Syst. Biol. 67(3):503–517, 2018

DOI:10.1093/sysbio/syx087
Advance Access publication December 13, 2017

Online Bayesian Phylogenetic Inference: Theoretical Foundations via Sequential Monte Carlo

Vu Dinh[1,*], Aaron E. Darling[2], and Frederick A. Matsen IV[3]

[1] *Department of Mathematical Sciences, University of Delaware, 312 Ewing Hall, Newark, DE 19716, USA;*
[2] *The ithree institute, University of Technology Sydney, 15 Broadway, Ultimo NSW 2007, Australia; and*
[3] *Program in Computational Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA*
**Correspondence to be sent to: Department of Mathematical Sciences, University of Delaware, 312 Ewing Hall, Newark, DE 19716, USA;*
Email: vucdinh@udel.edu.

Received 02 February 2017; reviews returned 08 November 2017; accepted 09 November 2017
Associate Editor: Edward Susko

Abstract.—Phylogenetics, the inference of evolutionary trees from molecular sequence data such as DNA, is an enterprise that yields valuable evolutionary understanding of many biological systems. Bayesian phylogenetic algorithms, which approximate a posterior distribution on trees, have become a popular if computationally expensive means of doing phylogenetics. Modern data collection technologies are quickly adding new sequences to already substantial databases. With all current techniques for Bayesian phylogenetics, computation must start anew each time a sequence becomes available, making it costly to maintain an up-to-date estimate of a phylogenetic posterior. These considerations highlight the need for an *online* Bayesian phylogenetic method which can update an existing posterior with new sequences. Here, we provide theoretical results on the consistency and stability of methods for online Bayesian phylogenetic inference based on Sequential Monte Carlo (SMC) and Markov chain Monte Carlo. We first show a consistency result, demonstrating that the method samples from the correct distribution in the limit of a large number of particles. Next, we derive the first reported set of bounds on how phylogenetic likelihood surfaces change when new sequences are added. These bounds enable us to characterize the theoretical performance of sampling algorithms by bounding the effective sample size (ESS) with a given number of particles from below. We show that the ESS is guaranteed to grow linearly as the number of particles in an SMC sampler grows. Surprisingly, this result holds even though the dimensions of the phylogenetic model grow with each new added sequence. [Bayesian inference; effective sample size; online inference; phylogenetics; sequential Monte Carlo; subtree optimality.]

Maximum likelihood and Bayesian methods currently form the most popular means of phylogenetic inference. The Bayesian methods in particular enjoy the flexibility to incorporate a wide range of ancillary model features such as geographical information or trait data which are essential for some applications (Lemey et al., 2009; Ronquist et al., 2012). However, Bayesian tree inference with current implementations is a computationally intensive task, often requiring days or weeks of CPU time to analyze modest data sets with 100 or so sequences.

New developments in DNA and RNA sequencing technology have led to sustained growth in sequence data sets. This advanced technology has enabled real time outbreak surveillance efforts, such as ongoing Zika, Ebola, and foodborne disease sequencing projects, which make pathogen sequence data available as an epidemic unfolds (Gardy et al., 2015; Quick et al., 2016). In general, these new pathogen sequences arrive one at a time (or in small batches) into a background of existing sequences. Most phylogenetic inferences, however, are performed “from scratch” even when an inference has already been made on the previously available sequences. Projects such as `nextflu.org` (Neher and Bedford, 2015) incorporate new sequences into trees as they become available, but do so by recalculating the phylogeny from scratch at each update using a fast approximation to maximum likelihood inference, rather than a Bayesian method.

Thus, modern researchers using phylogenetics are in the situation of having previous inferences, having new sequences, and yet having no principled method to incorporate those new sequences into existing inferences. Existing methods either treat a previous point estimate as an established fact and directly insert a new sequence into a phylogeny (Matsen et al., 2010; Berger et al., 2011) or use such a tree as a starting point for a new maximum-likelihood search (Izquierdo-Carrasco et al., 2014). There is currently no method to update posterior distributions on phylogenetic trees with additional sequences.

In this article, we develop the theoretical foundations for an online Bayesian method for phylogenetic inference based on Sequential and Markov Chain Monte Carlo (MCMC). Unlike previous applications of Sequential Monte Carlo (SMC) to phylogenetics (Bouchard-Côté et al., 2012; Bouchard-Côté, 2014; Wang et al., 2015), we develop and analyze online algorithms that can update a posterior distribution as new sequence data becomes available. We first show a consistency result, demonstrating that the method samples from the correct distribution in the limit of a large number of particles in the SMC. Next, we derive the first reported set of bounds on how phylogenetic likelihood surfaces change when new sequences are added. These bounds enable us to characterize the theoretical performance of sampling algorithms by developing a lower bound on the effective sample size (ESS) for a given number of particles. Surprisingly, this result holds even though the dimensions of the phylogenetic model grow with each new added sequence.

Mathematical Setting

Background and Notation

Throughout this article, a *phylogenetic tree* $t=(\tau,l)$ is a tree τ with leaves labeled by a set of taxon names (e.g., species names), such that each edge e is associated with a non-negative number l_e. These trees will be unrooted, although sometimes a root will be designated for notational convenience. For each phylogenetic tree (τ,l), we will refer to τ as its *tree topology* and to l as the vector of *branch lengths*. We denote by $E(\tau)$ the set of all edges in trees with topology τ; any edge adjacent to a leaf is called a *pendant edge*, and any other edge is called an *internal edge*.

We will employ the standard likelihood-based framework for statistical phylogenetics on discrete characters under the common assumption that alignment sites are independently and identically distributed (IID) (Felsenstein, 2004), which we now review briefly. Let Ω denote the set of character states and let $r=|\Omega|$. For DNA $\Omega=\{A,C,G,T\}$ and $r=4$. We assume that the mutation events occur according to a continuous time Markov chain on states Ω with instantaneous rate matrix Ξ and stationary distribution ω. This rate matrix Ξ and the branch length l_e on the edge e define the transition matrix $G(l_e)=e^{l_e\Xi}$ on edge e, where $G_{ij}(l_e)$ denotes the probability of mutating from state i to state j across the edge e (with length l_e).

In an online setting, the taxa $\{X_1,X_2,\ldots,X_N\}$ and their corresponding observed sequences $\{\psi_1,\psi_2,\ldots,\psi_N\}$, each of length S, arrive in a specific order, where N is a finite but large number. For all $n\leq N$, we consider the set of all phylogenetic trees that have $\{X_1,X_2,\ldots,X_n\}$ as their set of taxa and seek to sample from a sequence of probability distributions $\bar{\pi}_n$ of increasing dimension corresponding to phylogenetic likelihood functions (Felsenstein, 2004).

For a fixed phylogenetic tree (τ,l), the phylogenetic likelihood is defined as follows and will be denoted by $L(\tau,l)$. Given the set of observations $\psi(n)=(\psi_1,\psi_2,\ldots,\psi_n)\in\Omega^{S\times n}$ of length S up to sequence n, the likelihood of observing $\psi(n)$ given the tree has the form

$$L_n(\tau,l)=\prod_{u=1}^{S}\sum_{a^u}\omega(a^u_\sigma)\prod_{(i,j)\in E(\tau)}G_{a^u_i a^u_j}(l_{(i,j)}),$$

where a^u ranges over all extensions of ψ to the internal nodes of the tree, a^u_i denotes the assigned state of node i by a^u, and σ denotes the root of the tree. Although we designate a root for notational convenience, the methods and results we discuss apply equally to unrooted trees.

Given a proper prior distribution with density $\pi_0^{(n)}$ imposed on branch lengths and tree topologies, the target posterior distributions have densities $\bar{\pi}_n(\tau,l)\sim L_n(\tau,l)\pi_0^{(n)}(\tau,l)$. We will also use $\hat{\pi}_n(\tau,l)$ to denote the unnormalized density $L_n(\tau,l)\pi_0^{(n)}(\tau,l)$. Throughout the article, we assume that the phylogenetic trees of interest all have non-negative branch lengths bounded from above by $b>0$ and use $\mathcal{T}_n$ to denote the set of all such trees.

An important goal of this article is to show that the collection of particles generated by the online phylogenetic sequential Monte Carlo (OPSMC) algorithm do a good job of approximating a sample from the posterior distribution, in the same way that a sample from a MCMC algorithm approximates a sample from the posterior. We would like to show that this approximation becomes arbitrarily good as the number of particles goes to infinity, that is, that the particle distribution converges to the posterior distribution. The type of convergence we will demonstrate is called *weak convergence*, which means here that for any integrable real-valued function ϕ defined on the set of trees, the weighted average of the value of the function $\phi(\tau,l)$ over the trees (τ,l), and corresponding weights generated by the algorithm converges (with probability 1) to the to the posterior mean of $\phi(\tau,l)$. This convergence result also implies weak convergence of many key quantities of interest in phylogenetics: convergence of posterior probabilities on trees, branch lengths, and splits.

The proofs require a slightly more abstract measure-theoretic perspective. A *measure* in this context is a function from subsets of some set to the non-negative real numbers with nice properties, such as that the measure of a countable union of disjoint sets is the sum of the measures of the individual sets. A classic example is Lebesgue measure on subsets of real space, which simply gives the volume of each subset. We can extend this to $\mathcal{T}_n$, the product space of the space of all possible tree topologies and the space of all branch lengths $[0,b]^{2n-3}$, via the corresponding product measure, $\mu_n(dr)$, that is uniform on trees and Euclidean on each branch length space. In our presentation, we will use the notation $\|\phi\|_{\mathcal{T}_n}$ to denote the integration of a test function ϕ (defined on $\mathcal{T}_n$) with respect to the measure $\mu_n(dr)$. Specifically, this integration can be computed by

$$\|\phi\|_{\mathcal{T}_n}:=\frac{1}{V_n}\sum_{\tau}\int_{[0,b]^{2n-3}}\phi(\tau,l)dl$$

for any n and test function ϕ defined on $\mathcal{T}_n$, where $V_n=(2n-3)!!$ is the number of different topologies of $\mathcal{T}_n$. A posterior distribution on $\mathcal{T}_n$ can also be considered a measure, such that the value of the posterior measure on a subset of $\mathcal{T}_n$ is simply the integral of the posterior on that subset.

In this article, we will also be considering measures on collections of discrete particles, which are simply non-negative weights on those particles: the measure applied to a collection of particles is the sum of those particle weights. This measure will be called the empirical measure. Thus, said in measure-theoretic terms, we would like to show that the empirical measure on the particles converges weakly to the posterior measure on $\mathcal{T}_n$.

