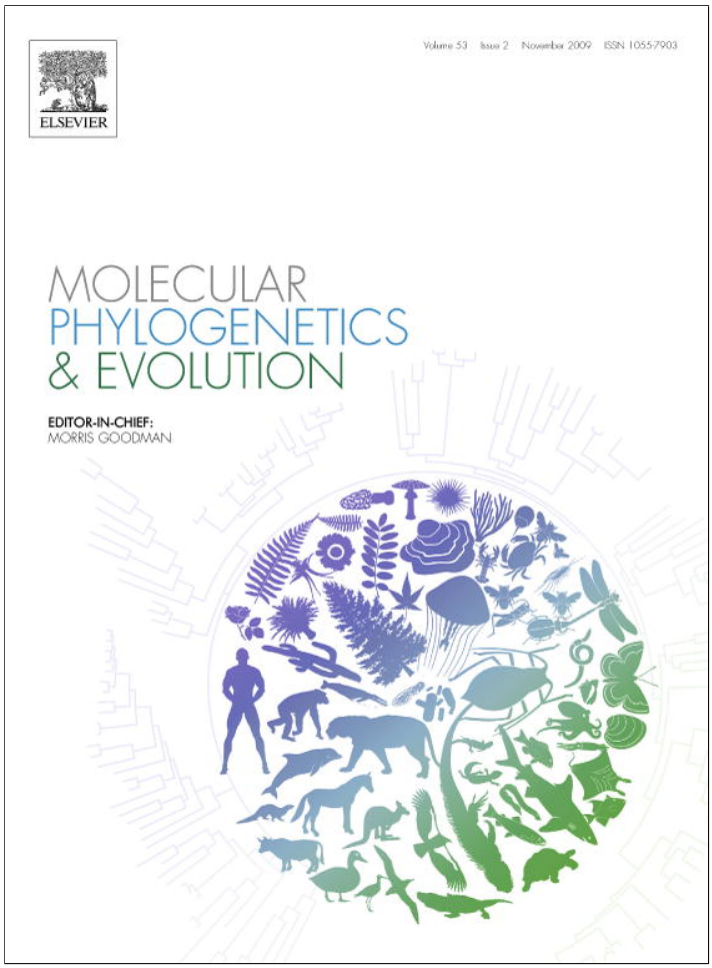

Molecular Phylogenetics and Evolution 53 (2009) 394–403

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Bayesian phylogeny analysis via stochastic approximation Monte Carlo

Sooyoung Cheon [a], Faming Liang [b,*]

[a] KU Industry-Academy Cooperation Group Team of Economics and Statistics, Korea University, Jochiwon 339-700, South Korea
[b] Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA

ARTICLE INFO

Article history:
Received 31 August 2008
Revised 15 June 2009
Accepted 30 June 2009
Available online 7 July 2009

Keywords:
Bayesian phylogeny analysis
Consensus tree
Markov chain Monte Carlo
Stochastic approximation Monte Carlo

ABSTRACT

Monte Carlo methods have received much attention in the recent literature of phylogeny analysis. However, the conventional Markov chain Monte Carlo algorithms, such as the Metropolis–Hastings algorithm, tend to get trapped in a local mode in simulating from the posterior distribution of phylogenetic trees, rendering the inference ineffective. In this paper, we apply an advanced Monte Carlo algorithm, the stochastic approximation Monte Carlo algorithm, to Bayesian phylogeny analysis. Our method is compared with two popular Bayesian phylogeny software, BAMBE and MrBayes, on simulated and real datasets. The numerical results indicate that our method outperforms BAMBE and MrBayes. Among the three methods, SAMC produces the consensus trees which have the highest similarity to the true trees, and the model parameter estimates which have the smallest mean square errors, but costs the least CPU time.

1. Introduction

Phylogenetic inference is one of fundamental topics in molecular evolution. The traditional methods select a single "best" tree, either by the neighbor joining (NJ) method (Saitou and Nei, 1987) or according to some optimality criterion, such as minimum evolution (Kidd and Sgaramella-Zonta, 1971; Rzhetsky and Nei, 1992), maximum parsimony (Fitch, 1971; Maddison, 1991), and maximum likelihood (Felsenstein, 1981, 1993; Kishino et al., 1990; Salter and Pearl, 2001). The neighbor joining method is a distance-based clustering method, which constructs phylogenetic trees by successively pairing the taxa with the smallest distance between sequences. The maximum evolution, maximum parsimony, maximum likelihood methods are in practice always combined with a search algorithm looking for the "best" tree. The maximum evolution method seeks for the tree with the smallest sum of branch lengths, the maximum parsimony methods seeks for the tree that requires the minimum number of mutations for reproducing the data, and the maximum likelihood method seeks for the tree that is most likely to have occurred given the observed data and the model of evolution. Although the traditional methods work well for many problems, they do not produce valid inferences beyond point estimates. Although uncertainty of the phylogeny estimation can be measured by the bootstrap resampling method (Felsenstein, 1985; Newton, 1996), the computational complexity of bootstrapping has constrained the applications of the method to very small problems.

Bayesian methods have received much attention in the recent literature of phylogeny analysis. Significant early work include Mau (1996), Rannala and Yang (1996), Yang and Rannala (1997), Mau and Newton (1997), Mau et al. (1999), Larget and Simon (1999), Newton et al. (1999), and Li et al. (2000), where the Metropolis–Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) is often employed to simulate from a posterior distribution defined on the parameter space that includes tree topologies as well as branch lengths and the parameters of the sequence evolutionary model. It is known that Bayesian phylogeny inference can depend on the prior distributions used in the analysis. Refer to Alfaro and Holder (2006) for discussions on how to set prior distributions for Bayesian phylogeny analysis.

Bayesian methods have several characteristics different from the traditional methods. Firstly, it automatically accounts for the uncertainty embedded in the structure of phylogenetic trees and the model parameter estimates. For example, it can provide a posterior probability that a particular tree or a portion of the tree represents the true phylogeny, and this in turn supplies with evolutionary biologists reasonable weighting of trees for further inference. The review of Huelsenbeck et al. (2001) puts more emphasis on this point. Secondly, Bayesian methods make analysis of large datasets more tractable. The samples simulated from the posterior distribution can be used to construct consensus trees, which can be much faster than the bootstrap resampling method (Larget and Simon, 1999). However, as pointed out by Suzuki et al. (2002), Bayesian methods may give over-credibility to the trees inferred.

* Corresponding author. Fax: +1 979 845 3144.
E-mail addresses: stat-csy@hanmail.net (S. Cheon), fliang@stat.tamu.edu (F. Liang).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.06.019

Another difficulty with Bayesian methods is the lack of efficient posterior samplers. Conventional MCMC algorithms, e.g., the MH algorithm used by most Bayesian phylogeny software, tend to get trapped in a local energy minimum, rendering ineffective inference for the phylogeny. In this paper, the energy function refers to the negative log-posterior distribution function of the phylogeny, and the local energy minimum refers to a local mode of the posterior distribution. We note that the local-trap problem has also occurred in applications of the optimization-based traditional methods, such as the maximum likelihood and maximum evolution methods. A number of authors have been aware of this difficulty and have tried to employ some advanced MCMC algorithms to resolve it. For example, Huelsenbeck and Ronquist (2001) and Altekar et al. (2004) employed parallel tempering (Geyer, 1991) and Feng et al. (2003) employed the multiple-try Metropolis algorithm (Liu et al., 2000). Cheon and Liang (2008) applied a sequential Monte Carlo algorithm to the problem, but focusing on the maximum *a posteriori* (MAP) trees. These advanced Monte Carlo algorithms can usually converge faster than the MH algorithm for simulating from the distributions for which the energy landscape is rugged. A recent work by Lakner et al. (2008) provides an analysis of efficiency of several MCMC moves in the tree space and suggests the use of a mixture of the moves that slightly perturb the tree and the moves that make drastic changes to the topology.

