to investigate covariance between two traits associated with HIV virulence, the ‘gold standard viral load’ (GSVL) and CD4 cell count slope decline as they evolve on the fixed topology presented by Blanquart et al. [26]. Previous analyses of these data study trait covariation while assuming a time homogeneous, strict Brownian diffusion process guides trait evolution [27]. HMC makes it possible to learn about trait covariation under the more general relaxed random walk (RRW) model [28]. Fisher et al. [29] develop HMC machinery to make inference under the RRW. We find that the posterior mean estimate of covariance between GSVL and CD4 cell count slope decline is $-0.163$ with 95% highest posterior density (HPD) interval $(-0.21, -0.11)$ and $-0.151 (-0.20, -0.10)$ under the strict Brownian diffusion and RRW, respectively. Similarly, 95% HPD intervals of individual variance estimates contain significant overlap between models as well. Overall, possible rate heterogeneity in the diffusion process does not influence the estimated covariance between GSVL and CD4 cell count slope decline. We, however, would not have known this without the availability of HMC to fit the richer model.

We additionally report the computational speed-up of estimating an internal node age using HMC versus uMH for the topologically fixed SARS-CoV-2 tree with 1000 full genome sequences (29 409 nt) of Lemey et al. [14]. We follow the model specifications of Lemey et al. [14] and model nucleotide evolution with an HKY85 with four category discrete-Γ site rate model and a coalescent tree prior for its induced prior on the node heights. We compare the accumulation of ESS per second of BEAST runtime using both traditional uMH BEAST default proposals as suggested by the BEAST graphical user interface, BEAUti, and recent HMC inference machinery developed by Ji et al. [30]. We estimate the age of the B.1.177 clade to be 0.45 years with 95% HPD interval $(0.40, 0.50)$ and we obtain this estimate within a couple of hours, accumulating 63 ESS $1000 \text{ s}^{-1}$ using the HMC machinery. Alternatively, under uMH, we do not obtain reliable estimates as quickly and only accumulate
Bayesian phylogenetic applications of variational inference employ a parametric family of distributions. Under this paradigm, one assumes parameters of interest, e.g. branch lengths in a phylogeny, are the parameters of the parametric family. Zhang & Matsen IV [39] develop a general variational phylogenetic framework to guide exploration in tree space while Dang & Kishino [40] offer a framework to make inference on the CAT model of amino acid evolution [41]. While these examples of phylogenetic variational inference show promising speed-up over some MCMC methods in their applications, current variational assumptions restrictively assume independence across variational parameters and examples of performance with more than a modest number of taxa (more than 100) remain open for exploration. It also remains an unexplored question, to what extent Bayesian phylogenetic variational inference may complement existing phylogenetic MCMC machinery. One possible direction of future exploration may employ fast variational approximation to set more informed starting points for MCMC analyses.

5. Discussion
Scalable Bayesian phylogenetics is growing to accommodate the challenge of analyzing increasingly large datasets under complex models. Some methods, e.g. adaptive MCMC, HMC, parallelized computing, and SMC are new additions to the Bayesian phylogenetic toolkit, yet these are not at all new tools. Bayesian phylogenetics has been late to adopt modern inference approaches. One might be inclined to ask ‘why?’ One probable reason is that under many phylogenetic models, model parameters grow faster than the observed data. For example, when one observation is added to the tree, the number of branches, and thus the number of branch lengths to estimate, increases by two. Compound this with the difficulty of integrating over uncertainty in the combinatorially complex phylogenetic tree space and the result is that many methods may work for small toy examples but fail when scaled to larger analyses.