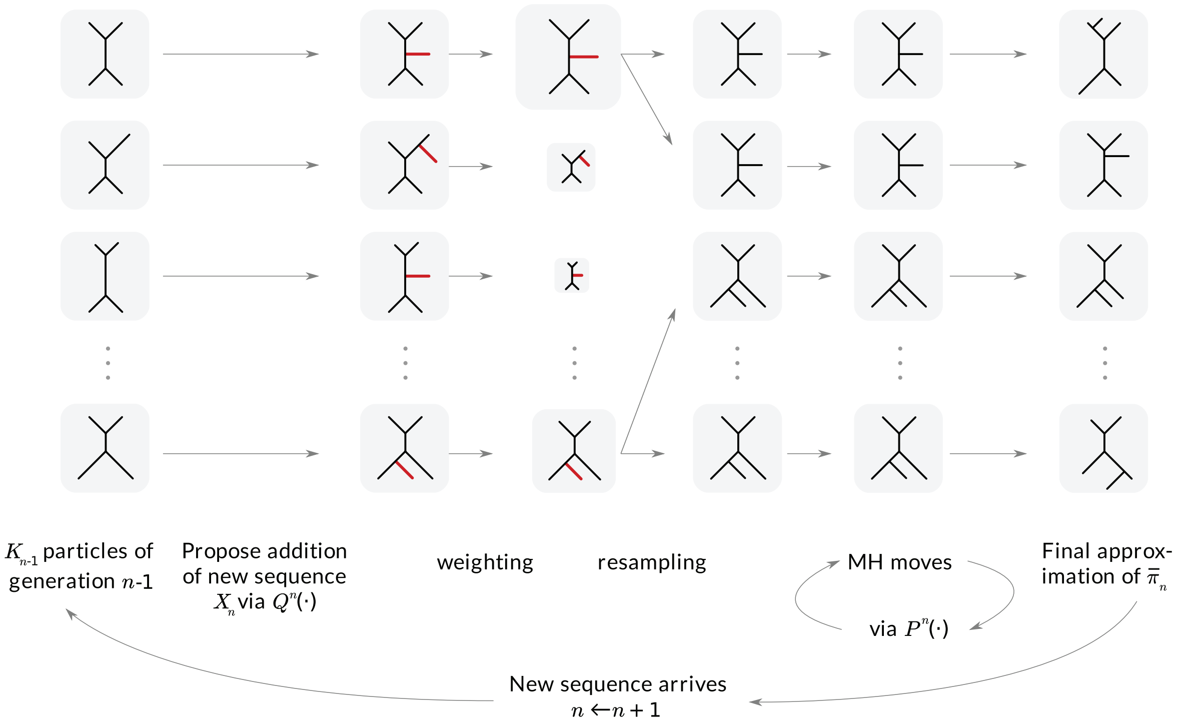
$K_{n\text{-}1}$ particles of generation n-1
Propose addition of new sequence X_n via $Q^n(\cdot)$
weighting
resampling
MH moves
via $P^n(\cdot)$
Final approximation of $\bar{\pi}_n$
New sequence arrives
$n \leftarrow n+1$

FIGURE 1. An overview of the OPSMC algorithm.

Sequential Monte Carlo

SMC methods are designed to approximate a sequence of probability distributions changing through time. These probability distributions may be of increasing dimension or complexity. They track the sequence of probability distributions of interest by producing a discrete representation of the distribution $\bar{\pi}_n$ at each generation n through a random collection of weighted particles. In the online phylogenetic context, these particles are trees; after each generation, new sequences arrive, and the collection of particles is updated to represent the next target distribution.

To make it easier to keep track of all variables involved in the description of SMC, we also provide a table of variables in the Appendix. Figure 1 can also be used as a reference to the steps of the algorithm. It is worth noting that the algorithm starts with a set of weighted particles, which are illustrated at the center of Figure 1, right before the "resampling step." While the details of the algorithms might vary, the main idea of SMC interspersed with MCMC sampling can be described as follows.

At the beginning of each generation n, a list of K_n particles $t_1^n,\ldots,t_{K_n}^n$ are maintained along with a positive weight w_i^n associated with each particle t_i^n. (In our phylogenetic context, each particle at this point is a phylogenetic tree with n taxa). Using a valid SMC algorithm, the weights of these particles can be normalized to approximate the integral of any integrable test function ϕ with respect to the target distribution, which in our case is:

$$\sum_{i=1}^{K_n} w_i^n \phi(t_i^n) \Big/ \sum_{i=1}^{K_n} w_i^n \approx \frac{1}{V_n} \sum_{\tau} \int_{[0,b]^{2n-3}} \bar{\pi}_n(\tau,l)\phi(\tau,l)\,dl. \tag{1}$$

This approximation will be formalized into a statement about weak convergence of the empirical particle distribution to the target (in our case the phylogenetic posterior).

A new list of K_{n+1} particles is then created in three steps: resampling, Markov transition, and mutation.

Resampling step The aim of the resampling step is to obtain an unweighted empirical distribution of the weighted measure $\hat{\pi}_{n,K_n}$ by discarding samples with small weights and allowing samples with large weights to reproduce (Fig. 1, "resampling" step). Particle sampling is done in proportion to particle weight, thus from an evolutionary perspective, this is a Wright–Fisher step using particle weights as fitnesses. After resampling, the weights on the particles have effectively been translated into their frequency, and so each particle is then assigned the same weight. A total of K_{n+1} particles are sampled from the weights w_i^n, and we denote the particles obtained after this step by s_i^n.

Markov transition The scheme employed in the resampling step introduces some Monte Carlo error. Moreover, when the distribution of the weights from the previous generation is skewed, the particles having high importance weights might be oversampled. This results in a depletion of samples (or *path degeneracy*): after some generations, numerous particles are in fact sharing the same ancestor as the result of a "selective sweep" in the SMC algorithm.

An optional Markov transition step can be employed to alleviate this sampling bias, during which MCMC steps are run separately on each particle s_i^n for a certain amount of time to obtain a new approximately independent sample m_i^n (Fig. 1, "MH moves" step). These transitions are done via the Metropolis–Hastings kernel P^n, which we will describe in more detail in the next section. Note that in our phylogenetic setting, the dimension of the state space does not change in the Resampling and Markov transition steps, and all particles have the same number of taxa as ones from the original list (n taxa).

Mutation step Finally, in the mutation step, new particles $t_1^{n+1},\ldots,t_{K_{n+1}}^{n+1}$ are created from the previous generation $s_1^n,\ldots,s_{K_{n+1}}^n$ using a proposal distribution Q^n (that may be dependent on data). That is, for each $i=1,\ldots,K_{n+1}$, the particle t_i^{n+1} is independently generated from the proposal distribution $Q^n(s_i^n,t)$. The particles are then weighted by an appropriate weight function h. If we assume further that for each state t, there exists a unique state, denoted by $\varrho(t)$, such that $Q^n(\varrho(t),t)>0$, then h can be chosen as

$$h(t)=\frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))\,Q^n(\varrho(t),t)}. \tag{2}$$

The new particles $t_1^{n+1},\ldots,t_{K_{n+1}}^{n+1}$ with their corresponding weights $w_1^{n+1},\ldots,w_{K_{n+1}}^{n+1}$ (where $w_i^{n+1}=h(t_i^{n+1})$ for $i=1,\ldots,K_{n+1}$) now represent the distribution $\bar{\pi}_{n+1}$ and act as the input for the next generation.

In our phylogenetic context, this mutation step adds a new sequence to the tree: the new particle t_i^{n+1} will be $(n+1)$-taxon tree obtained from the n-taxon tree t_i^n (Fig. 1, "propose addition of new sequence"). Then $\varrho(t)$ is simply t but with the most recently added pendant edge removed. To add a sequence, the proposal strategy Q^n will specify: an edge e to which the new pendant edge is added, the position x on that edge to attach the new pendant edge, and the length y of the pendant edge. Different ways of choosing (e,x,y) lead to different sampling strategies and performances. Note that such an addition changes not only the distribution on the state space, but also the state space itself.

The process is then iterated until $n=N$.

Online Phylogenetic Inference via Sequential Monte Carlo

Next, we more fully develop OPSMC and isolate technical conditions that will ensure a correct sampler. For these conditions, we make a distinction between Criteria, which are a more general set of conditions that imply consistency of OPSMC, and stronger Assumptions, which guarantee that the Criteria hold and to enable further analyses of the sampler.

In contrast to the traditional setting of SMC, for OPSMC when the number of leaves n of the particles increases, not only does the local dimension of the space $\mathcal{T}_n$ increase linearly, the number of different topologies in $\mathcal{T}_n$ also increases superexponentially in n. Careful constructions of the proposal distribution Q^n, which will build $(n+1)$-taxon trees out of n-taxon trees, and the Markov transition kernel P^n are essential to cope with this increasing complexity. This goes beyond simply satisfying the criteria guaranteeing a correct sampler (Fourment et al., 2017).

General Criteria for a Consistent OPSMC Sampler

The proposal distributions Q^n will be designed in such a way that the following criterion holds.

Criterion 1. *At every generation of the OPSMC sampling process, for any two trees r and r' in the tree space $\mathcal{T}=\bigcup\mathcal{T}_n$, the proposal density Q^n satisfies $Q^n(r,r')>0$ if and only if r can be obtained from r' by removing the taxon X_{n+1} and its corresponding edge.*

This formulation is analogous to the definition of "covering" from (Wang et al., 2015), although is distinct in the setting of online inference. In either case, for every tree $t\in\mathcal{T}_{n+1}$, there exists a unique tree $\varrho(t)$ in $\mathcal{T}_n$ such that $Q^n(\varrho(t),t)>0$ and thus a weight function of the form (2) can be used.

As we mentioned in the previous section, to obtain an $(n+1)$-taxon tree from an n-taxon tree, a proposal strategy Q^n must specify:

1. an edge e to which the new pendant edge is added,
2. the position x on that edge to attach the new pendant edge, and
3. the length y of the pendant edge.

The position x on an edge of a tree will be specified by its *distal* length, which is the distance from the attachment location to the end of the edge that is farthest away from the root of the tree.

Besides the SMC proposal strategy Q^n, an appropriate Markov transition kernel P^n can increase the effectiveness of an OPSMC algorithm. It is worth noting that the problem of sample depletion is even more severe for OPSMC than for typical applications of SMC, since after each generation, the sampling space actually expands in dimensionality and complexity. To alleviate this sampling bias, MCMC steps can be run separately on each particle for a certain amount of time to obtain new independent samples. We require the following criterion, which is as expected for any Markov transition kernel used in standard MCMC.

