Clock-constrained Tree Proposal Operators in Bayesian Phylogenetic Inference

Sebastian Höhna, Michael Defoin-Platel and Alexei J Drummond

Abstract— **Bayesian Markov chain Monte Carlo (MCMC) has become one of the principle methods of performing inference of phylogenetic trees. The MCMC algorithm requires the definition of a transition kernel over the state space, which depends on *Tree Proposal Operators*. So, the precise form of these operators has a large impact on the computational efficiency of the algorithm. In this paper we investigate the efficiency of different tree proposals specialized on *clock-constrained* phylogenetic trees. Two new operators are developed and their efficiency is compared to five standard operators. Each of the seven operators is tested individually on three synthetic datasets and eleven real datasets. In addition, the single operators are compared to different mixtures of operators. Results show that our new operators perform better than their standard counterparts, but no operator alone achieved a high efficiency on the full panel of data sets tested. Finally, our new proposed mixture using all operators together provides better performance than current techniques.**

I. Introduction

In evolutionary biology many research projects start with a phylogenetic analysis [1], *i.e.* with the reconstruction of an evolutionary history. The history is expressed for example as a genealogy or phylogeny, typically represented by a bifurcating tree. An estimate of the tree is obtained as a point estimate (e.g. by maximum likelihood) or by a set of trees (e.g. the 95% credible set of trees in a Bayesian estimate). Either way, one would really like to consider all possible trees when constructing this estimate. This gets exponentially more difficult when the numbers of species studied increases. In the past decades there have been several different approaches to find the best trees according to an optimality criterion such as the Parsimony Score or the likelihood of the tree. The main approaches are Maximum Parsimony, Maximum Likelihood and Bayesian Markov chain Monte Carlo (MCMC). Here we discuss details of the Bayesian MCMC approach which returns a posterior distribution of trees and divergence times. The distribution could be calculated directly using Bayes Theorem, which provides a means of calculating the probability of the hypothesis given the data. In context of phylogenetic inference the hypothesis is a phylogenetic tree and the data is for example a set of related molecular sequences from different species. Unfortunately, calculating the posterior distribution directly is not feasible in a reasonable amount of time. Instead, a sampling algorithm, such as the MCMC algorithm, can approximate the distribution via simulation of a Markov chain whose stationary distribution is the target posterior distribution of interest.

Manuscript received June 30, 2008.

S. Höhna, M. Defoin-Platel and A. Drummond are with the Department of Computer Science, University of Auckland, Private Bag 92019 Auckland, New Zealand
shhn001@aucklanduni.ac.nz
{michael.dp,alexei}@cs.auckland.ac.nz

For phylogenetic inference using Bayesian MCMC a new hypothesis h', *i.e.* a phylogenetic tree, is proposed by a Tree Proposal Operator (which we call simply operator in the rest of this paper). The proposal h' is only dependent on the current hypothesis h and is accepted if u, drawn uniformly from $U(0,1)$, satisfies:

$$u < \frac{P(h')P(h|h')}{P(h)P(h'|h)} = \frac{P(h')}{P(h)}\alpha_{\mathrm{H}}$$

with α_{H} known as the Hastings ratio [2]. Many operators are possible and the choice of operator has a large impact on the overall performance of the MCMC run, so that good operators can lead to much faster convergence to, and mixing in, the stationary distribution. The efficiency of different operators are the subject of this paper.

A. Tree Definitions

In this paper we represent the evolutionary history with a phylogenetic, bifurcating, rooted tree as shown in Fig. 1. A phylogenetic tree consists of $n-1$ internal nodes and

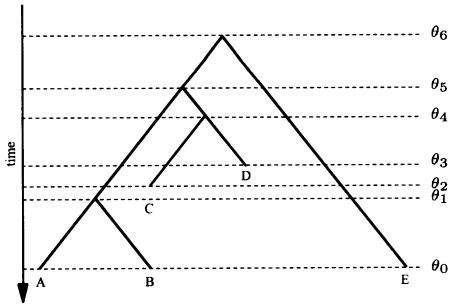

Fig. 1. Example of phylogenetic, bifurcated, rooted, clock-constrained tree with 5 species (A,B,C,D and E).

n leaves. Leaves, or tips, correspond to present species or individuals while internal nodes represent their ancestors and the root represents the common ancestor of all nodes. The arrangement of the nodes or branching pattern is called the tree topology τ. In clock-constrained trees, every node has a height θ (or age) which is the time elapsed between it and the youngest species. Therefore, the height of the youngest species is equal to zero and the height of the root is the time to the most recent common ancestor. Hence, the tree is

restricted so that every ancestor node has to be older than both its children.