In this paper, we apply the stochastic approximation Monte Carlo (SAMC) algorithm (Liang et al., 2007) to this problem. The SAMC algorithm has two nice features. Firstly, it possesses the self-adjusting mechanism and thus avoids essentially the local-trap problem suffered by other MCMC algorithms, e.g., the MH algorithm used in BAMBE (Larget and Simon, 1999) and the parallel tempering algorithm used in MrBayes (Huelsenbeck and Ronquist, 2001), in simulating from the posterior distribution of phylogenetic trees. Secondly, it falls into the category of dynamic importance sampling algorithms; the phylogeny and the model parameters can be inferred by weightedly averaging over the samples generated in simulations. The new method is compared with two popular Bayesian phylogeny software, BAMBE and MrBayes, on simulated and real datasets. The numerical results indicate that SAMC outperforms BAMBE and MrBayes for Bayesian phylogeny analysis. Among the three methods, SAMC produced the consensus trees which are most similar to the true trees, and the model parameter estimates which have the smallest mean square errors, but cost the least CPU time.

The remainder of this paper is organized as follows. In Section 2, we provide a brief description of Bayesian phylogeny analysis. In Section 3, we first give a brief review of the SAMC algorithm and then describe its implementation for Bayesian phylogeny inference. In Section 4, we present the numerical results on simulated and real data examples. In Section 5, we conclude the paper with a brief discussion.

2. Bayesian phylogeny analysis

A phylogenetic tree can be represented as a rooted binary tree, each node with descendants representing the most recent common ancestor of the descendants, and the root representing the most common ancestor of all the entities at the leaves of the tree. In general, a phylogenetic tree of n leaves has $n-2$ internal nodes (excluding the root node) and $2n-2$ branches. The length of branch represents the distance between two end node sequences, and it is often calculated from a model of substitution of residues over the course of evolution.

Suppose that we are interested in conducting a phylogeny analysis for n nucleotide sequences (taxa). The problem for protein sequences is similar. The nucleotide sequences can be arranged as a n by N matrix, where N is the common number of sites or the common length of the sequences. By assuming that evolution among sites is independent conditioned on the given genealogy, modeling is reduced to a single site. Although this assumption greatly simplifies the likelihood calculation, it is quite probably violated by most coding sequence datasets. As explained by Galtier et al. (2005), the sites in a protein (or a RNA sequence) interacts to determine the selected three-dimensional structure of the molecular, so the evolutionary process of interacting sites are not independent. Attempts have been made in the literature to relax the independence assumption by either introducing an autocorrelation parameter for the evolutionary rates of neighboring sites or directly modeling the joint evolutionary process of any two neighboring sites. Significant work in this respect include Yang (1995), Felsenstein and Churchill (1996), Thorne et al. (1996), Pollock et al. (1999), Duret and Galtier (2000), and Robinson et al. (2003). As our goal is to introduce an advanced Monte Carlo method for Bayesian phylogeny analysis, the independence assumption is still used in this paper for a demonstration purpose.

Under the independence assumption, several evolutionary models have been proposed for nucleotides, such as the one-parameter model (Jukes and Cantor, 1969), two-parameter model (Kimura, 1980), Felsenstein model (1981), and HKY85 model (Hasegawa et al., 1985). Among the four models, the HKY model is most flexible, which possesses a general stationary distribution of nucleotides and allows for different rates of transitional and transversional events. In this paper, we consider the HKY85 model, for which the elements of the transition probability matrix are given by

$$Q_{j|i}(h)=\begin{cases}\pi_j+\pi_j\left(\frac{1}{\lambda_j}-1\right)e^{-\alpha h}+\left(\frac{\lambda_j-\pi_j}{\lambda_j}\right)e^{-\alpha\gamma_j h} & \text{if } i=j,\\ \pi_j+\pi_j\left(\frac{1}{\lambda_j}-1\right)e^{-\alpha h}-\left(\frac{\pi_j}{\lambda_j}\right)e^{-\alpha\gamma_j h} & \text{if } i\neq j \text{ (transitional event)},\\ \pi_j(1-e^{-\alpha h}) & \text{if } i\neq j \text{ (transversional event)},\end{cases} \tag{1}$$

where h denotes the evolution time or the branch length of the phylogenetic tree, α denotes the evolutionary rate, $\lambda_j=\pi_A+\pi_G$ if base j is a purine (A or G) and $\pi_C+\pi_T$ if base j is a pyrimidine (C or T), $\gamma_j=1+(\kappa-1)\lambda_j$, and κ is a parameter responsible for distinguishing between transitions and transversions. The stationary probabilities of the four nucleotides are π_A,π_C,π_G, and π_T, respectively. The HKY85 model includes five free parameters, namely, $\alpha,\kappa,\pi_A,\pi_C$ and π_G, which satisfy the constraints $\alpha>0,\kappa>0,0<\pi_A,\pi_C,\pi_G<1$, and $0<\pi_A+\pi_C+\pi_G<1$.

Let $\omega=(\tau,\boldsymbol{h},\boldsymbol{\phi})$ denote a phylogenetic tree, where τ denotes the tree topology, $\boldsymbol{h}$ denotes the vector of branch lengths, and $\boldsymbol{\phi}$ denotes the vector of parameters of the evolutionary model. The likelihood can be calculated using the pruning method proposed by Felsenstein (1981). The pruning method produces a collection of partial likelihoods of subtrees, starting from the leaves and working recursively to the root for each site. Let $\mathscr{S}=\{A,C,G,T\}$ denote the set of nucleotides. For site k of a leaf e, define $L_e^k(i)=1$ if state i matches the base found in the sequence and 0 otherwise, where i indexes the elements of $\mathscr{S}$. At site k of an internal node v, the conditional probability of descendant data given state i is

$$L_v^k(i)=\left(\sum_{j\in\mathscr{S}}L_u^k(j)Q_{j|i}(h_{vu})\right)\times\left(\sum_{j\in\mathscr{S}}L_w^k(j)Q_{j|i}(h_{vw})\right),\quad i\in\mathscr{S},$$

where u and w denote the two children nodes of v, and h_{ab} denotes the length of the branch ended with the nodes a and b. The likelihood of the tree can then be written as

$$L(\omega|D)=\prod_{k=1}^{N}\sum_{i\in\mathscr{S}}\pi_0(i)L_\rho^k(i), \tag{2}$$

where D denotes the observed sequences of n taxa, ρ denotes the root node, and π_0 is the initial probability distribution assigned to the ancestral root sequence. In all simulations of this paper, π_0 is set to the observed frequency of the nucleotides of the given sequences.

Let $f(\omega)$ denote the prior distribution of ω. The posterior distribution of the phylogenetic tree can then be formed as

$$f(\omega|D) \propto L(\omega|D)f(\omega). \tag{3}$$

Various samplers can then be employed to sample from this posterior. For example, BAMBE employed the MH algorithm, and MrBayes employed the parallel tempering algorithm.

In this paper, we follow Mau et al. (1999) and Larget and Simon (1999) to place a uniform prior on ω. This can be understood that we place a uniform prior on both the parameter space of the HKY model and the joint space of tree topology and branch lengths induced by the HKY model; that is, we set $f(\boldsymbol{\phi}) \propto 1$ and $f(\boldsymbol{\tau}, \boldsymbol{h}|\boldsymbol{\phi}) \propto 1$. Under this setting, the MAP tree coincides with the maximum likelihood tree.

3. Bayesian phylogeny analysis via the SAMC algorithm

3.1. A review of the SAMC algorithm

Suppose that we are working on inference for a posterior distribution,

$$f(x) = \frac{1}{Z}\psi(x), \quad x \in \mathcal{X}, \tag{4}$$

where $\mathcal{X}$ is the sample space, and Z is the normalizing constant. In the context of Bayesian phylogeny analysis, $\psi(\cdot)$ corresponds to the unnormalized posterior density $L(\omega|D)f(\omega)$ as specified in (3), and $\mathcal{X}$ corresponds to the sample space of ω. Let $U(x) = -\log\psi(x)$, which is called the energy function in terms of physics. Suppose that the sample space has been partitioned according to the energy function into m disjoint subregions denoted by $E_1 = \{x : U(x) < u_1\}, E_2 = \{x : u_1 \leqslant U(x) < u_2\}, \ldots, E_{m-1} = \{x : u_{m-2} \leqslant U(x) < u_{m-1}\}$, and $E_m = \{x : U(x) \geqslant u_{m-1}\}$, where $u_1, \ldots, u_{m-1}$ are real numbers specified by the user. Issues on how to partition the sample space or how to specify the numbers $u_1, \ldots, u_{m-1}$ will be further discussed at the end of Section 3.2.