Criterion 2. *At every generation of the OPSMC sampling process, if a Markov transition kernel is used, the Markov transition kernel P^n has $\bar{\pi}_n$ as its stationary distribution.*

As we will see later in the proof of consistency of OPSMC, Criterion 2 is the only assumption required of the Markov transition kernel if it is applied. This leaves us with a great degree of freedom to improve the efficiency of the sampling algorithm without damaging its theoretical properties. For example, one can use global information provided by the population of particles, such as ESS (Beskos et al., 2014), to guide the proposal, or to define a transition kernel on the whole set (or some subset) of particles (Andrieu et al., 2001). In the context of phylogenetics, we can design a sampler that recognizes subtrees that have been insufficiently sampled, and samples more particles to improve the ESS within such regions. Similarly, one can use samplers that rearrange the tree structure in the neighborhood of newly added pendant edges.

We will prove in Theorem 10 that when Criteria 1 and 2 are satisfied, then the OPSMC sampler is consistent.

Designing an Effective OPSMC Sampler

Although Criteria 1 and 2 guarantee the consistency of OPSMC, they do not lead to any insights about the performance of the OPSMC sampler. From a practical point of view, one is interested in quantifying some measure of efficiency of an MCMC algorithm. This question is even more interesting for SMC (and in particular, for phylogenetic SMC algorithms), since in the general case it is known that SMC-type algorithms may suffer from the curse-of-dimensionality: when the dimension of the problem increases, the number of the particles must increase exponentially to maintain a constant effective sample size (Chopin, 2004; Bengtsson et al., 2008; Bickel et al., 2008; Snyder et al., 2008).

In this article, we are interested how the ESS of the OPSMC sampler behaves in the limit of a large number of particles. This task is challenging because we need to quantify how the target distribution changes with newly added sequences, and how the proposals should be designed to cope with these changes in an efficient way. Throughout this section, we will also derive several assumptions, imposed on both the target distribution and the design of the proposal, to enable such analyses.

First we assume the prior distribution does not change drastically when a new taxon is added:

Assumption 3. *There exists $C_0 > 0$ independent of n, and $A_1(n), A_2(n) > 0$ such that*

$$A_1(n) \le \frac{\pi_0^{(n+1)}(t)}{\pi_0^{(n)}(\rho(t))} \le A_2(n) \qquad \forall t \in \mathcal{T}_{n+1}$$

and $A_2(n)/A_1(n) \le C_0$ for all n.

In practice, the most common prior distribution on unrooted trees is the uniform distribution on topologies with an identical prior on each of the branches, which satisfy the Assumption 3 with $C_0 = 1$.

As we will discuss in later sections, to make sure that the proposals can capture the posterior distributions $\bar{\pi}_n$ efficiently, some regularity conditions on $\bar{\pi}_n$ are also necessary. These conditions are formalized in terms of a lower bound on the posterior expectation of $\zeta(r)$, the average branch length of r for a given tree $r \in \mathcal{T}_n$:

Assumption 4 (Assumption on the average branch length)**.** *There exist positive constants c (independent of n) such that for each n*

$$c \le \frac{1}{V_n} \sum_{\tau} \int_{[0,b]^{2n-3}} \bar{\pi}_n(\tau,l)\zeta(\tau,l)\, dl$$

where $V_n = (2n-3)!!$ is the number of tree topologies with n taxa, and $\zeta(\tau,l)$ denotes the average of branch lengths of the tree (τ,l).

In other words, we assume that on average (with respect to the posterior distribution), the average branch lengths of the trees are bounded from below by c. While the condition is technical, it can be verified in many realistic cases in phylogenetic inference. For example, in the absence of data, the posterior is just the prior, and Assumption 4 holds for any prior that put equal weights on topologies and selects branch lengths from a distribution with nonzero mean. Similarly, if the target distributions concentrate on a single "correct" tree generated from a Yule process with rate parameter λ, then the average branch length converges to $1/\lambda$.

As we discussed above, to quantify the efficiency of OPSMC, we need to provide some estimates on how the target distribution changes with newly added sequences. We will discuss in detail below that the most extreme changes on the target distributions happen when some of the target distribution concentrates on a set of trees with short edges, for which a single new sequence can shake up all previous beliefs about the phylogenies of interest. Assumption 4 rules out such pathological behaviors and provides a foundation for analyzing the ESS, although it does impose some limitations. For instance, if we think of a random process generating n taxon trees in a fixed time (e.g., since the dawn of life on earth) then for most natural models, the average branch length might go to zero as n goes to infinity. However, we note that both Assumptions 3 and 4 can be further relaxed to address more general cases (details in the section "Effective sample sizes of online phylogenetic SMC").

Throughout the article, we will investigate two different classes of sampling schemes: length-based proposals and likelihood-based proposals.

Length-based proposals For length-based proposals:

1. the edge e is chosen from a multinomial distribution weighted by length of the edges,
2. the distal position x is selected from a distribution $P_X^e(x)$ (with density p_X^e) across the edge length, and

3. the pendant length y is sampled from a distribution $P_Y(y)$ (with density p_Y) with support contained in $[0,b]$.

For example, if these P distributions are uniform, we obtain a uniform prior on attachment locations across the tree.

We assume that

Assumption 5. *The densities p_X^e of the distal position on edge e and p_Y of the pendant edge lengths are continuous on $[0,l_e]$ and $[0,b]$, respectively and satisfy*

$$\frac{1}{l_e^2}\int_0^{l_e}\frac{1}{p_X^e(x)}\,dx \le C \qquad \text{and} \qquad \int_0^b \frac{1}{p_Y(y)}\,dy < \infty,$$

where l_e denotes the length of edge e and C is independent of l_e.

One way to look at this assumption is that it is designed to guarantee Criterion 1: given two trees r and r' such that r can be obtained from r' by removing the taxon X_{n+1} and its corresponding edge, we want to guarantee that $Q^n(r,r')$ is positive. In the length-based proposal, this corresponds to guaranteeing that the proposal does not "miss" any possible choice of the edge, the distal position, and the pendant edge length, and one way to do so is to assume that the densities p_X^e and p_Y are bounded away from zero. The assumption is clearly implied by (and so is weaker than) a uniform positive lower bound on the densities p_X^e and p_Y. The constant C in the assumption quantifies how "spread out" the distribution is on $[0,l_e]$ and plays an important role in the efficiency of the algorithm.

Example 6. *For any density function ξ on $[0,1]$ such that $1/\xi$ is integrable, the family of proposals $\xi^e(x)=\frac{1}{l_e}\xi(\frac{x}{l_e})$ satisfies Assumption 5.*

Likelihood-based proposals We denote by $T(r,e,x,y)$ the tree obtained by adding an edge of length y to edge e of the tree r at distal position x. Any tree t can be represented by $t=T(\varrho(t),e(t),x,y)$, where $e(t)$ is the edge on which the pendant edge containing the most recent taxon is attached. In the likelihood-based approach, the edge e (from the tree r) is chosen from a multinomial distribution weighted by a user-defined likelihood-based utility function $f_n(r,e)$. Likelihood-based proposals are informed by data and are capable of capturing the posterior distribution more efficiently, but with an additional cost for computing the likelihoods.

In a Bayesian inference framework, the most natural utility function is the average likelihood utility function

$$\mathcal{G}_n(r,e)=\int_{x,y}\hat{\pi}_{n+1}(T(r,e,x,y))\,dx\,dy.$$

In this setting, for a fixed tree r, the utility function on each edge e of the tree is computed by $f_n(r,e)=\mathcal{G}_n(r,e)$. The utility quantifies the expected value of the unnormalized posterior if we indeed attach the new pendant edge to edge e. The sampler then uses these weights to choose the edge to attach the new sequence. By doing so, the proposals obtain a set of "better" trees that are more likely to survive in the subsequent generations.

Many other likelihood-based utility functions (that are quicker to compute) can be defined. We will assume that the likelihood-based utility function $f_n(r,e)$ satisfies the following assumption.

Assumption 7. *There exist $c_1,c_2>0$ independent of n,r,e such that*

$$c_1\mathcal{G}_n(r,e)\le f_n(r,e)\le c_2\mathcal{G}_n(r,e)$$

for all r,e.

This assumption ensures that the sampler is efficient as long as the utility function $f_n(r,e)$ is "comparable" to the average likelihood utility function up to some multiplicative constants c_1 and c_2. By analyzing OPSMC under this assumption rather than average likelihood utility function itself, we have the option to choose other (cheaper) utility functions for the proposal, the most important of which is the maximum *a posteriori* (MAP) utility function

$$M_n(r,e)=bl_e\sup_{x,y}\hat{\pi}_{n+1}(T(r,e,x,y)).$$

Here, we approximate an integral by the MAP multiplied by the area of integration (in this case, bl_e). This utility function can be computed via a simple optimization procedure that can be done quickly and efficiently in computational phylogenetics software and avoids the burden of sampling the posterior distribution required to compute the average likelihood utility function. The following lemma (proven in the Appendix) establishes that the MAP utility function also satisfies Assumption 7.

Lemma 8. *There exists $c_3>0$ independent of n,r,e such that*

$$\mathcal{G}_n(r,e)\le M_n(r,e)\le c_3\mathcal{G}_n(r,e)$$

for all r,e.

In a similar manner, the distributions $P_X^e(x)$ and $P_Y(y)$ might also be guided by information about the likelihood function. As for the length-based proposal, we assume the following conditions on the distal position and pendant edge length proposals for the likelihood-based approach.

Assumption 9. *The densities p_X^e and p_Y are absolutely continuous with respect to the Lebesgue measure on $[0,l_e]$ and $[0,b]$, respectively. Moreover, there exists a_0 independent of n such that*

$$\sup_{x,y}\frac{1}{p_X^e(x)}\,\frac{1}{p_Y(y)}\le a_0.$$

This condition plays the same roles as Assumption 5 in the length-based proposal: to guarantee Criterion 1 by ensuring that the densities p_X^e and p_Y are bounded away from zero. Since the likelihood-based proposals are more difficult to analyze than the length-based one, Assumption 9 is a bit stronger than Assumption 5. However, we note that the density functions from Example 6 satisfy both assumptions.