Although some phylogenetic inference programs estimate unrooted unconstrained phylogenies, in this paper we will only consider rooted clock-constrained trees, such as those estimated by software packages like BEAST [3], Batwing [4], LAMARC [5] and Migrate [6].

B. The Standard Tree Proposal Operators

In this section we describe some standard operators for clock-constrained Bayesian MCMC. These operators are implemented in different software packages and used in our analysis together with our new operators (*cf.* Section II). The purpose of an operator is to propose a new tree by modifying the current one. An operator is *local* if the produced modifications concern a small part of the tree or *global* if distant parts of the topology are affected. Existing operators include:

Narrow Exchange The Narrow Exchange operator [7] randomly selects a node i and exchanges this one with a second node j that is the sibling of its parent node p. Ages of the nodes are compared to ensure that the new tree does not violate the clock constraints ($\theta_j < \theta_p$). For example if j, which becomes a new child of the parent p, is older than p then this would result in an older child which is not permitted. Operators which exchange two nodes, as the Narrow Exchange, are called *subtree swap* operators.

Wide Exchange The Wide Exchange operator is a generalization of the Narrow Exchange where the second node j is chosen randomly [7]. Then, both subtended subtrees are exchanged if the proposal does not violate the clock constraints. Since both nodes can be arbitrarily far away from each other, this subtree swap operator is considered to be global.

cNNI The (clock-constrained) Nearest Neighbor Interchange (NNI) operator [8] is another extension of the Narrow Exchange operator where the only difference is that the node heights are changed too. This expands the neighbourhood of proposals. A random node is chosen and exchanged with the sibling node of the parent node. Finally, the height of the parent is set randomly to a value between the maximum height of its children and its parent.

Wilson-Balding The Wilson-Balding operator [9] belongs to the group of *prune and regraft* operators. It selects randomly a node i which is not the root and not a leaf. Then the subtree rooted at i is pruned. A second node j, which has to be older than the remaining child of i, is chosen. Next, the previously pruned subtree is reattached under the second node j. Finally, to be consistent with clock-constrained trees, the node height of i is adjusted randomly between its new parent j and its new children.

Subtree-Slide The purpose of the Subtree-Slide operator is primarily to change the branch length rather than to propose a new topology. Operators using this technique are called *branch-change* operators. Nevertheless, this operator can make changes to the topology (mostly local) as a side-effect of large node height moves. First, a node is randomly chosen. Then, a path from the root to a leaf passing this node is selected. Finally, the node is slid a certain distance on the path in one direction, either towards the root or towards the leaf. The effect of this operator is similar to the LOCAL operator proposed by Mau et al. [10], [11] and Larget and Simon[12].

C. Related Work

The standard operators are implemented in different software packages as PAML [13], Batwing [4], LAMARC [5], Migrate [6], MrBayes [14], [15] and BEAST [3]. Yet, in most of the available software packages only a different subset of the known operators is implemented. Furthermore, no one has yet compared all these operators for clock-constrained trees, although an analysis is essential to answer which operator should be favoured [8]. We note that Lakner et al. [16] discussed this problem but only for unrooted trees. Our framework is similar to theirs in terms of the datasets and of the convergence diagnostic we used. However, inasmuch as they studied unrooted trees and therefore they used non clock aware operators, their results can not be extended easily here. Lakner et al. concluded that global operators perform better than local ones. They showed that extending the global operators with a bias favouring local proposals increases the performance. Furthermore, a branch-change operator, in their case the "LOCAL and Continuous Change operator", were performing worst in their tests. In Section IV, we will compare these statements to our results.

II. New Tree Proposal Operators

In this section we introduce two new operators that we implemented in the open source project beast-mcmc [3].

Intermediate Exchange The Intermediate Exchange operator is intermediate between the Narrow and Wide Exchange operators. It can be considered as an adapted Wide Exchange where the second node is not chosen uniformly. Instead, the selection process is given a bias to prefer local nodes with a higher probability. The probability that a node gets selected as the second depends on its path length to the first node. The shorter the path, the higher the probability (see Fig. 2). Let i denote the first node, j the second node and l_{ij} the path length between them in the current tree, then the probability of choosing j given i is:

$$P_i(j) = \frac{l_{ij}}{\sum_{k=1}^{n_i} l_{ik}} \tag{1}$$

where n_i are all possible nodes to swap from i considering the node height constraint. Since this operator is

not symmetric we need to compute the Hastings ratio:

$$\alpha_{\mathrm{H}} = \frac{\frac{1}{\sum_{k=1}^{n_i'} l_{ik}} + \frac{1}{\sum_{k=1}^{n_j'} l_{jk}}}{\frac{1}{\sum_{k=1}^{n_i} l_{ik}} + \frac{1}{\sum_{k=1}^{n_j} l_{jk}}} \quad (2)$$

where n_j are all possible nodes which j can swap to, n_i' is the set of nodes i can swap to after the proposal and n_j' respectively.