SAMC seeks to draw samples from each of the subregions with a pre-specified frequency. Let $x^{(t+1)}$ denote a sample drawn from a MH kernel $K_{\theta^{(t)}}(x^{(t)}, \cdot)$ with the proposal distribution $q(x^{(t)}, \cdot)$ and the stationary distribution

$$f_{\theta^{(t)}}(x) \propto \sum_{i=1}^{m-1} \frac{\psi(x)}{e^{\theta_i^{(t)}}} I(x \in E_i) + \psi(x) I(x \in E_m), \tag{5}$$

where $\theta^{(t)} = (\theta_1^{(t)}, \ldots, \theta_{m-1}^{(t)})$ is an $(m-1)$-vector in a space Θ. For convenience, we set $\theta_m^{(t)} = 0$. Here, without loss of generality, we assume that E_m is non-empty; that is, $\int_{E_m} \psi(x)dx > 0$. A subregion E_i is called an empty subregion if $\int_{E_i} \psi(x)dx = 0$. In practice, E_m can be replaced by any subregion which is known to be non-empty.

Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$ be an m-vector with $0 < \pi_i < 1$ and $\sum_{i=1}^m \pi_i = 1$, which defines the desired sampling frequencies of the subregions. Henceforth, $\boldsymbol{\pi}$ will be called the desired sampling distribution. Define $H(\theta^{(t)}, x^{(t+1)}) = (\boldsymbol{e}^{(t+1)} - \boldsymbol{\pi})$, where $\boldsymbol{e}^{(t+1)} = (e_1^{(t+1)}, \ldots, e_m^{(t+1)})$ and $e_i^{(t+1)} = 1$ if $x^{(t+1)} \in E_i$ and 0 otherwise. Let $\{\gamma_t\}$ be a positive, non-decreasing sequence satisfying the conditions,

$$(i) \ \sum_{t=0}^{\infty} \gamma_t = \infty, \quad (ii) \ \sum_{t=0}^{\infty} \gamma_t^{\delta} < \infty, \tag{6}$$

for some $\delta \in (1, 2)$. In the context of stochastic approximation (Robbins and Monro, 1951), $\{\gamma_t\}_{t \geqslant 0}$ is called the gain factor sequence.

Let $J(x)$ denote the index of the subregion that the sample x belongs to. Let $\{\mathcal{K}_s, s \geqslant 0\}$ be a sequence of compact subsets of Θ such that

$$\bigcup_{s \geqslant 0} \mathcal{K}_s = \Theta, \quad \text{and} \quad \mathcal{K}_s \subset \text{int}(\mathcal{K}_{s+1}), \quad s \geqslant 0, \tag{7}$$

where $\text{int}(A)$ denotes the interior of set A. Let $\mathcal{X}_0$ be a subset of $\mathcal{X}$, and let $\mathbb{T} : \mathcal{X} \times \Theta \to \mathcal{X}_0 \times \mathcal{K}_0$ be a measurable function which maps a point in $\mathcal{X} \times \Theta$ to a random point in $\mathcal{X}_0 \times \mathcal{K}_0$. Let σ_k denote the number of truncations performed until iteration k. Let $\mathcal{S}$ denote the collection of the indices of the subregions that have been visited by SAMC. With above notations, one iteration of SAMC can be described as follows.

The SAMC algorithm

(a) (Sampling) Simulate a sample $x^{(t+1)}$ by a single MH update with the target distribution as defined in (5).
  (a.1) Generate y according to a proposal distribution $q(x_t, y)$.
  (a.2) Calculate the ratio

$$r = e^{\theta^{(t)}_{J(x^{(t)})} - \theta^{(t)}_{J(y)}} \frac{\psi(y)q(y, x^{(t)})}{\psi(x^{(t)})q(x^{(t)}, y)}. \tag{8}$$

  (a.3) Accept the proposal with probability $\min(1, r)$. If it is accepted, set $x^{(t+1)} = y$; otherwise, set $x^{(t+1)} = x^{(t)}$. If $J(x^{(t+1)}) \notin \mathcal{S}$, set $\mathcal{S} \leftarrow \mathcal{S} + \{J(x^{(t+1)})\}$.

(b) (Weight updating) For all $i \in \mathcal{S}$, set

$$\theta_i^{\left(t+\frac{1}{2}\right)} = \theta_i^{(t)} + a_{t+1}\left(I_{\{x^{(t+1)} \in E_i\}} - \pi_i\right) - a_{t+1}\left(I_{\{x^{(t+1)} \in E_m\}} - \pi_m\right). \tag{9}$$

(c) (Varying truncation) If $\theta^{\left(t+\frac{1}{2}\right)} \in \mathcal{K}_{\sigma_t}$, then set $(\theta^{(t+1)}, x^{(t+1)}) = \left(\theta^{\left(t+\frac{1}{2}\right)}, x^{(t+1)}\right)$ and $\sigma_{t+1} = \sigma_t$; otherwise, set $(\theta^{(t+1)}, x^{(t+1)}) = \mathbb{T}(\theta^{(t)}, x^{(t)})$ and $\sigma_{t+1} = \sigma_t + 1$.

The self-adjusting mechanism of the SAMC algorithm is obvious: If a proposal is rejected, the weight of the subregion that the current sample belongs to will be adjusted to a larger value, and thus the proposal of jumping out from the current subregion will be less likely rejected in the next iteration. This mechanism warrants the algorithm not to be trapped by local energy minima. The SAMC algorithm represents a significant advance in simulations of complex systems for which the energy landscape is rugged.

The proposal distribution $q(x, y)$ used in the MH updates is required to satisfy the following condition: For every $x \in \mathcal{X}$, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that

$$|x - y| \leqslant \epsilon_1 \Rightarrow q(x, y) \geqslant \epsilon_2, \tag{10}$$

where $|x - y|$ denotes the Euclidean distance between x and y. This is a natural condition in study of MCMC theory (Roberts and Tweedie, 1996). In practice, this kind of proposals can be easily designed for both discrete and continuum systems as discussed in Liang et al. (2007).

SAMC falls into the category of varying truncation stochastic approximation algorithms (Chen, 2002; Andrieu et al., 2005). Following Liang et al. (2007), we have the following convergence result:

Under the conditions (6) and (10), for all non-empty subregions,

$$\theta_i^{(t)} \to C + \log\left(\int_{E_i} \psi(x)dx\right) - \log(\pi_i + \pi_0), \tag{11}$$

as $t \to \infty$, where $\pi_0 = \sum_{j \in \{i : E_i = \emptyset\}} \pi_j/(m - m_0), m_0 = \#\{i : E_i = \emptyset\}$ is the number of empty subregions, and $C = -\log(\int_{E_m} \psi(x)dx) + \log(\pi_m + \pi_0)$.

Let $\hat{\pi}_i^{(t)} = P(x^{(t)} \in E_i)$ be the probability of sampling from the subregion E_i at iteration t. Eq. (11) implies that as $t \to \infty$, $\hat{\pi}_i^{(t)}$ will converge to $\pi_i + \pi_0$ if $E_i \neq \emptyset$ and 0 otherwise. With an appropriate specification of π, SAMC sampling can be biased to the low energy subregions to increase the chance of locating the global energy optimizer.