While the techniques presented here may alleviate computational burden in a variety of analyses, we have not provided an exhaustive overview of all scalable Bayesian phylogenetic approaches and their implementations. For example, Altekar et al. [42] develop an alternative adaptive MCMC approach within a Bayesian framework that runs multiple MCMC chains in parallel. Each chain is run at a different ‘temperature’. This effectively flattens the posterior so that more proposals are accepted and local extrema are avoided. Müller & Bouckaert [43] provide an implementation in BEAST. Another approach that handles multi-modality is the nested sampling method of Russel et al. [44]. Nested sampling, in a way conceptually similar to tempering, begins with a wide prior that grows successively more constrained as the sampler proceeds. Here the aim is model selection by marginal likelihood comparison and multi-modal inference is a by-product. Additionally, there are many recent developments aimed to improve tree sampling. We encourage readers interested in learning about efficient tree proposals to see Rannala & Yang [45] for inference under the multispecies coalescent, and Höhna & Drummond [46] for a comparison of the trade-offs between computationally expensive guided tree proposals and naive fast proposals.
Data accessibility. We provide reproducible instructions and open access to the BEAST XML files to recreate the analyses in this paper at: https://github.com/suchard-group/scalableBayesianPhylogenetics.

Authors’ contributions. A.F.: formal analysis, visualization, writing—original draft, writing—review and editing; G.H.: visualization, writing—original draft; X.J.: software, writing—original draft; G.B.: software, writing—review and editing; M.A.S.: conceptualization, writing—review and editing; P.L.: conceptualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interest.

Funding. The research leading to these results has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 752542—ReservoirDOCS). The Artic Network receives funding from the Wellcome Trust through project 206298/Z/17/Z. A.A.F., G.W.H., X.J. and M.A.S. are partially supported by NIH grant nos. U19 AI135995, R56 AI149004, R01 AI153044, F31 AI145824, T32-HG002536 and R01 AI162611. G.B. acknowledges support from the Interne Fondsen KU Leuven/Internal Funds KU Leuven under grant agreement no. C14/18/094 and the Research Foundation—Flanders (Fonds voor Wetenschappelijk Onderzoek—Vlaanderen, G0E1420N, G098321N). P.L. acknowledges support by the Research Foundation—Flanders (Fonds voor Wetenschappelijk Onderzoek—Vlaanderen, G0D5117N and G051322N).

References

1. Zhou Z et al. 2019 The Enterobase user’s guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia coli genomic diversity. Genome Res. 30, 138–152. (doi:10.1101/gx.251678.119)
2. Dudas G et al. 2017 Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature 544, 309–315. (doi:10.1038/nature22040)
3. Quick J et al. 2016 Real-time, portable genome sequencing for Ebola surveillance. Nature 530, 228–232. (doi:10.1038/nature16996)
4. Elbe S, Buckland-Merrett G. 2017 Data, disease and diplomacy: GIS4D’s innovative contribution to global health. Glob. Chang. 1, 33–46. (doi:10.1002/gch2.1018)
5. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2014 The Phylogenetic Likelihood Library. Mol. Biol. Evol. 37, 1970–1979. (doi:10.1093/molbev/msu240)
6. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018 Bayesian phylogenetic and phylogeographic data integration using BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15, e1006650. (doi:10.1371/journal.pcbi.1006650)
7. Nascimento FF, Dos Reis M, Yang Z. 2017 A formal definition of the SARS-CoV-2 viral load in Europe. Trends in Microbiol. 25, 349–367. (doi:10.1016/j.tim.2016.10.007)
8. Suchard MA, Lemey P, Baele G, Ayres DL. 2018 Adaptive Metropolis algorithm. Ann. Appl. Stat. 11, 20210242. (doi:10.1214/20-aoas1394)
9. Ji X, Zhang Z, Holbrook A, Nishimura A, Baele G, Rambaut A, Lemey P, Suchard MA. 2020 Graduate students on trees: a linear-time O(N)-dimensional gradient for statistical phylogenetics. Mol. Biol. Evol. 37, 3047–3060. (doi:10.1093/molbev/msaa130)
10. Blanchart F et al. 2017 Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. PLoS Biol. 15, e2001855. (doi:10.1371/journal.pbio.2001855)
11. Hassler G, Tolkoff MR, Allen WL, Ho LST, Lemey P, Suchard MA. 2022 Inferring phenotypic trait evolution on large trees with many incomplete measurements. J. Am. Stat. Assoc. 117, 678–692. (doi:10.1080/01621459.2020.1799812)
12. Lee P, Rambaut A, Welch RJ, Suchard MA. 2010 Phylogeography takes a relaxed random walk in continuous space and time. Mol. Biol. Evol. 27, 1877–1885. (doi:10.1093/molbev/msq067)
13. Fisher AA, Ji X, Zhang Z, Lemey P, Suchard MA. 2021 Relaxed random walks at scale. J. Biol. 70, 258–267. (doi:10.1093/ysb/gob056)
14. Ji X, Fisher AA, Su S, Thorne JL, Potter B, Lemey P, Baele G, Suchard MA. 2021 Scalable Bayesian divergence time estimation with ratio transformations. arXiv:2110.13298. (doi:10.48550/arXiv.2110.13298)
15. Dinh V, Darling AE, Matsen IV FA. 2018 Online Bayesian phylogenetic inference: theoretical foundations via Sequential Monte Carlo. Syst. Biol. 67, 503–517. (doi:10.1093/sysbio/sys087)
16. Fourment M, Claywell BC, Dinh V, McCoy C, Matsen IV FA, Darling AE. 2018 Effective online Bayesian phylogenetics via sequential Monte Carlo with guided proposals. Syst. Biol. 67, 490–502. (doi:10.1093/sysbio/sys090)
17. Wang L, Wang S, Bouchard-Côté A. 2019 An annealed sequential Monte Carlo method for Bayesian phylogenetics. Syst. Biol. 68, 155–183. (doi:10.1093/sysbio/sys028)
18. Andreieu C, Doucet A, Holenstein R. 2010 Particle Markov chain Monte Carlo methods. J. R. Stat. Soc.
35. Wang L, Bouchard-Côté A, Doucet A. 2015 Bayesian phylogenetic inference using a combinatorial sequential Monte Carlo method. J. Am. Stat. Assoc. 110, 1362–1374. (doi:10.1080/01621459.2015.1054487)