Fixed Nodeheight Prune and Regraft The FNPR prunes a subtree and reattaches it without changing any node heights. Hence, the FNPR is a subset of the Wilson-Balding operator. It selects randomly a node i which is not the root and not a leaf. Then the subtree rooted at i is pruned. A second node j with its parent node jP is chosen, where j has to be younger and jP older than i ($\theta_j < \theta_i < \theta_{jP}$). Next, the previously pruned subtree is reattached at the branch above j (see Fig. 2).
For the FNPR we do not need to define the Hastings ratio since this operator is symmetric.

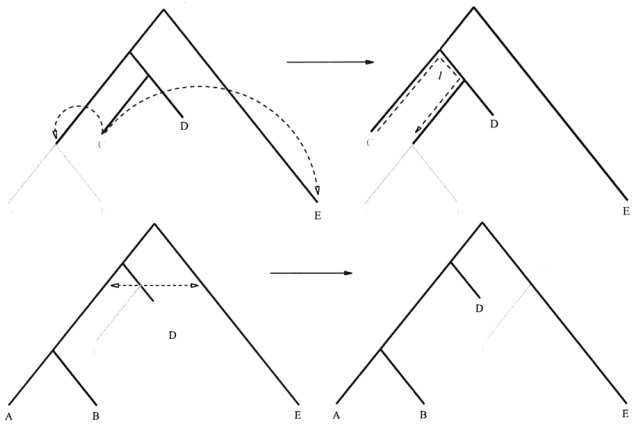

Fig. 2. Top: A proposal of the Intermediate Exchange operator. The left figure shows the possible nodes to swap if node C was chosen first. The second figure shows the tree after the swap and the path length l. Bottom: A proposal of the FNPR operator. The left side shows the tree before the topology change and the right side afterwards.

III. Measuring efficiency of Tree Proposal Operators

A. Convergence Diagnostics of MCMC runs

In Bayesian MCMC for phylogenetic inference the target distribution is the posterior probability distribution on trees. After an unknown amount of time, the distribution of the samples of the Markov chain corresponds, within a small error, to the target distribution. Estimating the performance of an MCMC algorithm is a difficult problem and it can be reformulated as the question: "How long does the MCMC algorithm takes to provide an estimate of the posterior within a certain error tolerance?". There are a number of methods for checking the convergence of an MCMC run, see for example [17] for a discussion about different convergence diagnostics. In this paper the convergence diagnostic of MCMC runs relies on a similarity measurement between the sampled distribution and the target distribution. However, instead of comparing the posterior distribution of trees directly, the marginal probability distribution on subtrees is used [8][16]. Each subtree is separated from the remaining tree by a particular branch. In an unrooted tree, this branch defines a bipartition (or split) of the taxa while in rooted trees it defines a clade, *i.e.* a particular subset of the species, without regard for the relationships within the subset. The marginal probability distribution of clades is a good proxy for the full distribution on trees since each rooted tree is defined by a compatible set of clades. We call C the set of possible clades. The advantage of using the clade probabilities instead of the posterior distribution on tree topologies is that the number of possible clades is much smaller than the number of trees. Hence, fewer samples are needed to obtain accurate estimates of clade probabilities. Moreover in evolutionary biology, the clade probabilities are often the statistics of interest. For each clade c, it is possible to compute the absolute difference between the clade frequency s_c in the sampled distribution and the clade frequency t_c in the target distribution. We call $\delta = \max_{c \in C}(|s_c - t_c|)$, the maximum deviation of the clade frequencies, *cf.* Fig. 3a. It should be noted that in practice the "true" target distribution of clades is unknown and in order to estimate this distribution a classical approach is to perform a set of extensive MCMC runs, the so-called "golden runs". Each "golden run" is actually an extremely long MCMC chain and therefore it produces samples which distribution reflects as accurately as possible the true target distribution.