Consistency of Online Phylogenetic SMC

We recall that the OPSMC sampler maintains a list of K_n particles $t_1^n,\ldots,t_{K_n}^n$ with a positive weight w_i^n associated with each particle t_i^n. We would like to evaluate whether a sample of particles provides a good approximation to the posterior distribution in the sense of (1), which is made rigorous in measure-theoretic terms as follows. We form the normalized empirical measure $\bar{\pi}_{n,K_n}$ and define the integral $\bar{\pi}_{n,K_n}(\phi)$ of a test function ϕ with respect to this measure by

$$\bar{\pi}_{n,K_n}(\phi):=\sum_{i=1}^{K_n}\bar{\pi}_{n,K_n}(t_i^n)\phi(t_i^n)=\sum_{i=1}^{K_n}w_i^n\phi(t_i^n)\Bigg/\sum_{i=1}^{K_n}w_i^n.$$

We will show that the normalized empirical measure $\bar{\pi}_{n,K_n}$ converges weakly to $\bar{\pi}_n$, that is the integral $\bar{\pi}_{n,K_n}(\phi)$ of a test function ϕ with respect to this measure converges to the integral of ϕ with respect to the posterior distribution with probability 1.

In this section, we will demonstrate this weak convergence by induction on the number of taxa n; that is, for every $n<N$, assuming that $\bar{\pi}_{n,K_n}\to\bar{\pi}_n$, we will prove that $\bar{\pi}_{n+1,K_{n+1}}\to\bar{\pi}_{n+1}$ (where $\to$ means weak convergence). We note that although the measures mentioned above are indexed by K_n, they implicitly depend on the number of particles from the previous generations. Thus, the convergence should be interpreted in the sense of when the number of particles of all generations approach infinity.

We note that when $n=0$, the set of all rooted trees with no taxa consists of a single tree σ. Thus, if we use this single tree as the ensemble of particles at $n=0$, then $\bar{\pi}_{0,K_0}$ is precisely $\bar{\pi}_0$. Alternatively, we can start with $n=n_0\in[1,N]$ and use an MCMC method to create an ensemble of particles with stationary distribution $\bar{\pi}_{n_0}$. In either case, an induction argument gives the main theorem:

Theorem 10 (Consistency). *If Criteria 1 and 2 are satisfied and the sampler starts at $n=n_0=0$ by a list consisting of a single rooted tree with no taxa, or at $n=n_0\in[1,N]$ with an ensemble of particles created by an ergodic MCMC method with stationary distribution $\bar{\pi}_{n_0}$, then $\bar{\pi}_{n,K_n}$ converges weakly to the posterior π_n. That is, for every integrable test function $\phi:\mathcal{T}_n\to\mathbb{R}$ and $n_0\le n\le N$*

$$\sum_{i=1}^{K_n}\bar{\pi}_{n,K_n}(t_i^n)\phi(t_i^n)\to\frac{1}{V_n}\sum_{\tau}\int_{[0,b]^{2n-3}}\bar{\pi}_n(\tau,l)\phi(\tau,l)dl \qquad a.s.$$

as $K_{n_0},K_{n_0+1},\ldots,K_n\to\infty$, where a.s. denotes almost sure convergence (i.e., convergence with probability 1).

Theorem 10 shows that approximation by the OPSMC sampler becomes arbitrarily good as the number of particles goes to infinity.

Remark 11. *Theorem 10 describes the asymptotic behavior of OPSMC in the limit when the number of particles of all generations (up to generation n) approaches infinity, and guarantees that the algorithm is consistent regardless of the relative rates with which the K_i approach infinity. For example, in the traditional setting when the number of particles are the same in every generation, that is, $K_1=K_2=\ldots=K_n=K$, we also have*

$$\sum_{i=1}^{K}\bar{\pi}_{n,K}(t_i^n)\phi(t_i^n)\to\frac{1}{V_n}\sum_{\tau}\int_{[0,b]^{2n-3}}\bar{\pi}_n(\tau,l)\phi(\tau,l)dl \qquad a.s.$$

as $K\to\infty$.

However, it is worth pointing out that because the sampler is built sequentially, to make a prediction with a given accuracy at generation n with finite data, we need to control the number of particles from all previous generations. These issues will be further addressed in the latter part of the article, where we uniformly bound the ESS of the sampler as the generations proceed.

Characterizing Changes in the Likelihood Landscapes When new sequences arrive

Although the consistency of OPSMC is guaranteed, and informative OPSMC samplers can be developed by changing the Markov transition kernels, its applicability is constrained by an implicit assumption: the distance between target distributions of consecutive generations is not too large. Since SMC methods are built upon the idea of recycling particles from one generation to explore the target distribution of the next generation, it is obvious that one would never be able to design an efficient SMC sampler if $\bar{\pi}_n$ and $\bar{\pi}_{n+1}$ are very different from one another.

While a condition on minor changes in the target distributions may be easy to verify in some applications, it is not straightforward in the context of phylogenetic inference. A similar question on how the "optimal" trees (under some appropriate measure of optimality) change has been studied extensively in the field, with negative results for almost all regular measures of optimality (Heath et al., 2008; Cueto and Matsen, 2011). However, to the best of our knowledge, no work has been done detailing how phylogenetic likelihood landscapes change when new sequences arrive.

In this section, we will establish that under some minor regularity conditions on the distribution described in the previous sections, the relative changes between target distributions from consecutive generations are uniformly bounded. This result enables us to provide a lower bound on the effective sample size of OPSMC algorithms in the next section.

Lemma 12. *Consider an arbitrary tree $t\in\mathcal{T}_{n+1}$ obtained from the parent tree $\varrho(t)$ by choosing edge e, distal position x and pendant length y. Denote*

$$M(y)=\max_{ij} G_{ij}(y), \qquad m(y)=\min_{ij} G_{ij}(y),$$

and

$$\mathcal{Z}_n=\frac{1}{V_n}\sum_{\tau}\int_{[0,b]^{2n-3}}\bar{\pi}_n(\tau,u)\zeta(\tau,u)\,du.$$

Recalling that our sequences are of length S and assuming that $A_1\le\frac{\pi_0^{(n+1)}(t)}{\pi_0^{(n)}(\rho(t))}\le A_2$ for all $t\in\mathcal{T}_{n+1}$, we have

$$\frac{\bar{\pi}_{n+1}(t)}{\bar{\pi}_n(\varrho(t))}\le\frac{A_2}{A_1}\frac{1}{\mathcal{Z}_n}\frac{M(y)^S}{\int_0^b m(y)^S dy},\qquad \forall t\in\mathcal{T}_{n+1}.$$

Sketch of proof. Using the 1D formulation of the phylogenetic likelihood function derived in Dinh and Matsen (2016), we can prove that

$$\frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))}=\frac{\pi_0^{(n+1)}(t)}{\pi_0^{(n)}(\rho(t))}\frac{L_{n+1}(t)}{L_n(\varrho(t))}\le A_2 M(y)^S,\qquad \forall t\in\mathcal{T}_{n+1}. \tag{3}$$

Similarly, we have $\hat{\pi}_{n+1}(t)/\hat{\pi}_n(\varrho(t))\ge A_1 m(y)^S$ for all $t\in\mathcal{T}_{n+1}$.

For any tree (τ_{n+1},u_{n+1}) with $(n+1)$ taxa, we let $(\tau_n,u_n)=\varrho((\tau_{n+1},u_{n+1}))$ and denote by $e,x,$ and y the chosen edge for attaching the last taxon, the distal length, and the length of the pendant edge, respectively. We deduce that

$$\begin{aligned}
&\|\hat{\pi}_{n+1}\|\\
&\quad=\frac{1}{V_{n+1}}\sum_{\tau_{n+1}}\int_{[0,b]^{2n-1}}\hat{\pi}_{n+1}(\tau_{n+1},u_{n+1})du_{n+1}\\
&\quad\ge\frac{1}{V_{n+1}}\sum_{\tau_{n+1}}\int_{[0,b]^{2n-1}}A_1 m(y)^S\hat{\pi}_n(\varrho(\tau_{n+1},u_{n+1}))du_{n+1}\\
&\quad=\frac{A_1}{V_{n+1}}\sum_{\tau_{n+1}}\int_{[0,b]^{2n-3}}\int_{x,y}m(y)^S\hat{\pi}_n((\tau_n,u_n))dx\,dy\,du_n.
\end{aligned}$$

Recall that $\zeta(r)$ is the average branch length of r. Using the fact that for a fixed tree r, $\int_0^{l_e}dx=l_e$ and $\sum_e l_e=(2n-3)\zeta(r)$, we have

$$\begin{aligned}
&\|\hat{\pi}_{n+1}\|\\
&\quad\ge\frac{A_1}{V_{n+1}}\left(\int_0^b m(y)^S dy\right)\sum_{\tau_n,e}\int_{[0,b]^{2n-3}}\hat{\pi}_n(\tau_n,u_n)\int_0^{l_e}dx\,du_n\\
&\quad\ge\frac{A_1}{V_{n+1}}\left(\int_0^b m(y)^S dy\right)\sum_{\tau_n}\int_{[0,b]^{2n-3}}\hat{\pi}_n(\tau_n,u_n)\sum_e l_e\,du_n\\
&\quad=\frac{(2n-3)V_nA_1}{V_{n+1}}\left(\int_0^b m(y)^S dy\right)\\
&\qquad\times\left(\frac{1}{V_n}\sum_{\tau_n}\int_{[0,b]^{2n-3}}\hat{\pi}_n(\tau_n,u_n)\zeta(\tau_n,u_n)du_n\right).
\end{aligned}$$

Recalling that $V_{n+1}=(2n-3)V_n$, we deduce that

$$\begin{aligned}
\frac{\bar{\pi}_{n+1}(t)}{\bar{\pi}_n(\varrho(t))}&=\frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))}\frac{\|\hat{\pi}_n\|}{\|\hat{\pi}_{n+1}\|}\\
&\le\frac{A_2}{A_1}\frac{1}{\mathcal{Z}_n}\frac{M(y)^S}{\int_0^b m(y)^S\,dy},\qquad \forall t\in\mathcal{T}_{n+1}.
\end{aligned}$$

This completes the proof. □

The main idea of Lemma 12 is to bound the posterior values of a tree t by that of its parent $\varrho(t)$ with an explicit constant independent of n. While the proof of the lemma is technical, its main insights can be explained from the observation that the most extreme changes of the posteriors happen when, either (1) the new pendant edge is small, or (2) the edge lengths of the parent tree $\rho(t)$ are small.