Let $(x^{(1)}, \theta^{(1)}), \ldots, (x^{(n)}, \theta^{(n)})$ denote a set of samples generated by SAMC. Let $y^{(1)}, \ldots, y^{(n')}$ denote the distinct samples among $x^{(1)}, \ldots, x^{(n)}$. Generate a random variable/vector Y such that

$$P(Y = y^{(i)}) = \frac{\sum_{t=1}^{n} e^{\theta^{(t)}_{J(x^{(t)})}} I(x^{(t)} = y^{(i)})}{\sum_{t=1}^{n} e^{\theta^{(t)}_{J(x^{(t)})}}}, \quad i = 1, \ldots, n', \tag{12}$$

where $I(\cdot)$ is the indicator function, and $J(x^{(t)})$ denotes the index of the subregion that the sample $x^{(t)}$ belongs to. Since the number of truncations in the varying truncation algorithm can only occur a finite number of times (Andrieu et al., 2005), $\theta^{(t)}_{J(x^{(t)})}$ can be bounded in a compact set and is thus finite. By calling some results from the literature of non-homogeneous Markov chains, Liang (2009a) showed that the random variable/vector Y generated in (12) is asymptotically distributed as $f(\cdot)$. Note that the samples $(x^{(1)}, \theta^{(1)}), \ldots, (x^{(n)}, \theta^{(n)})$ form a non-homogeneous Markov chain. Therefore, for an integrable function $g(x)$, the expectation $E_f g(x)$ can be estimated by

$$\widehat{E_f g(x)} = \frac{\sum_{t=1}^{n} e^{\theta^{(t)}_{J(x^{(t)})}} g(x^{(t)})}{\sum_{t=1}^{n} e^{\theta^{(t)}_{J(x^{(t)})}}}. \tag{13}$$

As $n \to \infty$, $\widehat{E_f g(x)} \to E_f g(x)$ for the same reason that the usual importance sampling estimate converges (Geweke, 1989).

3.2. Bayesian phylogeny inference

In this subsection, we first describe how to make SAMC moves over the space of feasible phylogenetic trees, and then discuss some practical issues on SAMC implementation.

In this paper, the local moves used in Larget and Simon (1999) was adopted for updating phylogenetic trees. The only difference is that the acceptance probability of those moves has been adjusted by the self-adjusting factor $e^{\theta^{(t)}_{J(x^{(t)})} - \theta^{(t)}_{J(y)}}$ as prescribed in Eq. (8). There are three types of local moves, the moves for model parameter updating, the moves for branch length updating, and the moves for tree topology rearrangement. Larget and Simon (1999) applied those moves to both types of trees with and without molecular clocks. The trees considered in this paper are without molecular clocks. We note that Larget and Simon (1999) also prescribed some global moves for updating phylogenetic trees. Since SAMC is capable of moving across high energy barriers (due to its self-adjusting mechanism) and the global moves are much more time consuming than the local moves, only the local moves were adopted in our simulations. To have the condition (10) satisfied, we constrained the model parameters and branch lengths to a compact set $\mathcal{X} = [-B, B]^d$, where $B = 1.0e+10$ and d is the total number of branches and model parameters. As a practical matter, this is equivalent to setting $\mathcal{X} = \mathbb{R}^m$.

Let $g(\omega)$ denote a quantity of interest for phylogeny analysis, such as the presence/absence of a branch or an evolutionary parameter. It follows from (13) that $E_f g(\omega)$, the expectation of $g(\omega)$ with respect to the posterior (3), can be estimated by

$$\widehat{E_f g(\omega)} = \frac{\sum_{k=n_0+1}^{n} g(\omega_k) e^{\theta^{(k)}_{J(\omega_k)}}}{\sum_{k=n_0+1}^{n} e^{\theta^{(k)}_{J(\omega_k)}}}, \tag{14}$$

where $\left(\omega_{n_0+1}, \theta^{(n_0+1)}_{J(\omega_{n_0+1})}\right), \ldots, \left(\omega_n, \theta^{(n)}_{J(\omega_n)}\right)$ denote the samples generated by SAMC, and n_0 denotes the number of burn-in iterations.

For an effective implementation of SAMC, several issues need to be considered.

- Sample space partitioning. In general, the sample space should be partitioned such that the MH updates within the same subregion have a reasonable acceptance rate. For Bayesian phylogeny analysis, the sample space is usually partitioned according to the energy function. The maximum energy difference in each subregion should be bounded by a reasonable number, say 2; that is $u_{i+1} - u_i \leqslant 2$. This ensures the MH moves within the same subregion to have a reasonable acceptance rate. Note that within the same subregion, the SAMC moves are reduced to the conventional MH moves. Since SAMC allows for the existence of empty subregions, u_1 can be set to a rather small number and u_{m-1} can be set to a rather large number.
- Choice of the desired sampling distribution π. This can be done according to our aims. For example, if one aims at the MAP tree, one may choose π to bias sampling to low energy regions to increase the chance of finding the global energy minima. In this paper, since we are interested in Bayesian phylogeny analysis, we set π to be uniform over all subregions, i.e., $\pi_1 = \cdots = \pi_m = \frac{1}{m}$. Our experience shows that setting π to be uniform often results in more robust estimates for the model parameters.
- Choice of the gain factor sequence and the total number of iterations. To have the condition (6) satisfied, we suggest to choose

$$\gamma_t = \left(\frac{T_0}{\max(T_0, t)}\right)^{\eta}, \quad t = 0, 1, 2, \ldots \tag{15}$$

for pre-specified values of $T_0 > 1$ and $\eta \in (0.5, 1]$. A large value of T_0 will allow the sampler to reach all subregions very quickly even for a large system. The appropriateness of the choice of T_0 and N can be diagnosed by checking the convergence of multiple runs (starting with different points) through an examination for the variation of $\widehat{\theta}$ or $\widehat{\pi}$, where $\widehat{\theta}$ and $\widehat{\pi}$ denote, respectively, the estimates of θ and π obtained at the end of a run. If only a single run was made, a practical guideline for the convergence diagnostic, as suggested by Wang and Landau (2001), is to examine flatness of the histogram of the samples drawn at different subregions. A histogram is said flat if the sampling frequency at each subregion is not less than 80% of the average sampling frequency of the subregions. If the simulation is diagnosed as unconverged, SAMC should be re-run with a larger value of T_0, a larger number of iterations, or both. In this paper, we set $T_0 = 50$ and $\eta = 0.6$ in all simulations.

4. Numerical examples

4.1. Simulated examples

In this study, we illustrate how SAMC can be used for Bayesian phylogeny inference. A total of 20 nucleotide sequences were generated according to a given tree (shown in Fig. 1(a)), a given root sequence (shown in Table 1), and a HKY85 model with parameters $\kappa = 2, \alpha = 1$, and $\pi_A = \pi_G = \pi_C = \pi_T = 0.25$. The length of each sequence is 300. A C-code used for generating the nucleotide sequences is available upon request from the authors.

SAMC was first applied to this example. The sample space was partitioned into 11 subregions, $E_1 = \{\omega : U(\omega) > 1830\}, E_2 = \{\omega : 1830 \geqslant U(\omega) > 1828\}, \cdots, E_{10} = \{\omega : 1814 \geqslant U(\omega) > 1812\}, E_{11} = \{\omega : U(\omega) \leqslant 1812\}$, where $U(\omega) = \log(L(\omega|D)f(\omega))$. SAMC was run 5 times independently, and each run consisted of 2.2×10^5 iterations. Table 2 shows the relative sampling frequency of each of the subregions realized in one run. The relative sampling frequency of subregion i is defined as $N_i/\overline{N} \times 100\%$, where N_i and $\overline{N}$ denote the sampling frequency of subregion i and the average sampling frequency over all non-empty subregions, respectively. Table 2

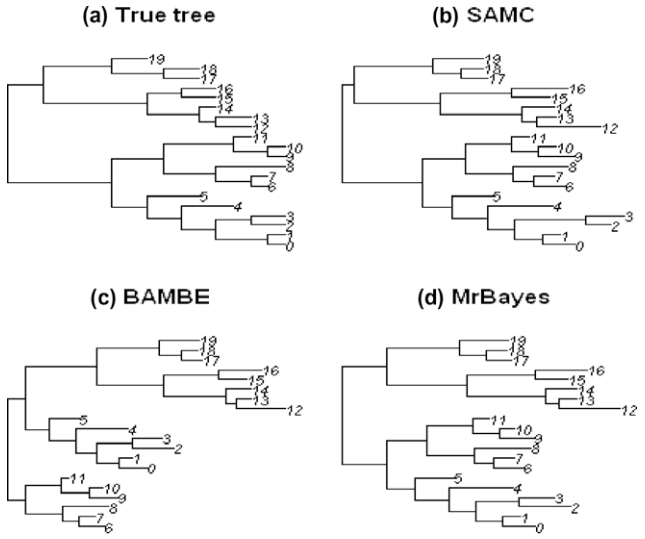

Fig. 1. Comparison of the consensus trees produced by SAMC, BAMBE and MrBayes for the simulated 20-taxon dataset.

Table 1
The root sequence for the simulated example.