Once the target distribution of clades is accurately estimated, it is then possible to monitor the convergence of any given MCMC run by computing the evolution of the maximum deviation δ after a number of samples are produced. Based on δ, we propose three different metrics to evaluate the efficiency of operators: firstly, for a particular operator, the percentage of MCMC runs that achieve a given accuracy, *i.e.* for which δ falls bellow a threshold ϵ, secondly, the convergence time, *i.e.* the average number of steps needed to reach ϵ for the first time[1] and lastly the average maximum deviation $\bar{\delta}$ after a limited number of steps. In Fig. 3b the convergence time and the average maximum deviation $\bar{\delta}$ are illustrated with ϵ equals to one percent and 10 millions of steps.

B. Datasets and target distributions

Three small synthetic datasets have been simulated. Three phylogenetic trees were constructed with respectively 6, 8 and 10 species according to a Yule process with a birth rate of one [3]. Then, given those tree topologies and branch lengths, we generated a random DNA sequence and applied mutations by simulating the process of evolution using the software Seq-Gen [18]. We fixed the sequences length to 40 nucleotides, the average number mutations per site to 0.05 and the F81 substitution model [19] was used. We first verified that every operator was able to sample the

[1]All operators have a similar computational time.

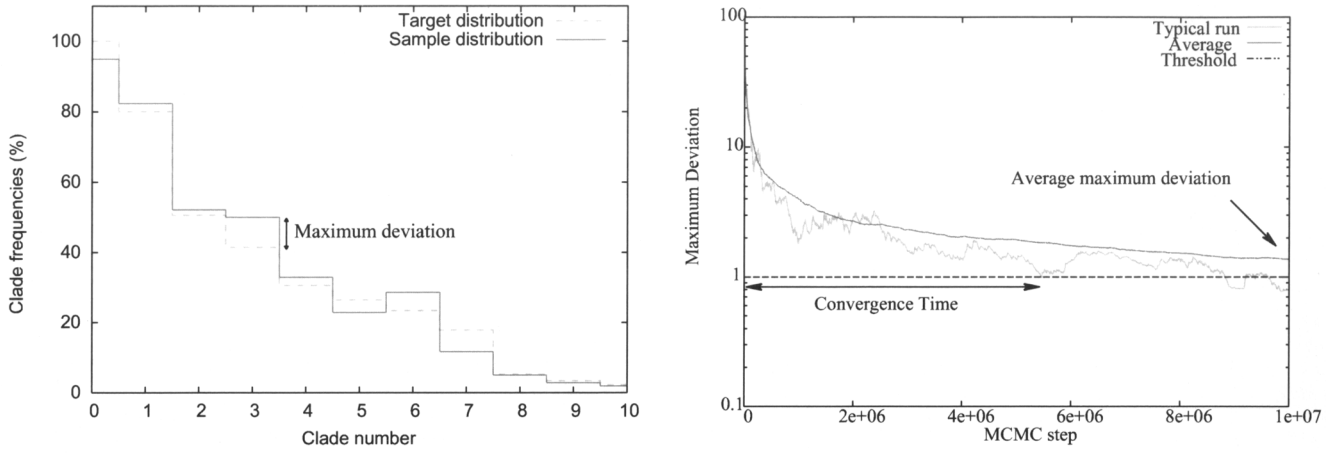

Fig. 3. Part a: An example of sampled and target distribution of clades where the maximum deviation is obtained for clade number 3. Part b: Evolution of the maximum deviation for a typical run and of the average maximum deviation $\bar{\delta}$ over 100 runs. From this analysis, the number of converged runs, the number of iterations needed to converge and the average maximum deviation of clade frequencies can be extracted.

TABLE I

WEIGHTS OF THE TWO DIFFERENT MIXTURE OF OPERATORS. THE NUMBER OF TIME EACH OPERATOR IS APPLIED IS PROPORTIONAL TO THE WEIGHT.

Mixture	**Narrow Exchange**	**NNI**	**SubtreeSlide**	**Intermediate Exchange**	**Wide Exchange**	**FNPR**	**Wilson-Balding**
Original (Default Settings)	15	-	15	-	3	-	3
All (New Proposed Setting)	10	10	10	5	3	3	3

whole set of 905 possible tree topologies on the 6 species dataset. We then performed ten full MCMC analyses on each synthetic dataset using the software BEAST [3] with the default mixture of operators (*cf.* Table I) and predefined standard settings. We verified that the original simulated true tree was present in the 95% credible set of trees. Finally we compared the ten sampled clade distributions and found that they were no more than plus or minus 0.02% different. We claim that these analysis are sufficient to prove the correctness of our implementations and validate the "golden run" approach. Another approach to obtain the posterior distribution is to calculate the marginal likelihood for every tree topology. The ratio of the marginal likelihoods is expected to be the same as the ratio of the samples per topology. To confirm our approach with this one we calculated the marginal likelihoods of the smallest dataset with 905 different topologies. The results agree with our claim but the details are beyond the scope of this paper.