36. Wang S, Wang L. 2021 Particle Gibbs sampling for Bayesian phylogenetic inference. Bioinformatics 37, 642–649. (doi:10.1093/bioinformatics/btaa867)

37. Bouchard-Côté A, Chen M, Kuo L, Lewis P. 2014 SMC (sequential Monte Carlo) for Bayesian phylogenetics. In Bayesian phylogenetics: methods, algorithms, and applications (eds M-H Chen, L Kuo, PO Lewis), pp. 163–185. London, UK: Chapman & Hall/CRC.

38. Blei DM, Kucukelbir A, McAuliffe JD. 2017 Variational inference: a review for statisticians. J. Am. Stat. Assoc. 112, 859–877. (doi:10.1080/01621459.2017.1285773)

39. Zhang C, Matsen IV FA. 2019 Variational Bayesian phylogenetic inference. In 7th Int. Conf. on Learning Representations, New Orleans, LA, 6–9 May 2019. OpenReview.net.

40. Dang T, Kishino H. 2019 Stochastic variational inference for Bayesian phylogenetics: a case of CAT model. Mol. Biol. Evol. 36, 825–833. (doi:10.1093/molbev/msz020)

41. Lartillot N, Philippe H. 2004 A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109. (doi:10.1093/molbev/msh112)

42. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004 Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415. (doi:10.1093/bioinformatics/btg427)

43. Müller NF, Bouckaert R. 2020 Adaptive parallel tempering for BEAST 2. bioRxiv. (doi:10.1101/603514)

44. Russel PM, Brewer BJ, Klaere S, Bouckaert RR. 2019 Model selection and parameter inference in phylogenetics using nested sampling. Syst. Biol. 68, 219–233. (doi:10.1093/sysbio/syy050)

45. Rannala B, Yang Z. 2017 Efficient Bayesian species tree inference under the multispecies coalescent. Syst. Biol. 66, 823–842. (doi:10.1093/sysbio/syw119)

46. Höhna S, Drummond AJ. 2012 Guided tree topology proposals for Bayesian phylogenetic inference. Syst. Biol. 61, 1–11. (doi:10.1093/sysbio/syr074)