AACAAAGCCA CAATTATTAA TACTCTTGCT ACATCCTGAG CAAAAGCCCC
CGCCCTCACA GCTCTCACCC CCCTTATTCT TCTTTCACTA GGGGGCCTCC
CCCCTCTCAC GGGCTTTATA CCAAAATGAC TGATTCTTCA AGAACTAACC
AAACAAGGCC TTGCCCCCAC CGCAACCCTA GCAGCCCTCT CAGCACTCCT
TAGCCTCTAT TTCTACCTGC GCCTCTCCTA CACAATAACC CTCACTATTT
CCCCCAACAG CCTTCTAGGT ACCACCCCCT GACGTTTGCC TTCTACCCAA

indicates that each of the subregions has been sampled approximately equally in the run. We have also explored the relative sampling frequencies realized in the other runs. The results are similar. This indicate that the simulation has converged, and our choices of T_0 and the number of iterations are appropriate for this example. Fig. 1(b) shows one consensus tree constructed based on the samples generated in the five runs.

Table 2
Relative sampling frequencies of the subregions for the simulated 20-taxon dataset.

Subregion	Frequency	Subregion	Frequency
$(-\infty,-1830)$	98.2702	$[-1830,-1828)$	99.3434
$[-1828,-1826)$	99.1771	$[-1826,-1824)$	99.0483
$[-1824,-1822)$	99.3917	$[-1822,-1820)$	99.3917
$[-1820,-1818)$	99.4507	$[-1818,-1816)$	100.7815
$[-1816,-1814)$	101.3717	$[-1814,-1812)$	101.5488
$[-1812,\infty)$	102.2249		

For comparison, two popular Bayesian phylogeny software, MrBayes and BAMBE, were also applied to this example. The reasons why we chose these two software for comparison are 2-fold. Firstly, they are available to the public and have been tested extensively. Secondly, also more importantly, they focus on the Bayesian phylogeny analysis instead of MAP trees. Each software was run 5 times independently with its default setting, and each run consisted of 2.2×10^5 iterations. Fig. 1(c) and (d) show the consensus trees produced by BAMBE and MrBayes, respectively. Fig. 1 indicates that all the consensus trees produced by the three methods are similar, but the tree produced by SAMC is most similar to the true one. Similar results can also be found in Figs. 3 and 4 for the 30-taxon and 40-taxon examples, where the consensus trees produced by SAMC are most similar to the true trees. To explain why SAMC tends to produce better consensus trees, we compare in Fig. 2 the progression curves of the best log-likelihood values produced by the above three methods. It indicates that SAMC tends to produce phylogenetic trees with higher likelihood values than do BAMBE and MrBayes.

The Bayesian analysis has also been done for the parameters of the HKY85 model. In each run, the first 2×10^4 iterations were discarded for the burn-in process, and the samples generated in the remaining iterations were used for the inference. The numerical results were summarized in Table 3. Since each site of the taxon sequences is modeled equally in this example, the parameter α is restricted to be 1. This is the same for all the three Bayesian methods under comparison. Table 3 shows that SAMC produces the highest averaged log-likelihood value and the most accurate estimate for the parameter κ. Note that the estimates of the nucleotide

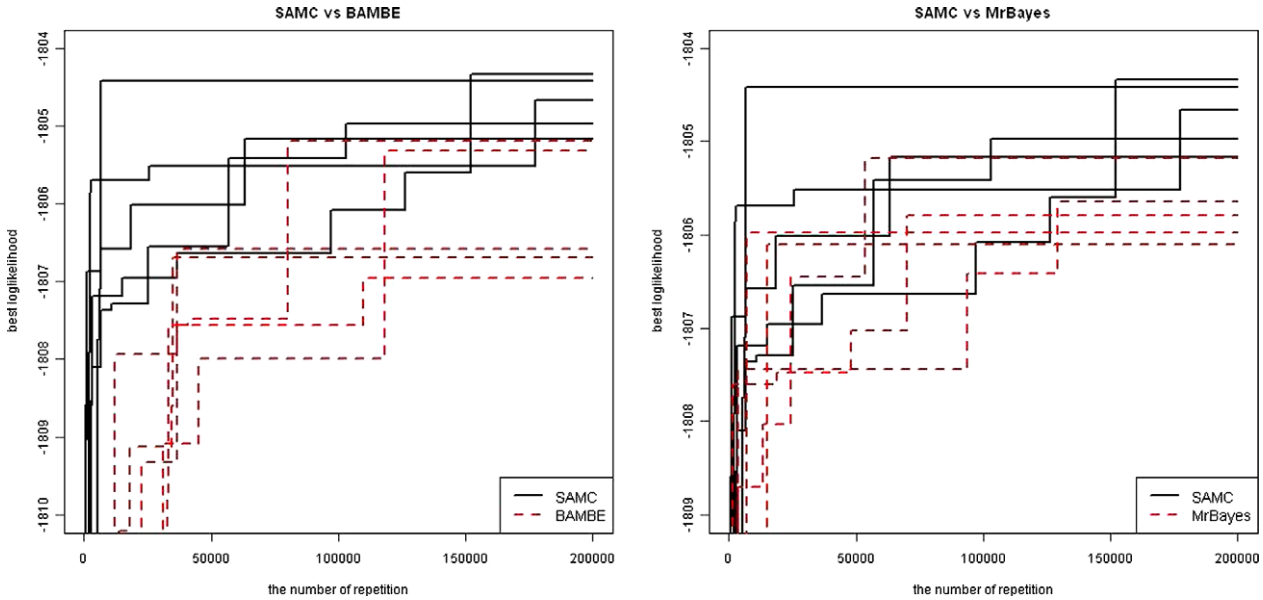

Fig. 2. Comparison of the progression curves of the best log-likelihood values produced by SAMC, MrBayes, and BAMBE. Left: SAMC versus BAMBE. Right: SAMC versus MrBayes.

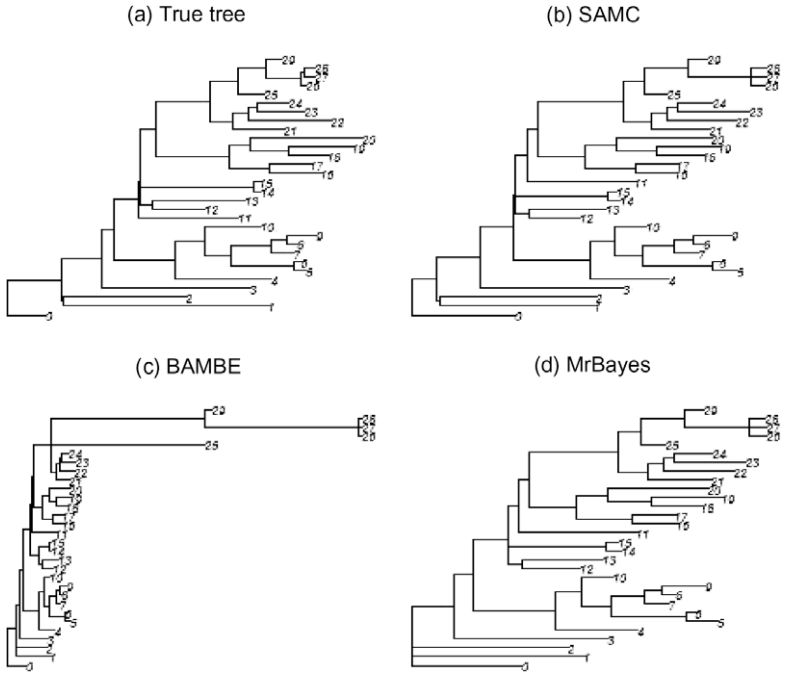

Fig. 3. Comparison of the consensus trees produced by SAMC, BAMBE and MrBayes for the simulated 30-taxon dataset.

frequencies produced by the three methods are similar. It is interesting to point out that SAMC produced the best results among the three methods, but it cost the least CPU time. MrBayes employed the parallel tempering algorithm for the simulation, where multiple Markov chains were run in parallel at different temperatures, so it cost more CPU time than the single chain methods, BAMBE and SAMC, for the same number of iterations. BAMBE was a little more time consuming than SAMC, as it included in the simulation about ten percents of global moves; 20,000 global moves were performed in the 220,000 iterations. As aforementioned, the global move is more time consuming than the local move. Later, MrBayes and BAMBE were re-run independently 5 times, and each run consisted of 2.2×10^6 iterations, a 10-fold increase of the number of iterations. However, neither MrBayes and BAMBE could produce better averaged likelihood values or estimates of κ than those produced by SAMC in the previous runs.