For a more realistic evaluation, we used 11 datasets containing real data, *cf.* [16]. The datasets range from 27 to 71 species and from 378 to 2520 nucleotides, see Table II. For each dataset, ten very long runs, with the predefined standard settings of BEAST were performed. Each run had 100 million iterations with samples every 1000 iterations, which is by far longer than the usual analysis. We observed that the maximal estimated error was below 0.8% (as shown in Table II). It is not expected that the estimated error increases proportional to the size of the dataset. Furthermore, the distribution of trees with high posterior probability has a large impact on the error. The consensus of the ten runs were taken as the "golden runs" which gave the target distributions in our evaluations.

C. Experiments

For each dataset and for each of the seven operators described, we performed 100 runs of 10 million iterations starting from different random trees. In each run a sample tree was taken every 100 steps and every operator was executed the same number of times. In addition, we evaluated the performance of a mixture of four operators predefined in BEAST and a new mixture including all seven operators together (see Table I). We created the new mixture according to the properties of the operators keeping the balance between the local and global operators but without further optimization. We note that for all experiments, we used an additional scaling operator to change the node heights. Indeed most of the operators do not alter the node heights, so they can not sample the whole state space when used alone. To assess the convergence of MCMC runs, the threshold ϵ was set to 1% for the maximum deviation of clade frequencies δ.

The three metrics introduced in section III-A were used to evaluate the performance of the seven operators and of the two mixtures. In Fig. 4, we give the counts of the number of converged runs ($\delta < 1\%$) before 10 million iterations. We see that the two mixtures report the best results. More precisely,

TABLE II

DETAILS OF THE THREE SIMULATED DATASETS AND OF THE ELEVEN REAL DATASETS USED IN THIS STUDY.

Dataset	Number of Species	Number of Nucleotides	Type of Data	TreeBase	Estimated Error (in %)
S 1	6	40	Simulated dataset with 6 Species	-	0.01
S 2	8	40	Simulated dataset with 8 Species	-	0.02
S 3	10	40	Simulated dataset with 10 Species	-	0.02
DS 1	27	1949	rRNA, 18s	M336	0.44
DS 2	29	2520	rDNA, 18s	M501	0.06
DS 3	36	1812	mtDNA, COII (1 -678),cytb (679 -1812)	M1510	0.16
DS 4	41	1137	rDNA, 18s	M1366	0.21
DS 5	50	378	Nuclear protein coding, wingless	M3475	0.77
DS 6	50	1133	rDNA, 18s	M1044	0.17
DS 7	59	1824	mtDNA, COII and cytb	M1809	0.05
DS 8	64	1008	rDNA, 28s	M755	0.10
DS 9	67	955	Plastid ribosomal protein, s16 (rps16)	M1748	0.18
DS 10	67	1098	rDNA, 18s	M520	0.74
DS 11	71	1082	rDNA, internal transcibed spacer (ITS)	M767	0.25

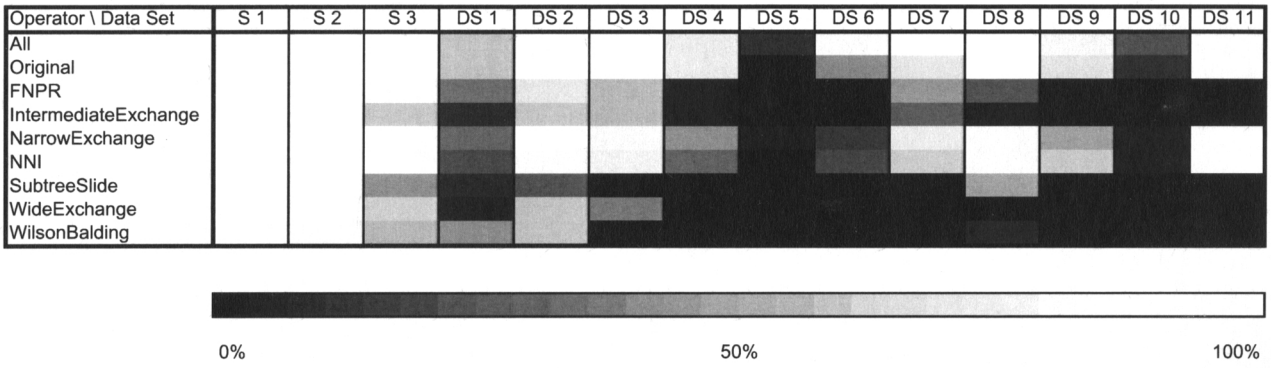

Fig. 4. Percentage of converged runs ($\delta < 1\%$) starting with random trees and for a maximum of 10 million iterations. Constitutions of the mixtures "All" and "Original" are given in Table I.