Indeed, in the simplest case where both priors on the sets of trees with n and $(n+1)$-taxa are uniform, we can choose $A_1=A_2$. Equation (3) and the subsequent equation establish that the ratio between the likelihood values of tree t and its parent $\rho(t)$ can be bounded from above as long as the length of the pendant edge is not too small. Similarly, the ratio between the total masses of $\hat{\pi}_{n+1}$ and $\hat{\pi}_n$ can also be controlled if the target distribution $\bar{\pi}_{n+1}$ does not concentrate on a set of trees with short edges. This motivates Assumption 4 to provide a lower bound on the total edge lengths and to rule out such pathological behaviors.

ESSs of Online Phylogenetic SMC

In this section, we are interested in the asymptotic behavior of OPSMC in the limit of large K_n, that is, when the number of particles of the sampler approaches infinity. This asymptotic behavior is illustrated via estimates of the ESS of the sampler with large numbers of particles. We note that although there are several studies on the stability of SMC as the generations proceed, most of them focus on cases where the sequence of target distributions have a common state space of fixed dimension (Del Moral, 1998; Douc and Moulines, 2008; Künsch, 2005; Oudjane and Rubenthaler, 2005; Del Moral et al., 2009; Beskos et al., 2014). In general, establishing stability bounds for SMC requires imposing some conditions on the effect of data at any generation k to the target distribution at generation $n\ge k$ (Crisan and Doucet, 2002; Chopin, 2004; Doucet and Johansen, 2009). Lemma 12 helps validate a condition of this type.

In this section, we use the ESS (Kong et al., 1994; Liu and Chen, 1995) of the particles at generation $n+1$

as a measure of the sampler's efficiency. The ESS is computed as

$$\mathrm{ESS}_{n+1} = \frac{\left(\sum_{i=1}^{K} w_i^{n+1}\right)^2}{\sum_{i=1}^{K} (w_i^{n+1})^2}.$$

The detailed derivation of the formula is provided in Kong et al. (1994) and Liu and Chen (1995), but a simple intuition is as follows: if the weights of the particles are roughly of the same fitness (weight), then the ESS is equal to the number of particles; on the other hand, if M of the particles share almost all of the weight equally, $\mathrm{ESS} \approx M$. This formulation of the ESS is usually used to do adaptive resampling, whereas some additional resampling steps are done when the ESS drops below certain threshold. We emphasize that as with other measures of efficiency of MCMC methods, good ESS is necessary but not sufficient to ensure a good quality posterior. From the definition, it is also clear that the ESS could not exceed the number of particles.

The following result, proven in the Appendix, enables us to estimate the asymptotic behavior of the sample's ESS in various settings.

Theorem 13. *In the limit as the number of particles approaches infinity, we have*

$$\lim_{K_{n+1}\to\infty} \frac{K_{n+1}}{ESS_{n+1}} = \frac{1}{V_{n+1}} \sum_{\tau} \int_{[0,b]^{2n-1}} \frac{\bar{\pi}_{n+1}^2(\tau,l)}{\bar{\pi}_n(\varrho(\tau,l))\; Q^n(\varrho(\tau,l),(\tau,l))}\, dl$$

with probability 1.

This asymptotic estimate and the results on likelihood landscapes from the previous section allow us to prove the following Theorem (see Appendix for proof).

Theorem 14 (ESS of OPSMC for likelihood-based proposals)**.** *If Assumptions 3, 4, 7, and 9 hold, then with probability 1, there exists* $\alpha \in (0,1]$ *independent of n such that* $ESS_n \geq \alpha K_n$*. That is, the effective sample size of an OPSMC sampler with likelihood-based proposals is bounded below by a constant multiple of the number of particles. Moreover, if Assumption 4 does not hold, the ESS of OPSMC algorithms decays at most linearly as the dimension (the number of taxa) increases.*

We also have similar estimates for length-based proposals (see Appendix for proof):

Theorem 15 (ESS of OPSMC for length-based proposals)**.** *If Assumptions 3, 5, and 4 hold, then with probability 1, the ESS of OPSMC with length-based proposals are bounded below by a constant multiple of the number of particles. Moreover, if Assumption 4 does not hold, the ESS of OPSMC algorithms decays at most quadratically as the dimension (the number of taxa) increases.*

In summary, we are able to prove that in many settings, the ESS of OPSMC is bounded from below. These results are surprising, since in the general case it is known that SMC-type algorithms may suffer from the curse-of-dimensionality: when the dimension of the problem increases, the number of the particles must increase exponentially to maintain a constant ESS (Chopin, 2004; Bengtsson et al., 2008; Bickel et al., 2008; Snyder et al., 2008). We further note that although the Markov transition kernels P^n after the mutation step are not involved in the theoretical analysis of the ESS in this section, the results hold true for all kernels that satisfy Criterion 2.

Discussion

In this article, we establish foundations for OPSMC, including essential theoretical convergence results. We prove that under some mild regularity conditions and with carefully constructed proposals, the OPSMC sampling algorithm is consistent. This includes relaxing the condition used in (Bouchard-Côté et al., 2012), in which the authors assume that the weight of the particles are bounded from above. We then investigate two different classes of sampling schemes for online phylogenetic inference: length-based proposals and likelihood-based proposals. In both cases, we show the ESS to be bounded below by a multiple of the number of particles.

The consistency and convergence results in this article apply to a variety of sampling strategies. One possibility would be for an algorithm to use a large number of particles, directly using the SMC machinery to approximate the posterior. Alternatively, the SMC part of the sampler could be quite limited, resulting in an algorithm which combines many independent parallel MCMC runs in a principled way. As described above, the SMC portion of the algorithm enables MCMC transition kernels that would normally be disallowed by the requirement of preserving detailed balance. For example, one could use a kernel that focuses effort around the part of the tree which has recently been disturbed by adding a new sequence.

In the future we will develop efficient and practical implementations of these ideas, and a first step in this direction has already been made (Fourment et al., 2017). Many challenges remain. For example, the exclusive focus of this article has been on the tree structure, consisting of topology, and branch lengths. However, Bayesian phylogenetics algorithms typically coestimate mutation model parameters along with the tree structures. Although proposals for other model parameters can be obtained by particle MCMC (Andrieu et al., 2010), we have not attempted to incorporate it into the current SMC framework. In addition, we note that the input for this type of phylogenetics algorithm consists of a *multiple sequence alignment* (MSA) of many sequences, rather than just individual sequences themselves. This raises the question of how to maintain an up-to-date

MSA. Programs exist to add sequences into existing MSAs (Caporaso et al., 2010; Katoh and Standley, 2013), although from a statistical perspective, it could be preferable to jointly estimate a sequence alignment and tree posterior (Suchard and Redelings, 2006). It is an open question how that could be done in an online fashion, although in principle it could be facilitated by some modifications to the sequence addition proposals described here.

Funding

This work was funded by National Science Foundation grants [DMS-1223057, CISE-1564137 to V.D. and F.A.M.]; National Institutes of Health grant [U54-GM111274 to V.D. and F.A.M.]; and Faculty Scholar grant from the Howard Hughes Medical Institute and the Simons Foundation to F.A.M.

Appendix

Notation and variables

t_i^n	Particle in generation n, weighted by w_i^n
s_i^n	Particle obtained after the "resampling step," unweighted
s_i^n	Particle obtained after the "Markov transition step," unweighted
t_i^{n+1}	Particle in generation $(n+1)$, obtained after the "mutation step," weighted by w_i^{n+1}
$\hat{\pi}_{n,K_n}$	Unnormalized measure, induced by the particles $\{t_i^n\}$
$\hat{\alpha}_{n,K_{n+1}}$	Unnormalized measure, induced by the particles $\{s_i^n\}$
$\hat{\beta}_{n,K_{n+1}}$	Unnormalized measure, induced by the particles $\{m_i^n\}$
$\hat{\lambda}_{n,K_{n+1}}$	Unnormalized measure, induced by the particles $\{t_i^{n+1}\}$
$\bar{\pi}_n$	The posterior distribution at generation n
Q^n	The proposal distribution at generation n
P^n	The Markov transition kernel at generation n
$p_X^e(x)$	Proposal distribution of the distal position
$p_Y(y)$	Proposal distribution of the pendant edge length

Integration on $\mathcal{T}_n$

As described in the introduction, to do integration on $\mathcal{T}_n$, the product space of the space of all possible tree topologies and the space of all branch lengths $[0,b]^{2n-3}$, we consider the corresponding product measure, $\mu_n(dr)$, that is uniform on trees and Euclidean on each branch length space. Integration using this measure is then just the average of the integrals for each topology:

$$\int_{r\in\mathcal{T}_n}\phi(r)\ \mu_n(dr) := \frac{1}{V_n}\sum_{\tau}\int_{[0,b]^{2n-3}}\phi(\tau,l)dl$$

for any n and test function ϕ defined on $\mathcal{T}_n$, where $V_n = (2n-3)!!$ is the number of different topologies of $\mathcal{T}_n$.

Consistency of Online Phylogenetic SMC

For clarity, we restate below the main steps in the OPSMC and introduce some important notations. We recall that at the beginning of each generation n, a list of K_n particles $t_1^n,\dots,t_{K_n}^n$ are maintained along with a positive weight w_i^n associated with each particle t_i^n. These weighted particles form an unnormalized measure and a corresponding normalized empirical measure

$$\hat{\pi}_{n,K_n} = \sum_{i=1}^{K_n} w_i^n\delta_{t_i^n}(\cdot) \ \text{ and } \ \bar{\pi}_{n,K_n} = \frac{1}{\sum_{i=1}^{K_n} w_i^n}\hat{\pi}_{n,K_n}$$

such that $\bar{\pi}_{n,K_n}$ approximates $\bar{\pi}_n$. A new list of K_{n+1} particles is then created in three steps: resampling, Markov transition and mutation.

Resampling step (the "resampling" step illustrated in Fig. 1).

A total of K_{n+1} particles are sampled from the distribution $\bar{\pi}_{n,K_n}$, and after resampling we obtain the unweighted measure

$$\hat{\alpha}_{n,K_{n+1}} = \sum_{i=1}^{K_n} K_{n+1,i}\delta_{t_i^n}(\cdot),$$

where $K_{n+1,i}$ is the multiplicity of particle t_i^n (i.e., the number of times t_i^n arose in the sample), sampled from a multinomial distribution parameterized by the weights w_i^n. We denote the particles obtained after this step by s_i^n.

Markov transition (the "MH moves" step illustrated in Fig. 1). MCMC steps can be run separately on each particle s_i^n for a certain amount of time to obtain a new approximately independent sample m_i^n with (unweighted) measure denoted $\hat{\beta}_{n,K_{n+1}}$.