To figure out the impact of the number of taxa on phylogeny estimation, the three Bayesian methods were applied to two more datasets with 30 and 40 taxa, which are generated using the same program as described before. For each dataset, each method was

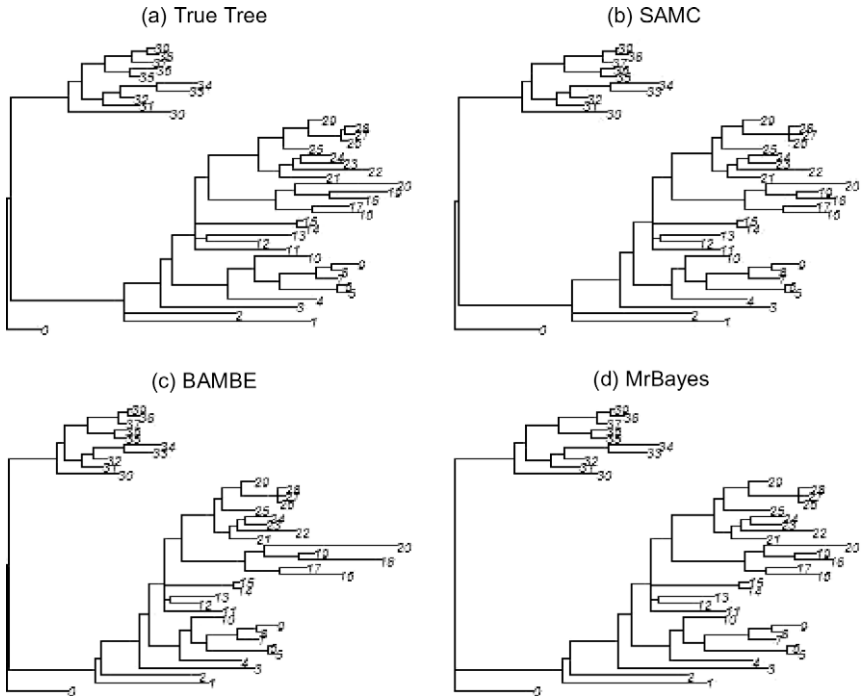

Fig. 4. Comparison of the consensus trees produced by SAMC, BAMBE and MrBayes for the simulated 40-taxon dataset.

Table 3
Bayesian analysis for the parameters of the HKY85 model used for the simulated 20-taxon dataset. CPU: CPU time (in minutes) cost by a single run of the algorithm on an Intel Pentium III computer. Averaged log-likelihood: the difference of the averaged log-likelihood value produced by the respective method and that produced by BAMBE. Each entry of the table is calculated by averaging over five independent runs, and the number in the parentheses represent the standard deviation of the corresponding average.

Methods	CPU (m)	Average log-likelihood	κ	π_A	π_G	π_C	π_T
SAMC	2.1	1.69 (0.07)	2.017 (9.9e−4)	0.257 (1.9e−4)	0.157 (1.6e−4)	0.321 (3.0e−4)	0.264 (2.4e−4)
MrBayes	10.2	0.62 (0.19)	1.894 (6.5e−3)	0.255 (7.7e−4)	0.158 (4.9e−4)	0.322 (7.0e−4)	0.266 (5.9e−4)
BAMBE	2.6	0 (0.52)	1.627 (4.0e−2)	0.248 (2.9e−3)	0.156 (3.4e−3)	0.330 (4.3e−3)	0.266 (5.1e−3)

run 5 times independently, and each run consisted of 2.2×10^5 iterations. The results were summarized in Table 4 and Figs. 3 and 4. They indicate again that SAMC outperforms BAMBE and MrBayes. SAMC produced the highest averaged likelihood value, the most accurate estimate of κ, and the consensus trees that have the highest similarity to the true trees.

4.2. Extensive simulation studies

Our simulation studies in Section 4.1 indicate that SAMC tends to outperform BAMBE and MrBayes in Bayesian phylogeny analysis. Comparing to BAMBE and MrBayes, SAMC tends to produce phylogenetic trees with higher likelihood values, more accurate model parameter estimates, and consensus trees more similar to the true trees. To concrete this result, more extensive simulation studies were conducted in this subsection.

For each of the true trees shown in Figs. 1, 3 and 4, 100 datasets were simulated according to a HKY85 model with the root sequence as shown in Table 1 and the parameters $\kappa = 2, \alpha = 1$, and $\pi_A = \pi_G = \pi_C = \pi_T = 0.25$. SAMC, BAMBE and MrBayes were applied to each of the 300 datasets with the same respective settings as described in Section 4.1. In all simulations, α was restricted to be 1, as the evolutionary rate of each site of the nucleotide sequences was modeled equally.

The results were summarized in Table 5, where MAST, standing for Maximum Agreement Subtree, provides a measurement for the similarity between the true trees and the consensus trees constructed by the respective methods. The MAST scores were calculated using the software `TreeAnalyzer`, which was developed by Dong and Kraemer (2004) based on the tree comparison algorithms by Farach et al. (1995) and Goddard et al. (1994), and was available at `http://www.cs.uga.edu/~eileen/TreeAnalyzer`. Note that `TreeAnalyzer` has normalized its output by 100; that is, the MAST score of two identical trees is 100. We also note that for our datasets, the consensus trees for the 30-taxon data have lower MAST values than those for the 40-taxon data. This is due to the structure difference of the corresponding true trees. The true 40-taxon tree has two well separated branches, making it relatively easier to be recovered.

The significance of the MAST scores produced by SAMC was assessed using the two-sample t-test. The test p-values were summarized in Table 6. The hypotheses corresponding to the first entry of Table 6 are H_0: *SAMC and MrBayes produced the same MAST scores for the 20-taxon data* versus H_1: *SAMC produced higher MAST scores for the 20-taxon data*. The hypotheses corresponding to other entries are similar. Table 6 indicates that the consensus trees produced by SAMC tend to be more resembled to the true trees than those produced by MrBayes and BAMBE, especially when the number of taxa is large. The tests are generally significant at a level of 0.01, except that the case SAMC versus BAMBE with 20 taxa is significant at a level of 0.05. This is reasonable, as the SAMC and BAME employ the same local moves and the example is relatively simple. Later, we simulated independently another 100 datasets for the case of 20 taxa. Putting the results of 200 datasets together, we got the p-values 3.57×10^{-4} and 8.71×10^{-3} for the tests SAMC versus MrBayes and SAMC versus BAMBE, respectively.

In addition to MAST scores, SAMC also produced more accurate estimates of κ than did MrBayes and BAMBE for these datasets. This can be easily seen from Table 7, which showed the root mean square errors of the estimates. It is known that root mean square error calibrates both the bias and variance of the estimates.

Table 4
Bayesian analysis for the parameters of the HKY85 model used for the simulated 30-taxon and 40-taxon datasets. CPU: CPU time (in minutes) cost by a single run of the algorithm on an Intel Pentium III computer. Averaged log-likelihood: the difference of the averaged log-likelihood value produced by the respective method and that produced by BAMBE. Each entry of the table is calculated by averaging over five independent runs, and the number in the parentheses represent the standard deviation of the corresponding average.