for datasets DS3, DS6, DS7 and DS8, a binomial test of the percentage of converged runs has shown that the new mixture performs significantly better than the original one (*i.e.* the mixture "All" has 83% converged runs for DS6 whereas the mixture "Original" has only 42%). The two local operators, *i.e.* NNI and Narrow Exchange, are the only operators able to report comparably good results when used alone, while the remaining five operators demonstrate comparatively poor results. We note that no experiments successfully tackled the datasets DS1, DS5 and DS10. This result was actually expected since for these 3 datasets a high estimated error was found (*cf.* Table II) and so the target clade frequencies were not estimated properly. Further studies of these three difficult datasets have shown that they contain islands of trees with high posterior probability. The islands are separated by several NNI transformations. Therefore the MCMC runs spent more time in part of the tree space until they move to another part. Hence, runs with the same number over iterations are varying more in their results and the estimated error is larger.

In Fig. 5, the number of iterations needed to reach the threshold accuracy is used as a measure of performance on the simulated dataset S3 where S3 was arbitrary chosen. The boxes show the central 50% and the whiskers the maximum and minimum values. The operators are ranked according to their performance on this dataset. We performed this test for all other datasets where just minor variations in the performance of the operators was observed. The results here are in accordance with the percentage of converged runs.

The evolution of $\bar{\delta}$ on dataset DS8 is plotted in Fig. 6. While the two mixtures display a fast and accurate convergence behaviour, the two local operators, NNI and Narrow Exchange are the worse operators according to this metric. Since we have seen in Figure 4 that the number of converged runs is reasonably high on this dataset for these operators, we have suspected that a reduced number of runs have found extremely bad estimates of the clade frequencies. This assumption was confirmed by looking at individual runs.

IV. DISCUSSION

Two new operators for clock-constrained trees have been introduced and evaluated. The FNPR is a global prune and regraft operator and therefore it should be compared with the Wilson-Balding operator. The performance of these two operators are indeed similar but in terms of number of converged runs the FNPR outperforms the Wilson-Balding operator. The Intermediate Exchange operator is comparable

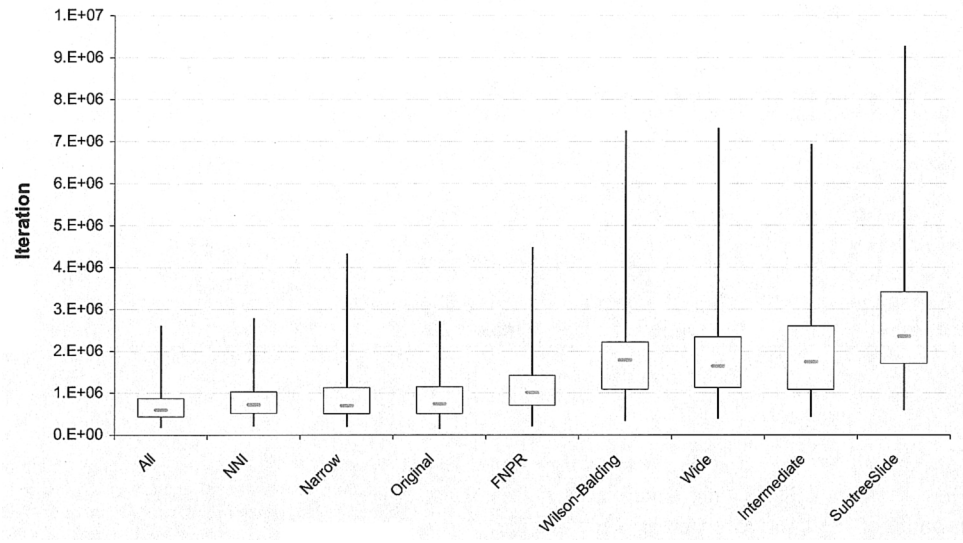

Fig. 5. The average, central 50%, min and max number of iterations needed to reach the threshold of 1% on the dataset S3.