Mutation step (the "propose addition of new sequence" step illustrated in Fig. 1).

In the mutation step, new particles $t_1^{n+1},\dots,t_{K_{n+1}}^{n+1}$ are created from a proposal distribution Q^n and are weighted by the weight function h defined by equation (2). The new particles $t_1^{n+1},\dots,t_{K_{n+1}}^{n+1}$ with their corresponding weights now represent the distribution $\bar{\pi}_{n+1}$ and act as the input for the next generation.

The proposal Q^n is assumed to be normalized, and the unnormalized measure over the particles t_i^{n+1} can be computed by

$$\hat{\lambda}_{n,K_{n+1}}(A) = \sum_{i=1}^{K_{n+1}} Q^n(m_i^n, A)\ \hat{\beta}_{n,K_{n+1}}(m_i^n).$$

for any measurable set $A\subset\mathcal{T}_{n+1}$.

The process is then iterated until $n=N$. For convenience, we will denote the unnormalized empirical measures of the particles right after generation n by $\hat{\alpha}_{n,K_{n+1}}$, $\hat{\beta}_{n,K_{n+1},}$ and $\hat{\lambda}_{n,K_{n+1}}$, respectively. Similarly, the corresponding normalized distributions will be denoted by $\bar{\alpha}_{n,K_{n+1}}$, $\bar{\beta}_{n,K_{n+1},}$ and $\bar{\lambda}_{n,K_{n+1}}$.

For convenience, let L and K be the number of particles at the nth and $(n+1)$st generations, respectively. Recall that the normalized distributions after the substeps of OPSMC are denoted by $\bar{\alpha}_{n,K}$, $\bar{\beta}_{n,K}$, and $\bar{\lambda}_{n,K}$, we have the following lemma.

Lemma 16. *Assume that Criteria 1 and 2 are satisfied. If we define*

$$\bar{\lambda}_n(t) := \bar{\pi}_n(\varrho(t))Q^n(\varrho(t),t) \qquad \textit{and}$$

$$h(t) := \frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))\; Q^n(\varrho(t),t)}$$

then the following statements hold, where $\rightarrow$ *means weak convergence*

1. *If* $\bar{\pi}_{n,L} \rightarrow \bar{\pi}_n$, *then* $\bar{\alpha}_{n,K} \rightarrow \bar{\pi}_n$.
2. *If* $\bar{\alpha}_{n,K} \rightarrow \bar{\pi}_n$, *then* $\bar{\beta}_{n,K} \rightarrow \bar{\pi}_n$.
3. *If* $\bar{\beta}_{n,K} \rightarrow \bar{\pi}_n$, *then* $\bar{\lambda}_{n,K} \rightarrow \bar{\lambda}_n$.
4. $h(t)\bar{\lambda}_n(t)$ *is proportional to* $\bar{\pi}_{n+1}(t)$.
5. *If* $\bar{\lambda}_{n,K} \rightarrow \bar{\lambda}_n$, *then* $\bar{\pi}_{n+1,K} \rightarrow \bar{\pi}_{n+1}$.

Proof of Lemma 16. (1). Assume that $\bar{\pi}_{n,L}$ converges to $\bar{\pi}_n$ a.s.,

$$|\bar{\alpha}_{n,K}(\phi) - \bar{\pi}_n(\phi)| \leq \left| \frac{1}{K}\sum_{i=1}^{L} K_{n+1,i}\; \phi(t_i^n) - \frac{1}{\|w\|}\sum_{i=1}^{L} w_i\; \phi(t_i^n) \right|$$
$$+ |\bar{\pi}_{n,L}(\phi) - \bar{\pi}_n(\phi)|$$

where here $\|w\|$ denotes the total mass of the empirical measure w (i.e., the sum of all of all particle weights). Since the particles in generation $(n+1)$ are sampled independently from a multinomial distribution with fixed weights, by the strong law of large numbers, we have

$$\lim_{K\to\infty} \frac{K_{n+1,i}}{K} = \frac{w_i}{\|w\|} \qquad \forall i = 1,\ldots,L.$$

This implies

$$\limsup_{K\to\infty} |\bar{\alpha}_{n,K}(\phi) - \bar{\pi}_n(\phi)| \leq |\bar{\pi}_{n,L}(\phi) - \bar{\pi}_n(\phi)| \qquad \textit{a.s.}$$

This implies that when $K, L \to \infty$, we have $\bar{\alpha}_{n,K}(\phi) \to \bar{\pi}_n(\phi)$ a.s..

(2). The rationale behind the use of MCMC moves is based on the observation that if the unweighted particles are distributed according to $\bar{\pi}_n$, then when we apply a Markov transition kernel P of invariant distribution $\bar{\pi}_n$ to any particle, the new particles are still distributed according to the posterior distribution of interest.

Formally, if $\bar{\alpha}_{n,K}(\phi) \to \bar{\pi}_n(\phi)$ a.s. for every integrable test function $\phi : \mathcal{T}_n \to \mathbb{R}$, by choosing $\phi = P^r(\cdot, A)$ for any measurable set $A \subset \mathcal{T}_{n+1}$, we deduce that

$$\bar{\beta}_{n+1,K}(A) = \sum_{i=1}^{K} P^r(s_i^n, A)\; \bar{\alpha}_{n,K}(s_i^n)$$
$$\xrightarrow{K\to\infty} \int_{\mathcal{T}_n} P^r(s,A)\; \bar{\pi}_n(s)\; \mu_n(ds) = \bar{\pi}_n(A)$$

a.s., since the Markov kernel P is invariant with respect to $\bar{\pi}_n$. Therefore, for any measurable function $\phi : \mathcal{T}_n \to \mathbb{R}$, we have that $\bar{\beta}_{n,K}(\phi)$ converges to $\bar{\pi}_n(\phi)$ almost surely.

(3). Since $\bar{\beta}_{n,K}(\phi) \to \bar{\pi}_n(\phi)$ a.s. for every integrable test function $\phi : \mathcal{T}_n \to \mathbb{R}$, by choosing $\phi = Q^n(\cdot, A)$ for any measurable set $A \subset \mathcal{T}_{n+1}$, we deduce that

$$\bar{\lambda}_{n,K}(A) = \sum_{i=1}^{K} Q^n(m_i^n, A)\bar{\beta}_{n,K}(m_i^n)$$
$$\xrightarrow{K\to\infty} \int_{\mathcal{T}_n} Q^n(m,A)\bar{\pi}_n(dm)$$
$$= \int_A \bar{\pi}_n(\varrho(t))Q^n(\varrho(t),t)\mu_{n+1}(dt) = \bar{\lambda}_n(A).$$

Therefore, for any measurable function $\phi : \mathcal{T}_{n+1} \to \mathbb{R}$, we also have $\bar{\lambda}_{n,K}(\phi)$ converges to $\bar{\lambda}_n(\phi)$ almost surely.

(4) We note that

$$\begin{aligned} h(t)\bar{\lambda}_n(t) &= h(t)\bar{\pi}_n(\varrho(t))Q^n(\varrho(t),t) \\ &= \frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))Q^n(\varrho(t),t)}\frac{1}{\|\hat{\pi}_n\|}\hat{\pi}_n(\varrho(t))Q^n(\varrho(t),t) \\ &= \frac{1}{\|\hat{\pi}_n\|}\hat{\pi}_{n+1}(t). \end{aligned} \tag{A.1}$$

(5). Since the proposal Q^n and the Markov kernel P are assumed to be normalized, we have $\|\hat{\lambda}_{n,K}\| = \|\hat{\beta}_{n,K}\| = \|\hat{\alpha}_{n,K}\| = K$.

We have,

$$\frac{1}{K}\|\hat{\pi}_{n+1,K}\|$$
$$= \frac{1}{\|\hat{\lambda}_{n,K}\|}\sum_{i=1}^{K}\hat{\pi}_{n+1,K}(t_i^{n+1}) = \sum_{i=1}^{K} h(t_i^{n+1})\bar{\lambda}_{n,K}(t_i^{n+1}) \tag{A.2}$$
$$\xrightarrow{K\to\infty} \int_{\mathcal{T}_{n+1}} h(t)\bar{\lambda}_n(t)\mu_{n+1}(dt)$$
$$= \frac{1}{\|\hat{\pi}_n\|}\int_{\mathcal{T}_{n+1}} \hat{\pi}_{n+1}(t)\mu_{n+1}(dt) = \frac{\|\hat{\pi}_{n+1}\|}{\|\hat{\pi}_n\|} \qquad \textit{a.s.}$$

By a similar argument, we have

$$\bar{\pi}_{n+1,K}(\phi) = \frac{\frac{1}{K}\sum_{i=1}^{K}\phi(t_i^{n+1})\hat{\pi}_{n+1,K}(t_i^{n+1})}{\frac{1}{K}\|\hat{\pi}_{n+1,K}\|}$$
$$= \frac{\sum_{i=1}^{K}\phi(t_i^{n+1})h(t_i^{n+1})\bar{\lambda}_{n,K}(t_i^{n+1})}{\frac{1}{K}\|\hat{\pi}_{n+1,K}\|}$$

$$\xrightarrow{K\to\infty} \frac{\|\hat{\pi}_n\|}{\|\hat{\pi}_{n+1}\|}\int_{\mathcal{T}_n}\phi(t)h(t)\bar{\lambda}_n(t)\mu_{n+1}(dt)$$

$$=\frac{\|\hat{\pi}_n\|}{\|\hat{\pi}_{n+1}\|}\int_{\mathcal{T}_n}\phi(t)\frac{\hat{\pi}_{n+1}(t)}{\|\hat{\pi}_n\|}\mu_{n+1}(dt)=\bar{\pi}_{n+1}(\phi).$$

In other words, $\bar{\pi}_{n+1,K}$ converges to $\bar{\pi}_{n+1}$. □

Proof of Theorem 13. By definition, we have

$$h(t)=\frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))\,Q^n(\varrho(t),t)}, \qquad \text{and}$$

$$w_i^{n+1}=\hat{\pi}_{n+1,K_{n+1}}(t_i^{n+1})=h(t_i^{n+1}).$$

Thus,

$$\bar{\pi}_{n+1,K_{n+1}}(h)=\frac{\sum_{i=1}^{K_{n+1}}h(t_i^{n+1})\hat{\pi}_{n+1,K}(t_i^{n+1})}{\|\hat{\pi}_{n+1,K_{n+1}}\|}$$

$$=\frac{\sum_{i=1}^{K_{n+1}}(w_i^{n+1})^2}{\sum_{i=1}^{K_{n+1}}w_i^{n+1}}$$

$$=\frac{\sum_{i=1}^{K_{n+1}}w_i^{n+1}}{\text{ESS}_{n+1}}=\frac{\|\hat{\pi}_{n+1,K_{n+1}}\|}{\text{ESS}_{n+1}}.$$

On the other hand, by applying Theorem 10 for $\phi\equiv h$, we have

$$\bar{\pi}_{n+1,K_{n+1}}(h)$$

$$\to \bar{\pi}_{n+1}(h)$$

$$=\frac{\|\hat{\pi}_{n+1}\|}{\|\hat{\pi}_n\|}\int_{t\in\mathcal{T}_{n+1}}\frac{\bar{\pi}^2_{n+1}(t)}{\hat{\pi}_n(\varrho(t))\,Q^n(\varrho(t),t)}\,\mu_{n+1}(dt) \qquad a.s.,$$

which completes the proof via the convergence result (A.2). □

An induction argument with Lemma 16 gives the main theorem.