Methods	CPU (m)	Averaged log-likelihood	κ	π_A	π_G	π_C	π_T
30 taxa							
SAMC	6.2	4.51 (0.30)	2.01 (2.3e−3)	0.248 (1.4e−4)	0.158 (1.9e−4)	0.330 (2.4e−4)	0.265 (1.1e−4)
MrBayes	18.7	0.45 (0.95)	1.72 (1.4e−3)	0.245 (3.3e−4)	0.158 (2.8e−4)	0.331 (3.7e−4)	0.267 (4.5e−4)
BAMBE	6.4	0 (1.53)	1.35 (1.0e−1)	0.241 (1.7e−3)	0.151 (5.5e−3)	0.337 (4.6e−3)	0.271 (1.4e−3)
40 taxa							
SAMC	6.7	5.67 (0.51)	1.70 (1.5e−2)	0.251 (3.5e−4)	0.161 (2.6e−4)	0.315 (2.4e−4)	0.273 (2.8e−4)
MrBayes	25.5	2.07 (1.84)	1.68 (2.7e−3)	0.250 (3.1e−4)	0.162 (3.2e−4)	0.315 (4.6e−4)	0.273 (4.9e−4)
BAMBE	8.0	0 (1.53)	1.39 (9.3e−2)	0.247 (2.3e−3)	0.156 (2.4e−3)	0.320 (4.4e−3)	0.276 (1.6e−3)

Table 5
Bayesian analysis for the parameters of the HKY85 model used in the extensive simulation study. Averaged log-likelihood: the difference of the averaged log-likelihood value produced by the respective method and that produced by BAMBE. Each entry of the table is calculated by averaging the results over 100 independent datasets. and the number in the parentheses represent the standard deviation of the corresponding average.

Methods	Averaged log-likelihood	κ	π_A	π_G	π_C	π_T	MAST
20 taxa							
SAMC	2.427	1.998	0.283	0.181	0.266	0.271	98.940
	(3.512)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.228)
MrBayes	0.771	1.855	0.279	0.181	0.267	0.272	97.950
	(3.652)	(0.024)	(0.002)	(0.001)	(0.001)	(0.002)	(0.344)
BAMBE	0	1.810	0.283	0.177	0.270	0.270	98.280
	(3.653)	(0.019)	(0.002)	(0.001)	(0.001)	(0.002)	(0.292)
30 taxa							
SAMC	9.743	1.998	0.277	0.178	0.294	0.250	78.123
	(3.28)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.333)
MrBayes	5.146	2.154	0.274	0.181	0.295	0.252	75.182
	(3.166)	(0.032)	(0.002)	(0.001)	(0.001)	(0.001)	(0.350)
BAMBE	0	1.514	0.271	0.174	0.303	0.252	72.413
	(4.522)	(0.016)	(0.002)	(0.001)	(0.002)	(0.002)	(0.312)
40 taxa							
SAMC	30.084	1.952	0.288	0.183	0.298	0.232	85.704
	(3.535)	(0.010)	(0.002)	(0.001)	(0.001)	(0.001)	(0.432)
MrBayes	22.838	2.410	0.283	0.185	0.296	0.236	83.948
	(3.533)	(0.048)	(0.002)	(0.001)	(0.001)	(0.001)	(0.434)
BAMBE	0	1.389	0.288	0.171	0.305	0.236	81.716
	(7.727)	(0.021)	(0.002)	(0.001)	(0.001)	(0.002)	(0.526)

Table 6
The p-values of the two-sample t-tests (with unequal variances) for the MAST values produced by SAMC versus those produced by MrBayes and BAMBE.

Data	SAMC		
	20 Taxa	30 Taxa	40 Taxa
MrBayes	8.73×10^{-3}	2.88×10^{-9}	2.30×10^{-3}
BAMBE	3.83×10^{-2}	0.00	1.01×10^{-8}

Table 7
The root mean square errors (RMSEs) of the estimates of κ produced by SAMC, MrBayes and BAMBE for the 300 datasets. For an estimate of κ, the RMSE is defined as $\sqrt{(\hat{\kappa}-2)^2+s_{\hat{\kappa}}^2}$, where $\hat{\kappa}$ denotes the estimate and $s_{\hat{\kappa}}^2$ denotes the variance of the estimate.

Data	20 Taxa	30 Taxa	40 Taxa
SAMC	0.0019	0.0023	0.0485
MrBayes	0.1472	0.1569	0.4132
BAMBE	0.1904	0.4867	0.6117

Finally, we would like to point out that Table 5 indicates a positive correlation between the averaged log-likelihood values and the averaged MAST scores. That is, the consensus trees constructed from higher likelihood trees tend to be more similar to the true trees. This further implies that the higher likelihood trees tend to be more similar to the true trees. This also explains why SAMC tends to outperform BAMBE and MrBayes; SAMC less likely gets trapped in local energy minima, and thus has more chance to infer correctly from the posterior distribution of the phylogeny.

4.3. Cichlid fishes

In this subsection, we analyzed aligned protein coding mitochondrial DNA sequences obtained from 32 species of cichlid fishes (Kocher et al., 1995). Table 1 of the supporting document shows the tribal classification of 32 species of African cichlid fish. Taxa 1–5 form a flock from Lake Malawi. The remainder from Lake Tanganyika (6–31 taxa) constitute a Tanganyikan flock. The Malawi, Ectodini, and Lamprologini tribes are represented by the letters A, C, and D, respectively. Class B consists of $\{6,7,8,9\}$, a combination of most of Tropheini and one species of Limnochromini. Classes $E = \{22,23,24,26,27\}$ and $F = \{28,29,30,31\}$ are convenient conglomerations (pseudoclades) of the remaining tribes. Taxon $\{25\}$ is not grouped. Taxon $\{32\}$ is an outgroup from cichlid America. Each DNA sequence consists of 1044 sites. Across all species, identical nucleotides are observed on 567 sites, and the nucleotides on the remaining sites are used for phylogenetic tree construction. The HKY85 model was considered for the real dataset.

Table 8
Bayesian analysis for the parameters of the HKY85 model used for the cichlid fish example. CPU: CPU time (in minutes) cost by a single run of the algorithm on an Intel Pentium III computer. Averaged log-likelihood: the difference of the averaged log-likelihood value produced by the respective method and that produced by BAMBE. Each entry of the table is calculated by averaging over five independent runs, and the number in the parentheses represent the standard deviation of the corresponding average.

Methods	CPU (m)	Averaged log-likelihood	κ	π_A	π_G	π_C	π_T
SAMC	9.3	2.83	6.03	0.243	0.158	0.347	0.253
		(0.34)	(8.2e−2)	(4.5e−4)	(1.7e−4)	(2.5e−4)	(4.1e−4)
MrBayes	22.5	1.22	6.83	0.246	0.158	0.344	0.252
		(1.06)	(1.1e−2)	(2.6e−4)	(2.2e−4)	(2.7e−4)	(4.6e−4)
BAMBE	10.1	0.0	5.73	0.246	0.150	0.348	0.256
		(2.32)	(2.7e−1)	(3.2e−3)	(5.8e−3)	(2.0e−3)	(2.4e−3)

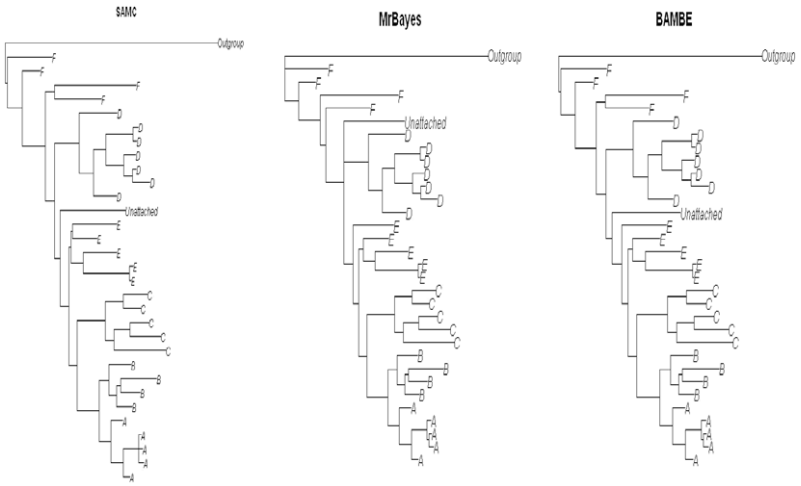

Fig. 5. Comparison of the consensus trees produced by SAMC, MrBayes and BAMBE for the cichlid fish example.