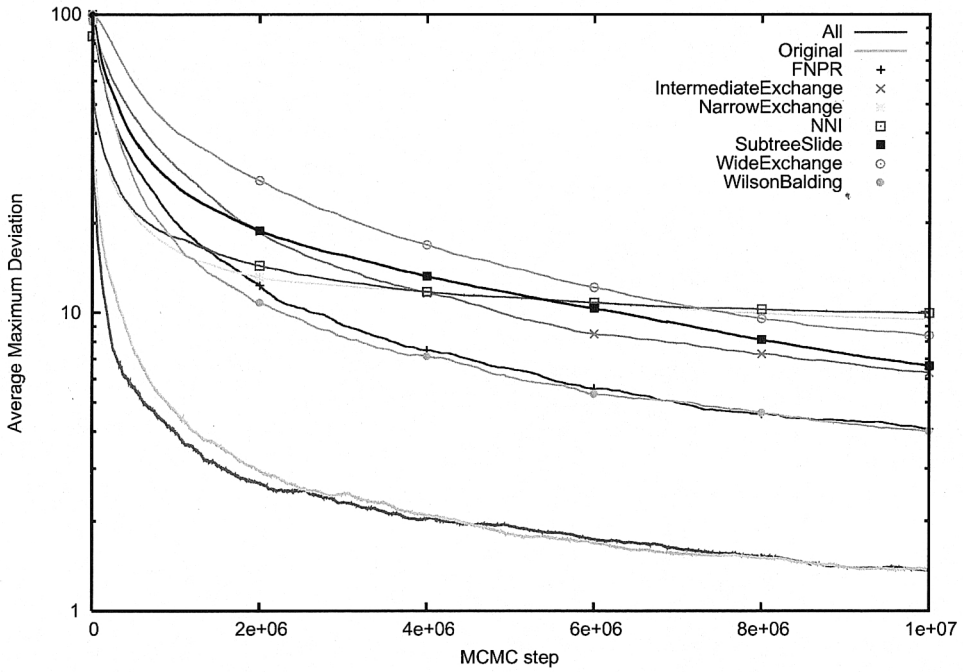

Fig. 6. Evolution of the average maximum deviation $\bar{\delta}$ from the target clade frequencies on dataset DS8.

to the Wide Exchange but is much more robust in terms of average maximum deviation. A summary of the operators is provided in Table III. Local operators are generally faster in computation and have a higher acceptance rate but they are more likely to get stuck in a local part of the tree space. Nevertheless, in our study single local operators have a higher mixing efficiency than single global operators and there is no dataset which clearly uncovers a deficiency in the best local operators. The comparison of the prune and regraft operators with the subtree swap operators has shown that the prune and regraft operators should be favoured. Another characteristic for the operators is whether they change the node heights or not. The FNPR, which does not change the node heights, performs better than the Wilson-Balding operator, which changes the node heights and tree topology simultaneously. Hence, separating the topology change and the branch length change seems to achieve better efficiency.

We conclude from our study that none of the operators performs reliably when used alone. A mixture of several operators proved to be more robust and provides better

TABLE III
CHARACTERISTICS OF THE SEVEN OPERATORS.

Operator	Changes on Topology	Affect on Tree	Branch Length	Proposal Technique	Rank
Narrow Exchange	Direct	Local	Preserved	Subtree Swap *	1
NNI	Direct	Local	Changed	Subtree Swap *	1
FNPR	Direct	Global	Preserved	Prune and Regraft	3
Wilson-Balding	Direct	Global	Changed	Prune and Regraft	4
Intermediate Exchange	Direct	Global	Preserved	Subtree Swap	5
Wide Exchange	Direct	Global	Preserved	Subtree Swap	6
Subtree-Slide	Side Effect	mostly Local	Changed	Branch Change	7

* in this case equal to a Prune and Regraft

performance than any single operator. Further, adding our newly developed operators, FNPR and Intermediate Exchange, to the mixture of the operators increases the average performance, as evaluated by the data sets we analyzed. We therefore hypothesize that these mixed proposal distributions are less likely to fail on an unknown dataset. However finding the most efficient mixture, in terms of how much weight each operator should have, will need further investigation.

V. Conclusion and Future Work

Our results show that the conclusions of [16] on unrooted trees can not be directly extended to clock-constrained trees. Since none of the available software packages have been analysed in a comparable way before, there is a lack of systematic evaluation of clock-constrained tree operators for phylogenetic inference. Therefore, this work should be seen as an initial step towards a systematic validation and evaluation of such operators. In this paper we have already demonstrated an improvement in the mixing efficiency of the software BEAST, and we believe that more generally this work will also improve our understanding of the convergence of the phylogenetic MCMC algorithm in practice.