Other Proofs

Proof of Lemma 8. The lower bound is straightforward. For the upper bound, consider $(x,y)\in[0,l_e]\times[0,b]$ and fix $\delta>0$; by the same arguments as in the proof of Lemma 12, we have

$$\hat{\pi}_{n+1}(T(r,e,x,y))\geq m(\delta)^S\hat{\pi}_n(r) \qquad \forall y\geq\delta.$$

Thus, if we define

$$A=\{(x,y)\in[0,l_e]\times[0,b]:\hat{\pi}_{n+1}(T(r,e,x,y))\geq m(\delta)^S\hat{\pi}_n(r)\},$$

then we have $|A|\geq(b-\delta)l_e$ and

$$\mathcal{G}_n(r,e)=\int_{x,y}\hat{\pi}_{n+1}(T(r,e,x,y))\,dx\,dy$$

$$\geq\int_A\hat{\pi}_{n+1}(T(r,e,x,y))\,dx\,dy\geq(b-\delta)l_e\,m(\delta)^S\hat{\pi}_n(r).$$

On the other hand, from Lemma 12, we have $f_n(r,e)\leq b l_e\,\hat{\pi}_n(r)M(b)^S$.

By choosing $\delta=b/2$, we obtain

$$f_n(r,e)\leq 2\frac{M(b)^S}{m(b/2)^S}\,\mathcal{G}_n(r,e)$$

which completes the proof. □

Proof of Lemma 12. Let l_e be the length of the edge e and $G(\alpha)$ be the transition matrix across an edge of length α and k_u the observed value at site u of the newly added taxon. We follow the formulation of 1D phylogenetic likelihood function as in Dinh and Matsen (2016, Section 2.2) to fix all parameters except l_e and consider the likelihood of $\varrho(t)$ a function of l_e, we have

$$L_n(\varrho(t))=\prod_{u=1}^{S}\left(\sum_{ij}d_{ij}^uG_{ij}^e(l_e)\right)$$

where d_{ij}^u the probability of observing i and j at the left and right nodes of e at the site index u, respectively (note that in Dinh and Matsen 2016, it is called d_{ij}^u). Similarly, by representing the likelihood of the tree t in terms of x, y, and l_e, we have

$$L_{n+1}(t)=\prod_{u=1}^{S}\left(\sum_{ij}d_{ij}^uG_{ij}^e(l_e,x,y)\right)$$

$$=\prod_{u=1}^{S}\left(\sum_{ij}d_{ij}^u\sum_{m}G_{im}(x)G_{mj}(l_e-x)G_{mk_u}(y)\right), \tag{A.3}$$

where the indices i,j,m are looped over all possible state characters. Since $G_{mk_u}(y)\leq M(y)$ for all m and k_u, we deduce that

$$\frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))}=\frac{\pi_0^{(n+1)}(t)}{\pi_0^{(n)}(\rho(t))}\frac{L_{n+1}(t)}{L_n(\varrho(t))}\leq A_2M(y)^S, \qquad \forall t\in\mathcal{T}_{n+1}. \tag{A.4}$$

Similarly, we have $\hat{\pi}_{n+1}(t)/\hat{\pi}_n(\varrho(t))\geq A_1m(y)^S$ for all $t\in\mathcal{T}_{n+1}$.

Recall that $\zeta(r)$ is the average branch length of r. Using the fact that for a fixed tree r, $\int_0^{l_e}dx=l_e$ and $\sum_e l_e=(2n-3)\zeta(r)$, we have

$$\|\hat{\pi}_{n+1}\|=\int_{t\in\mathcal{T}_{n+1}}\hat{\pi}_{n+1}(t)\mu_{n+1}(dt)$$

$$\geq\int_{t\in\mathcal{T}_{n+1}}A_1m(y)^S\hat{\pi}_n(\varrho(t))\mu_{n+1}(dt)$$

$$=\frac{V_n}{V_{n+1}}\int_{r\in\mathcal{T}_n}\sum_e\int_{x,y}A_1m(y)^S\hat{\pi}_n(r)dxdy\,\mu_n(dr)$$

$$=\frac{(2n-3)V_nA_1}{V_{n+1}}\left(\int_0^b m(y)^S\,dy\right)\int_{r\in\mathcal{T}_n}\hat{\pi}_n(r)\zeta(r)\mu_n(dr).$$

Noting that $V_{n+1}=(2n-3)V_n$, we obtain

$$\frac{\|\hat{\pi}_{n+1}\|}{\|\hat{\pi}_n\|}\geq A_1\left(\int_0^b m(y)^S\,dy\right)\mathcal{Z}_n \qquad \text{(A.5)}$$

which implies

$$\begin{aligned}\frac{\bar{\pi}_{n+1}(t)}{\bar{\pi}_n(\varrho(t))}&=\frac{\hat{\pi}_{n+1}(t)}{\hat{\pi}_n(\varrho(t))}\,\frac{\|\hat{\pi}_n\|}{\|\hat{\pi}_{n+1}\|}\\&\leq\frac{A_2}{A_1}\,\frac{1}{\mathcal{Z}_n}\,\frac{M(y)^S}{\int_0^b m(y)^S\,dy},\qquad \forall t\in\mathcal{T}_{n+1}.\end{aligned}$$

□

Proof of Theorem 14. We recall that $T(r,e,x,y)$ denotes the tree obtained by adding an edge of length y to edge e of the tree r at distal position x. Any tree t can be represented by $t=T(\varrho(t),e(t),x,y)$, where $e(t)$ is the edge on which the pendant edge containing the most recent taxon is attached.

Define

$$f_n(r)=\sum_e f_n(r,e).$$

Since edge e is chosen from a multinomial distribution weighted by $f_n(r,e)$, given any tree $t\in\mathcal{T}_{n+1}$ obtained from the parent tree $\varrho(t)$, chosen edge $e(t)$, distal position x, and pendant length y, we have

$$Q^n(\varrho(t),t)=\frac{V_{n+1}}{V_n}\,\frac{f_n(\varrho(t),e(t))}{f_n(\varrho(t))}\,p_X(x)\,p_Y(y).$$

On the other hand, by Lemma 12 and the fact that $M(y)\leq 1$, we have

$$\frac{\bar{\pi}_{n+1}(t)}{\bar{\pi}_n(\varrho(t))}\leq\frac{A_2(n)}{A_1(n)}\,\frac{1}{\mathcal{Z}_n}\,\frac{M(y)^S}{\int_0^b m(y)^S dy}\leq\frac{1}{u_1\mathcal{Z}_n},$$

where C_0 is the constant from Assumption 3, $u_1=(1/C_0)\int_0^b m(y)^S\,dy$ and $\mathcal{Z}_n$ are defined as in Lemma 12.

These two identities and Assumption 9 imply that

$$\begin{aligned}&\int_{t\in\mathcal{T}_{n+1}}\frac{\bar{\pi}^2_{n+1}(t)}{\bar{\pi}_n(\varrho(t))\,Q^n(\varrho(t),t)}\,\mu_{n+1}(dt)\\&\quad\leq\frac{1}{u_1\mathcal{Z}_n}\,\frac{V_n}{V_{n+1}}\int_{t\in\mathcal{T}_{n+1}}\frac{f_n(\varrho(t))}{f_n(\varrho(t),e(t))}\\&\qquad\times\frac{1}{p_X(x)}\,\frac{1}{p_Y(y)}\,\bar{\pi}_{n+1}(t)\,\mu_{n+1}(dt)\\&\quad=\frac{a_0}{u_1\mathcal{Z}_n}\left(\frac{V_n}{V_{n+1}}\right)^2\frac{1}{\|\hat{\pi}_{n+1}\|}\\&\qquad\times\int_{r\in\mathcal{T}_n}\sum_e\frac{f_n(r)}{f_n(r,e)}\int_{x,y}\hat{\pi}_{n+1}(T(r,e,x,y))\,dx\,dy\,\mu_n(dr).\end{aligned}$$

Note that from Assumption 7, we have

$$f_n(r,e)\geq c_1\mathcal{G}_n(r,e)=c_1\int_{x,y}\hat{\pi}_{n+1}(T(r,e,x,y))\,dx\,dy$$

and

$$\begin{aligned}&\int_{r\in\mathcal{T}_n}f_n(r)\,\mu_n(dr)\\&\quad\leq c_2\int_{r\in\mathcal{T}_n}\sum_e\int_{x,y}\hat{\pi}_{n+1}(T(r,e,x,y))\,dx\,dy\,\mu_n(dr)\\&\quad=c_2(2n-3)\|\hat{\pi}_{n+1}\|.\end{aligned}$$

Thus

$$\begin{aligned}&\int_{t\in\mathcal{T}_{n+1}}\frac{\bar{\pi}^2_{n+1}(t)}{\bar{\pi}_n(\varrho(t))\,Q^n(\varrho(t),t)}\,\mu_{n+1}(dt)\\&\quad\leq\frac{a_0}{c_1}\,\frac{1}{u_1\mathcal{Z}_n}\,\frac{1}{\|\hat{\pi}_{n+1}\|}\left(\frac{V_n}{V_{n+1}}\right)^2\left((2n-3)\int_{r\in\mathcal{T}_n}f_n(r)\,\mu_n(dr)\right)\\&\quad\leq(2n-3)^2a_0\frac{c_2}{c_1}\,\frac{1}{u_1\mathcal{Z}_n}\left(\frac{V_n}{V_{n+1}}\right)^2=a_0\frac{c_2}{c_1}\,\frac{1}{u_1\mathcal{Z}_n}.\end{aligned}$$

Now by Theorem 13, there exists $\alpha_1>0$ independent of K_n and n such that with probability one, there exists N such that we have

$$\frac{\mathrm{ESS}_n}{K_n}\geq\alpha_1\qquad\forall n\geq N.$$

Let

$$\alpha_2=\inf_{1\leq n\leq N}\frac{\mathrm{ESS}_n}{K_n},\qquad\alpha=\min\{\alpha_1,\alpha_2\}.$$

Note that since ESS_n and K_n are positive (the ESS is at least 1), and the infimum is taken over a finite set, α_2 is positive and does not depend on n. Thus $\alpha>0$ is independent of n and satisfies $\mathrm{ESS}_n\geq\alpha K_n$.