SAMC was first applied to this example. The sample space was partitioned as follows: $E_1 = \{\omega : U(\omega) > 8701\}, E_2 = \{\omega : 8701 \geqslant U(\omega) > 8699\}, \cdots, E_{39} = \{\omega : 8627 \geqslant U(\omega) > 8625\}, E_{40} = \{\omega : U(\omega) \leqslant 8625\}$. SAMC was run five times and each run consisted of 2.2×10^5 iterations. For comparison, MrBayes and BAMBE were also applied to this example. Each software was run five times, and each run consisted of 2.2×10^5 iterations. The first 2×10^4 samples were discarded for the burn-in process, and the samples generated in the remaining iterations were used for phylogeny inference. The results were summarized in Table 8. Again, SAMC produced the highest averaged likelihood value among the three Bayesian methods. In Fig. 5, we compared the consensus trees produced by the three methods. It is interesting to note that these trees are all very similar.

5. Conclusion

In this paper, we have applied the stochastic approximation Monte Carlo algorithm to Bayesian phylogeny analysis. The new method was compared with two popular Bayesian phylogeny software, BAMBE and MrBayes, on simulated and real datasets. The numerical results indicate that our method outperforms BAMBE and MrBayes. Among the three methods, SAMC produced the consensus trees which are most similar to the true trees, and the model parameter estimates which have the smallest mean square errors, but cost the least CPU time.

Comparing to BAMBE and MrBayes, SAMC less likely gets trapped by local energy minima, and thus has more chance of inferring correctly from the posterior distribution for the phylogeny. This is also the reason why SAMC tends to outperform BAMBE and MrBayes. This work can be extended in various ways, e.g., applying SAMC to more complicated evolutionary models. We note that efficiency of SAMC can be improved further by incorporating some smoothing techniques into its sampling procedure as suggested by Liang (2009b).

Acknowledgments

Liang's research was partially supported by the grant (DMS-0607755) made by the National Science Foundation and the award (KUS-C1-016-04) made by King Abdullah University of Science and Technology (KAUST). The authors thank the editor Professor A.L. Hughes and the referees for their comments which have led to significant improvement of this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.06.019.

References

Alfaro, M.E., Holder, M.T., 2006. The posterior and the prior in Bayesian phylogenetics. Annu. Rev. Ecol. Syst. 37, 19–42.
Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415.
Andrieu, C., Moulines, É., Priouret, P., 2005. Stability of stochastic approximation under verifiable conditions. SIAM J. Control Optimization 44, 283–312.
Chen, H.F., 2002. Stochastic Approximation and its Applications. Kluwer Academic Publishers, Dordrecht.
Cheon, S., Liang, F., 2008. Phylogenetic tree construction using sequential stochastic approximation Monte Carlo. BioSystems 91, 94–107.
Dong, S., Kraemer, E., 2004. Calculation, visualization, and manipulation of MASTs (maximum agreement subtrees). In: Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference. IEEE Computer Society, Washington, DC, pp. 405–414.
Duret, L., Galtier, N., 2000. The covariation between TpA deficiency, CpG deficiency, and G + C content of human isochores is due to a mathematical artifact. Mol. Biol. Evol. 17, 1620–1625.
Farach, M., Przytycka, T., Thorup, M., 1995. On the agreement of many trees. Inform. Process.Lett. 55, 297–301.
Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Felsenstein, J., 1993. PHYLIP (Phylogenetic inference package), version 3.5. Univ. Washington, Seattle.
Felsenstein, J., Churchill, G.A., 1996. A hidden Markov model approach to variation among sites in rate of evolution. Mol. Biol. Evol. 13, 93–104.
Feng, X., Buell, D.A., Rose, J.R., Waddell, P.J., 2003. Parallel algorithms for Bayesian phylogenetic inference. J. Parallel Distrib. Comput. 63, 707–718.
Fitch, W.M., 1971. Toward defining the course of evolution: minimal change for a specific tree topology. Syst. Zool. 20, 406–416.
Galtier, N., Gascuel, O., Jean-Marie, A., 2005. Markov models in molecular evolution. In: Nielsen, R. (Ed.), Statistical Methods in Molecular Evolution. Springer, pp. 3–24.
Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.
Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood. In: Keramigas, E.M. (Ed.), Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. Interface Foundation, Fairfax, pp. 156–163.
Goddard, W., Kubicka, E., Kubicki, G., McMorris, F.R., 1994. The agreement metric for labeled binary trees. Math. Biosci. 123, 215–226.
Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.
Jukes, T., Cantor, C., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), Mammalian Protein Metabolism. Academic Press, New York.
Kidd, K.K., Sgaramella-Zonta, L.A., 1971. Phylogenetic analysis: concepts and methods. Am. J. Hum. Genet. 23, 235–252.
Kimura, M., 1980. A simple method for estimating evolutionary rate in a finite population due to mutational production of neutral and nearly neutral base substitution through comparative studies of nucleotide sequences. J. Mol. Biol. 16, 111–120.
Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 31, 151–160.
Kocher, T.D., Conroy, J.A., MacKaye, K.R., Stauffer, J.R., Lockwood, S.F., 1995. Evolution of NADH dehydrogenase submit 2 in East African cichlid fish. Mol. Phylogenet. Evol. 4, 420–432.
Lakner, C.P., van der Mark, P., Huelsenbeck, J.P., Larget, B., Ronquist, F., 2008. Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. Syst. Biol. 57, 86–103.
Larget, B., Simon, D., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16, 750–759.
Li, S., Pearl, D., Doss, H., 2000. Phylogenetic tree construction using Markov chain Monte Carlo. J. Am. Stat. Assoc. 95, 493–508.
Liang, F., 2009a. On the use of stochastic approximation Monte Carlo for Monte Carlo Integration. Stat. Probability Lett. 79, 581–587.
Liang, F., 2009b. Improving SAMC using smoothing methods: theory and applications to Bayesian model selection problems. Ann. Statist. 37, 2626–2654.
Liang, F., Liu, C., Carroll, R.J., 2007. Stochastic approximation in Monte Carlo computation. J. Am. Stat. Assoc. 102, 305–320.
Liu, J.S., Liang, F., Wong, W.H., 2000. The use of multiple-try method and local optimization in Metropolis sampling. J. Am. Stat. Assoc. 95, 121–134.
Maddison, D.R., 1991. The discovery and importance of multiple islands of most parsimonious trees. Syst. Zool. 40, 315–328.
Mau, B., 1996. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Ph.D. thesis. University of Wisconsin-Madison, Department of Statistics.
Mau, B., Newton, M.A., 1997. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. J. Comp. Graph. Stat. 6, 122–131.
Mau, B., Newton, M.A., Larget, B., 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo. Biometrics 55, 1–12.
Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1091.
Newton, M.A., 1996. Bootstrapping phylogenies: large deviations and dispersion effects. Biometrika 83, 315–328.
Newton, M.A., Mau, B., Larget, B., 1999. Markov chain Monte Carlo for the Bayesian analysis of evolutionary trees from aligned molecular sequences. In: Seillier-Moseiwitch, F. (Ed.), Statistics in Molecular Biology and Genetics. IMS Lecture Notes-Monograph Series, vol. 33, pp.143–162.
Pollock, D.D., Taylor, W.R., Goldman, N., 1999. Coevolving protein residues: maximum likelihood identification and relationship to structure. J. Mol. Biol. 287, 187–198.
Rannala, B., Yang, Z., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304–311.
Robbins, H., Monro, S., 1951. A stochastic approximation method. Ann. Math. Stat. 22, 400–407.
Roberts, G.O., Tweedie, R.L., 1996. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. Biometrika 83, 95–110.
Robinson, D.M., Jones, D.T., Kishino, H., Goldman, N., Thorne, J.L., 2003. Protein evolution with dependence among codons due to tertiary structure. Mol. Biol. Evol. 20, 1692–1704.
Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum evolution trees. Mol. Biol. Evol. 9, 945–967.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Salter, L.A., Pearl, D.K., 2001. Stochastic search strategy for estimation of maximum likelihood phylogenetic trees. Syst. Biol. 50, 7–17.
Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. 99, 16138–16143.
Thorne, J.L., Goldman, N., Jones, D.T., 1996. Combining protein evolution and secondary structure. Mol. Biol. Evol. 13, 666–673.
Wang, F., Landau, D.P., 2001. Efficient, multiple-range random walk algorithm to calculate the density of states. Phys. Rev. Lett. 86, 2050–2053.
Yang, Z., 1995. A space-time process model for the evolution of DNA sequences. Genetics 139, 993–1005.
Yang, Z., Rannala, B., 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol. Biol. Evol. 14, 717–724.