The operators we used are unguided by the data. They have the advantage that they are very fast and easy to compute. But they do not use knowledge of the data to inform their choice of newly proposed trees. In practice, they all propose low-probability trees more often than high-probability ones and the average quality of the proposals, especially for operators with a large neighbourhood like the Wide Exchange, is very low. This problem can be overcome using Gibbs Sampling [20] and Data Driven MCMC [21]. A short pre-sampling step could be used to obtain clade probabilities which then could give weights for each possible proposal. We plan on investigating such alternatives within the framework developed in this paper.

References

[1] J. Huelsenbeck, F. Ronquist, R. Nielsen, and J. Bollback, "Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology," *Science*, vol. 294, p. 2310, 2001.

[2] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[3] A. Drummond and A. Rambaut, "BEAST: Bayesian evolutionary analysis sampling trees," *BMC Evolutionary Biology*, vol. 7, p. 214, 2007.

[4] I. Wilson, M. Weale, and D. Balding, "Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 166, pp. 155–188, 2003.

[5] M. K. Kuhner, "LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters," *Bioinformatics*, vol. 22, pp. 768–770, 2006.

[6] P. Beerli and J. Felsenstein, "Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach," *Proceedings of the National Academy of Sciences*, p. 081068098, 2001.

[7] A. J. Drummond, G. K. Nicholls, A. G. Rodrigo, and W. Solomon, "Estimating Mutation Parameters, Population History and Genealogy Simultaneously From Temporally Spaced Sequence Data," *Genetics*, vol. 161, pp. 1307–1320, 2002.

[8] S. Li, D. K. Pearl, and H. Doss, "Phylogenetic tree construction using markov chain monte carlo," *Journal of the American Statistical Association*, vol. 95, pp. 493–508, 2000.

[9] I. J. Wilson and D. J. Balding, "Genealogical Inference From Microsatellite Data," *Genetics*, vol. 150, pp. 499–510, 1998.

[10] B. Mau and M. A. Newton, "Phylogenetic inference for binary data on dendograms using markov chain monte carlo," *Journal of Computational and Graphical Statistics*, vol. 6, pp. 122–131, 1997.

[11] B. Mau, M. A. Newton, and B. Larget, "Bayesian phylogenetic inference via markov chain monte carlo methods," *Biometrics*, vol. 55, pp. 1–12, 1999.

[12] B. Larget and D. Simon, "Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees," *Mol. Biol. Evol*, vol. 16, pp. 750–759, 1999.

[13] Z. Yang, "PAML: a program package for phylogenetic analysis by maximum likelihood," *Comput. Appl. Biosci.*, vol. 13, pp. 555–556, 1997.

[14] J. Huelsenbeck and F. Ronquist, "MRBAYES: Bayesian inference of phylogenetic trees," *Bioinformatics*, vol. 17, pp. 754–755, 2001.

[15] F. Ronquist and J. P. Huelsenbeck, "MrBayes 3: Bayesian phylogenetic inference under mixed models," *Bioinformatics*, vol. 19, pp. 1572–1574, 2003.

[16] C. Lakner, P. van der Mark, J. P. Huelsenbeck, B. Larget, and F. Ronquist, "Efficiency of markov chain monte carlo tree proposals in bayesian phylogenetics," *Sys. Biol.*, vol. 57, pp. 86–103, 2008.

[17] M. K. Cowles and B. P. Carlin, "Markov chain monte carlo convergence diagnostics: A comparative review," *Journal of the American Statistical Association*, vol. 91, pp. 883–904, 1996.

[18] A. Rambaut and N. C. Grass, "Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees," *Comput. Appl. Biosci.*, vol. 13, pp. 235–238, 1997.

[19] J. Felsenstein, "Evolutionary trees from DNA sequences: A maximum likelihood approach," *Journal of Molecular Evolution*, vol. 17, no. 6, pp. 368–376, 1981.

[20] S. Geman, D. Geman, K. Abend, T. Harley, and L. Kanal, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*," *Journal of Applied Statistics*, vol. 20, pp. 25–62, 1993.

[21] E. Clark and A. Quinn, "A data-driven bayesian sampling scheme for unsupervised image segmentation," *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, vol. 6, pp. 3497–3500 vol.6, 1999.