We also note that without the assumption on average branch lengths, a crude estimate gives $\mathcal{Z}_n\geq 2\mathcal{Z}_2/n$, which leads to a linear decay in the upper bound on the ESS. □

Proof of Theorem 15. Since the edge e is chosen from a multinomial distribution weighted by length of the edges, then given any tree $t\in\mathcal{T}_{n+1}$ obtained from the parent tree $\varrho(t)$ by choosing edge e, distal position x and pendant length y, we have

$$Q^n(\varrho(t),t)=\frac{V_{n+1}}{V_n}\,\frac{l_e(\varrho(t))}{l(\varrho(t))}\,p_X(x)\,p_Y(y),$$

where $l_e(r)$ and $l(r)$ are the length of edge e and the total tree length, respectively, and V_n and V_{n+1} are the numbers of tree topologies of $\mathcal{T}_n$ and $\mathcal{T}_{n+1}$.

We denote

$$u_1=\frac{1}{C_0}\int_0^b m(y)^S\,dy,\qquad u_2=\int_0^b\frac{M(y)^{2S}}{p_Y(y)}\,dy,$$

and recall that

$$\sum_e \frac{1}{l_e(r)} \int_0^{l_e(r)} \frac{1}{p_X^e(x)}\, dx \le C \sum_e l_e(r) = Cl(r),$$

where C is the constant from Assumption 5, and

$$\mathcal{Z}_n = \int_{r\in\mathcal{T}_n} \bar{\pi}_n(r)\zeta(r)\, dr \ge c$$

from the assumption on the average branch length (Assumption 4). We have

$$\begin{aligned}
&\int_{t\in\mathcal{T}_{n+1}} \frac{\bar{\pi}_{n+1}^2(t)}{\bar{\pi}_n(\varrho(t))\, Q^n(\varrho(t),t)}\, \mu_{n+1}(dt)\\
&= \int_{t\in\mathcal{T}_{n+1}} \frac{\bar{\pi}_{n+1}^2(t)}{\bar{\pi}_n^2(\varrho(t))} \frac{1}{Q^n(\varrho(t),t)}\, \bar{\pi}_n(\varrho(t))\, \mu_{n+1}(dt)\\
&\le \left(\frac{V_n}{V_{n+1}}\right)^2 \frac{1}{\mathcal{Z}_n^2} \int_{r\in\mathcal{T}_n} \sum_e\\
&\quad \times \int_{x,y} \frac{M(y)^{2S}}{u_1^2} \frac{l(r)}{l_e(r)} \frac{1}{p_X^e(x)} \frac{1}{p_Y(y)} \bar{\pi}_n(r)\, dx\, dy\, \mu_n(dr)\\
&= \left(\frac{V_n}{V_{n+1}}\right)^2 \frac{1}{\mathcal{Z}_n^2} \frac{u_2}{u_1^2} \int_{r\in\mathcal{T}_n} \left(\sum_e \frac{1}{l_e(r)} \int_0^{l_e(r)} \frac{1}{p_X^e(x)}\, dx\right)\\
&\quad \times l(r)\, \bar{\pi}_n(r)\, \mu_n(dr).
\end{aligned}$$

By the assumption of maximum branch length b, we have

$$\begin{aligned}
&\int_{t\in\mathcal{T}_{n+1}} \frac{\bar{\pi}_{n+1}^2(t)}{\bar{\pi}_n(\varrho(t))\, Q^n(\varrho(t),t)}\, \mu_{n+1}(dt)\\
&\le C(2n-3)^2 \left(\frac{V_n}{V_{n+1}}\right)^2 \frac{1}{\mathcal{Z}_n^2} \int_{r\in\mathcal{T}_n} \bar{\pi}_n(r)\zeta^2(r)\, \mu_n(dr)\\
&\le \frac{Cb^2}{c^2}.
\end{aligned}$$

Thus by Theorem 13 there exists $\alpha > 0$ independent of K_n and n such that $\mathrm{ESS}_n \ge \alpha K_n$. □

References

Andrieu, C., Doucet, A., Holenstein, R. 2010. Particle Markov chain Monte Carlo methods. J. R. Stat. Soc. Series B Stat. Methodol. 72:269–342.

Andrieu, C., Doucet, A., Punskaya, E. 2001. Sequential Monte Carlo methods for optimal filtering. In: Doucet, A., de Freitas, N., Gordon, N., editors. *Sequential Monte Carlo Methods in Practice*. New York: Springer. p. 79–95.

Bengtsson, T., Bickel, P., Li, B. 2008. Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. In: *Probability and Statistics: Essays in Honor of David A. Freedman*. Institute of Mathematical Statistics, p. 316–334.

Berger, S. A., Krompass, D., Stamatakis, A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst. Biol. 60:291–302.

Beskos, A., Crisan, D., Jasra, A. 2014. On the stability of sequential Monte Carlo methods in high dimensions. Ann. Appl. Probab. 24:1396–1445.

Bickel, P., Li, B., Bengtsson, T. 2008. Sharp failure rates for the bootstrap particle filter in high dimensions. In: *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*. Institute of Mathematical Statistics, p. 318–329.

Bouchard-Côté, A. 2014. SMC (Sequential Monte Carlo) for Bayesian phylogenetics. In: Chen, M.-H., Kuo, L., Lewis, P.O., editors. *Bayesian Phylogenetics: Methods, Algorithms, and Applications*. Boca Raton (FL): CRC Press. p. 163–186.

Bouchard-Côté, A., Sankararaman, S., and Jordan, M.I. 2012. Phylogenetic inference via sequential monte carlo. Syst. Biol. 61:579–593.

Caporaso, J. G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266–267.

Chopin, N. 2004. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. Ann. Stat. 32:2385–2411.

Crisan, D., Doucet, A. 2002. A survey of convergence results on particle filtering methods for practitioners. IEEE Trans. Signal Process. 50:736–746.

Cueto, M. A., Matsen, F.A. 2011. Polyhedral geometry of phylogenetic rogue taxa. Bull. Math. Biol. 73:1202–1226.

Del Moral, P. 1998. A uniform convergence theorem for the numerical solving of the nonlinear filtering problem. J. Appl. Probab. 35:873–884.

Del Moral, P., Patras, F., Rubenthaler, S. et al. 2009. Tree based functional expansions for Feynman–Kac particle models. Ann. Appl. Probab. 19:778–825.

Dinh, V., Matsen, F.A. 2016. The shape of the one-dimensional phylogenetic likelihood function. Ann. Appl. Probab. (in press) http://arxiv.org/abs/1507.03647.

Douc, R, Moulines, E. 2008. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. Ann. Stat. 36:2344–2376.

Doucet, A., Johansen, A.M. 2009. A tutorial on particle filtering and smoothing: Fifteen years later." In: *The Oxford handbook of nonlinear filtering*. Oxford: Oxford University Press. p 656–704.

Felsenstein, J. 2004. Inferring phylogenies, vol. 2. Sunderland (MA): Sinauer Associates Sunderland.

Fourment, M., Claywell, B.C., Dinh, V., McCoy, C., Matsen IV, F.A., Darling, A. E. 2017. Effective online Bayesian phylogenetics via sequential Monte Carlo with guided proposals. *bioRxiv*, submitted to Syst. Biol. 145219. https://www.biorxiv.org/content/early/2017/06/02/145219.

Gardy, J., Loman, N.J., Rambaut, A. 2015. Real-time digital pathogen surveillance—the time is now. Genome Biol. 16:155.

Heath, T.A., Hedtke, S.M., Hillis, D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Systemat. Evol. 46:239–257.

Izquierdo-Carrasco, F., Cazes, J., Smith, S.A., Stamatakis, A. 2014. PUmPER: phylogenies updated perpetually. Bioinformatics 30:1476–1477.

Katoh, K., Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kong, A., Liu, J.S., Wong, W.H. 1994. Sequential imputations and Bayesian missing data problems. J. Am. Stat. Assoc. 89:278–288.

Künsch, H. R. 2005. Recursive Monte Carlo filters: algorithms and theoretical analysis. Ann. Stat. 33:1983–2021.

Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A. 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5:e1000520.

Liu, J. S., Chen, R. 1995. Blind deconvolution via sequential imputations. J. Am. Stat. Assoc. 90:567–576.

Matsen, F., Kodner, R., Armbrust, E.V. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538.

Neher, R.A., Bedford, T. 2015. nextflu: real-time tracking of seasonal influenza virus evolution in humans. Bioinformatics. 31:3546–3548.

Oudjane, N., Rubenthaler, S. 2005. Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. Stoch. Anal. Appl. 23:421–448.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J.H.J., Becker-Ziaja, B., Boettcher, J.P., Cabeza-Cabrerizo, M., Camino-Sánchez, A., Carter, L.L., Doerrbecker, J., Enkirch, T., García-Dorival, I., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L.E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C.H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallasch, E., Patrono, L.V., Portmann, J., Repits, J.G., Rickett, N.Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R.L., Zekeng, E.G., Racine, T., Bello, A., Sall, A.A., Faye, O., Faye, O., Magassouba, N., Williams, C.V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K.Y., Diarra, A., Savane, Y., Pallawo, R.B., Gutierrez, G.J., Milhano, N., Roger, I., Williams, C.J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., Turner, D.J., Pollakis, G., Hiscox, J.A., Matthews, D.A., M. K. O'Shea, Johnston, A.M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Wölfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S.A., Koivogui, L., Diallo, B., Keïta, S., Rambaut, A., Formenty, P., S. Günther, Carroll, M. W. 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature 530:228–232.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. 2012. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Snyder, C., Bengtsson, T., Bickel, P., Anderson, J. 2008. Obstacles to high-dimensional particle filtering. Mon. Weather Rev. 136:4629–4640.

Suchard, M.A., Redelings, B.D. 2006. BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. Bioinformatics 22:2047–2048.

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.